



Répertoire international des technologies langagières

Présenté au
Centre de recherche en technologies langagières

Préparé par Infolog Inc.
Mars 2006

TABLE DES MATIÈRES

AVANT-PROPOS	8
ALLEMAGNE	9
LABORATOIRES	9
<i>Applied Informatics / Cooperative Systems, Technical University of Munich</i>	9
<i>Artificial Intelligence Research Koblenz</i>	10
<i>Chair of Computer Science VI: Human Language Technology and Pattern Recognition</i>	11
<i>Competence Center Semantic Web</i>	12
<i>Computational Linguistics Lab, University of Potsdam</i>	13
<i>Computational Linguistics, University of Erlangen-Nuremberg</i>	14
<i>Computational Psycholinguistics Group, University of Saarlandes</i>	16
<i>Experimental Phonetics Group at the Institute of Language Processing, University of Stuttgart</i>	17
<i>German Research Center for Artificial Intelligence</i>	18
<i>Institute of Communication Acoustics</i>	19
<i>Institute of Human-Machine Communication</i>	21
<i>Intelligent Information and Communication Systems</i>	22
<i>Jenna University Language and Information Engineering Lab</i>	23
<i>Language Technology Lab, DFKI</i>	24
<i>Linguistics Department, University of Konstanz</i>	25
<i>Manfred Pinkal Computational Linguistics Group</i>	26
<i>Phonetics Institute, University of Saarlandes</i>	28
ORGANISMES ET ASSOCIATIONS	29
<i>Bavarian Archive for Speech Signals</i>	29
AUSTRALIE	30
LABORATOIRES	30
<i>Centre for Language Technology, Macquarie University</i>	30
<i>Department of Computer Science and Software Engineering, University of Melbourne</i>	32
<i>Knowledge Management Research Group, University of Sydney</i>	33
<i>Speech Hearing and Language Research Centre, Macquarie University</i>	34
<i>Sydney Language Technology Research Group, University of Sydney</i>	35
ORGANISMES ET ASSOCIATIONS	36
<i>Australian Computer Science Society</i>	36
<i>Australian Speech Science and Technology Association</i>	37
<i>KnowledgeNet (ARC Research Network for Next Generation Web Technologies)</i>	38
<i>National Information and Communication Technology Association of Australia</i>	39
<i>Smart Internet Technology CRC</i>	40
BELGIQUE	41
LABORATOIRES	41
<i>Center for Computational Linguistics, Université Catholique de Louvain</i>	41
<i>Center for Processing Speech and Images (PSI)</i>	43
<i>Department of Electronics and Information Processing, Vrije Universiteit Brussels</i>	44
<i>Digital Speech and Signal Processing, Ghent University</i>	45
<i>Théorie des Circuits et Traitement du Signal, Faculté Polytechnique de Mons</i>	46
ORGANISMES ET ASSOCIATIONS	47
<i>Belgian Acoustical Society (ABAV)</i>	47
CHINE ET TAIWAN	48

LABORATOIRES	48
<i>Institute of Information Science, Academia Sinica</i>	<i>48</i>
<i>National Taiwan University , Department of Computer Science and Information Engineering</i>	<i>49</i>
<i>National Tsing Hua University, National Tsing Hua University.....</i>	<i>50</i>
<i>Spoken Language Group, Aacademia Sinica</i>	<i>51</i>
CORÉE	52
LABORATOIRES	52
<i>Postech Intelligent Software Lab.....</i>	<i>52</i>
DANEMARK.....	53
LABORATOIRES	53
<i>Center for Language Technology, Faculty of Humanities, University of Copenhagen.....</i>	<i>53</i>
<i>Department of Computational Linguistics, Copenhagen Business School.....</i>	<i>54</i>
<i>Department of Nordic Studies and Linguistics, University of Copenhagen</i>	<i>55</i>
ESPAGNE.....	56
LABORATOIRES	56
<i>Artificial Intelligence section, Software Dept., Technical University of Catalonia.....</i>	<i>56</i>
<i>Compilers and Languages, Escuela Superior de Ingeniería Informática, Orense</i>	<i>57</i>
<i>Database Lab, University of A Coruña.....</i>	<i>58</i>
<i>Department of Artificial Intelligence, Technical University of Madrid</i>	<i>59</i>
<i>IXA Group.....</i>	<i>60</i>
<i>Laboratory for Research and Development in Artificial Intelligence.....</i>	<i>61</i>
<i>Natural Language Processing and Information Retrieval Group at National Distance Learning University (UNED).....</i>	<i>62</i>
<i>Natural Language processing Group, Technical university of Catalonia.....</i>	<i>63</i>
<i>Speech Processing Group (VEU), Technical University of Catalonia.....</i>	<i>64</i>
<i>The Research Group on Signals, Telematic and Communications, University of Granada.....</i>	<i>65</i>
ORGANISMES ET ASSOCIATIONS	66
<i>Catalan Association of Artificial Intelligence</i>	<i>66</i>
ÉTATS-UNIS.....	67
LABORATOIRES	67
<i>Brain and cognitive sciences, Rochester University.....</i>	<i>67</i>
<i>Cambridge research Lab</i>	<i>68</i>
<i>Center for advanced information processing.....</i>	<i>69</i>
<i>Center for automated learning and discovery.....</i>	<i>70</i>
<i>Human Language Technology Research institute</i>	<i>71</i>
<i>Center for Human-Computer Communication.....</i>	<i>72</i>
<i>Center for intelligent information retrieval.....</i>	<i>73</i>
<i>Center for language and speech processing.....</i>	<i>74</i>
<i>Center for machine translation</i>	<i>76</i>
<i>Center for spoken language research.....</i>	<i>77</i>
<i>Center for Study of Language and Information.....</i>	<i>78</i>
<i>Center of Excellence for Document Analysis and Recognition</i>	<i>79</i>
<i>Computational Linguistics and Information Processing Laboratory.....</i>	<i>80</i>
<i>Computational Linguistics at Stanford University.....</i>	<i>81</i>
<i>Columbia Natural Language Processing Group.....</i>	<i>82</i>
<i>Computing research laboratory</i>	<i>83</i>
<i>Department of Cognitive and Linguistic Sciences, Brown University</i>	<i>84</i>
<i>Haskins Laboratories.....</i>	<i>85</i>
<i>Human-Computer Interaction Institute.....</i>	<i>86</i>

<i>USC Natural Language Group</i>	87
<i>International Computer Science institute, Speech Group</i>	88
<i>Linguistic Grammars Online</i>	89
<i>Maryland Information and Network Dynamics Lab (MIND)</i>	91
<i>MIT Media Laboratory</i>	92
<i>Research Cluster on Language, Center for Cognitive Science, Rutgers State University</i>	93
<i>Robust Speech Processing Group Center for Spoken Language Research</i>	94
<i>The Signal, Speech and Language Interpretation Lab at the University of Washington</i>	95
<i>Speech Technology and Research Laboratory</i>	97
<i>Spoken Language Systems</i>	98
ORGANISMES DE FINANCEMENT ET ASSOCIATIONS	99
<i>American association of artificial intelligence</i>	99
<i>Applied Voice Input/Output Society (AVIOS)</i>	100
<i>Association for machine translation of America</i>	101
<i>IEEE Signal Processing Society</i>	102
<i>Linguistic data Consortium</i>	103
<i>Retrieval Group</i>	104
<i>Open Mind Initiative</i>	105
FINLANDE	106
LABORATOIRES	106
<i>From data to Knowledge research unit</i>	106
<i>Helsinki Institute for Information Technology, Basic Research Unit</i>	107
<i>Laboratory of Acoustics and Audio Signal Processing</i>	108
<i>Multimodal Interfaces Group, Adaptive Informatics Research center</i>	109
<i>Semantic Computing Research Group</i>	110
FRANCE	111
LABORATOIRES	111
<i>Analyse et traitement informatique de la langue française (ATILF)</i>	111
<i>École nationale supérieure de télécommunications, Groupe Information, Interaction, Intelligence</i>	112
<i>Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Département Communication Homme-Machine</i>	113
<i>Laboratoire des sciences de l'information et des systèmes, INCOD</i>	114
<i>Laboratoire d'Automatique Documentaire et Linguistique</i>	115
<i>Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur</i>	116
<i>Laboratoire de Linguistique Formelle</i>	117
<i>LORIA Laboratoire lorrain de recherche en informatique et ses applications</i>	118
<i>Laboratoire informatique d'Avignon</i>	119
<i>Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen</i>	120
<i>Groupe de traitement du langage parlé, Laboratoire d'Informatique et de Mécanique pour les Sciences de l'Ingénieur</i>	121
ORGANISMES ET ASSOCIATIONS FRANÇAISES	122
<i>Association francophone de la communication parlée</i>	122
<i>Association pour la Recherche Cognitive</i>	123
<i>Association pour le traitement automatique des langues</i>	124
<i>European Language Resources Association (ELRA)</i>	125
<i>Association Internationale de la communication parlée</i>	126
<i>Réseau Francophone de l'Ingénierie de la Langue (FRANCIL)</i>	127
<i>Portail Technolangue</i>	128
HONG KONG	129
LABORATOIRES	129
<i>Haliday Centre for Intelligent Applications of Language Studies</i>	129

INDE 130

LABORATOIRES 130
AU-KBC Research Center, Madras Institute of Technology..... 130
Center for Development of Advanced Computing 132
Computer Vision and Pattern Recognition Unit, Indian Statistical Institute of Calcutta..... 133

IRLANDE 134

LABORATOIRES 134
Digital Signal Processing, University College Dublin 134
Information Management Group, Dublin College University..... 135
Knowledge and Data Engineering Group, Trinity College Dublin 136
Language and Intelligence Group, Dublin City University..... 137
National Center for Language Technologies 138
TCD Computational Linguistics Group..... 139

ISRAËL 140

LABORATOIRES 140
Department of Computer Science, Bar-Ilan University..... 140
Department of Computer Science, University of Haifa..... 141
Israël Institute of Technology, Computational Linguistics Group..... 142
Knowledge Center for Processing Hebrew 143

ITALIE 144

LABORATOIRES 144
Center for Scientific and Technological Research, Cognitive and Communication Technologies Division.... 144
Cognitive and Communication Technologies Division, ITC-Irst..... 146
Department of computer and systems science, University of Rome..... 147
Department of Information Engineering, University of Padova..... 148
European Academy Bolzano, Institute for Specialised Communication and Multilingualism..... 149
Institute of Cognitive Sciences and Technologies..... 151
Istituto di Elaborazione della Informazione..... 152
Interactive Sensory System Division, SSI..... 153
ISTI, Knowledge Discovery and Delivery Laboratory..... 156
Istituto di Linguistica Computazionale del CNR..... 157
Knowledge Representation Research Center, Free University of Bozen-Bolzano..... 158
Laboratory fo Applied Ontology..... 160
ORGANISMES ET ASSOCIATIONS 161
Center for the Evaluation of Language and Communication Technologies..... 161
FoLLI, Association of Logic, Language and Information..... 162
Tuscan Word Centre..... 163

JAPON 164

LABORATOIRES 164
Advanced Telecommunications Research Institute International ATR..... 164
Computational Linguistics Laboratory..... 165
Itaku Laboratory 166
Kasuya Lab (Klab)..... 167
Language Media Laboratory 168
National Institute of Advanced Industrial Science and Technology (AIST)..... 169
Shikano Lab 170
Tokyo Institute of Technology 171
ORGANISATIONS ET ASSOCIATIONS 172
Asian-Pacific Association for Machine Translation..... 172

<i>Information Processing Society of Japan</i>	173
PAYS BAS	174
LABORATOIRES	174
<i>Computational Linguistics and AI Group, Tilburg University</i>	174
<i>Mathematics and Computer Science? Faculty, University of Twente</i>	175
<i>Institute for Logic, Language and Computation</i>	176
<i>Dutch Research School for Information and Knowledge Systems</i>	177
<i>Human Computer Studies Laboratory, University of Amsterdam</i>	178
<i>IRIS Information Retrieval and Information, Universtiy of Nijmegen</i>	179
<i>The Leiden University Centre for Linguistics</i>	180
<i>Utrecht Istitute of Linguistics OTS</i>	181
ORGANISMES ET ASSOCIATIONS	182
<i>Acoustical Society of the Netherlands</i>	182
<i>European Network of Excellence in Human Language Technologies</i>	183
ROYAUME UNI	184
LABORATOIRES	184
<i>Center for communication Interface Research</i>	184
<i>The Center for Speech Technology Research</i>	185
<i>Center for vision, Speech and Signal Processing</i>	187
<i>Natural Language and Information Processing Group</i>	189
<i>Natural Language Processing and AI</i>	190
<i>Natural Language Processing Research Group</i>	191
<i>Institute for Communication and Collaborative Systems</i>	192
<i>School of Computer Science, University of Manchester, IMG, NLE</i>	193
<i>Scool of computing, Robert Gordon University</i>	194
<i>Computing Science at Aberdeen</i>	195
<i>Language and computation group</i>	197
<i>National Center for Text Mining</i>	198
<i>Natural Language and Computational Linguistics</i>	199
<i>Oxford Computational Linguistics Group</i>	201
<i>UCL Department of Phonetics and Linguistics</i>	202
<i>University Center for Computer Corpus Research on Language</i>	204
SUÈDE	206
LABORATOIRES	206
<i>Department of Linguistics, Lund University</i>	206
<i>Department of Linguistics, Stockholm University</i>	207
<i>Graduate School of Language Technology</i>	208
<i>Human Computer Interaction and Language Engineering Laboratory (HUMBLE), Swedish Institute of Computer Science</i>	209
<i>Interaction and Presentation Laboratory</i>	210
<i>Natural Language Processing Laboratory, Linköping University</i>	211
<i>Speech and Communications Technology Group</i>	212
SUISSE	213
LABORATOIRES	213
<i>Artificial intelligence Laboratory, Federal Institute of Technology in Lausanne</i>	213
<i>Département de linguistique, Université de Genève, Laboratoire d'analyse et de technologie du langage</i>	215
<i>Institute of Computational Linguistics, University of Zurich</i>	216
ORGANISMES ET ASSOCIATIONS	217
<i>European Association for Machine Translation</i>	217

Avant-propos

Le présent document est la résultante d'un mandat de recherche accordé par le Centre de recherche en technologies langagières (CRTL) à Infolog Inc., entreprise de recherche et de communication. Il s'agit d'un annuaire répertoriant les principaux laboratoires universitaires étrangers effectuant des travaux de recherche et développement dans le domaine de l'industrie des technologies langagières. Il rassemble de plus les institutions de financement majeures et organismes associatifs de ce même domaine.

Les travaux de recherche ont été consacrés à un nombre limité de pays, afin de circonscrire le contenu de ce document aux nations comportant les plus importants bassins de recherche dans le domaine des technologies langagières. La majorité des institutions ci-regroupées se trouvent donc en Amérique du Nord et en Europe de l'Ouest. Sont de ce fait incluses les institutions se trouvant aux États-Unis, au Royaume-Uni, en France, en Italie, en Allemagne, en Espagne, en Irlande, en Belgique, aux Pays-Bas ainsi que dans la péninsule scandinave. Des laboratoires d'Australie, d'Israël, et d'Asie (Chine, Japon, Corée, Inde et Hong Kong) sont également décrits dans ce document. Il est toutefois important de souligner que seuls les institutions disposant de textes informatifs en version anglaise ou française ont été prises en compte dans ce répertoire. Cette contrainte n'a cependant eu qu'une influence mineure sur la collecte de données, la majorité des activités du domaine des technologies langagières se déroulant en anglais, indépendamment de la langue officielle du pays hôte.

Ce répertoire détaille les travaux de recherche dans l'industrie de la langue ayant cours dans ces diverses régions. La publication de cette information vise à engendrer une plus grande coopération entre les différents intervenants du domaine et de permettre aux organisations canadiennes d'établir d'éventuels Personne-ressources avec leurs pairs étrangers. Ultimement, cet outil facilitera les actions du CRTL, de l'Association canadienne de l'industrie de la langue (AILIA), du Centre national de recherches du Canada (CNRC) ainsi que des acteurs gouvernementaux qui supportent les initiatives dans l'industrie de la langue.

Le répertoire est structuré de façon à séparer en dix-neuf (19) sections distinctes les fiches descriptives relatives aux divers pays étudiés. Chacune de ces sections est à son tour subdivisée en deux parties. On retrouve en premier lieu la liste des laboratoires de recherche, puis celle des organismes et associations.

Toutes ces institutions sont décrites au sein de ce répertoire au moyen de fiches individuelles. Ces fiches comportent les coordonnées civiques et électroniques utiles, le nom des responsables à Personne-ressourcer, ainsi qu'une brève description des projets de recherche en cours, des produits en découlant ou de la mission et de la portée, dans le cas des organismes.

L'élaboration de ce répertoire s'est effectuée en plusieurs étapes échelonnées du mois de janvier 2006 au mois de mars de la même année. Les acteurs de l'industrie langagière ci-présentés ont été répertoriés à partir de listes fournies par les divers organismes de technologies langagières nationaux des pays concernés. Le portail de l'Union Européenne a été particulièrement utile à l'équipe de recherche, démontrant un excellent degré de cohésion au sein de l'Europe. Le critère pris en considération afin de déterminer si un laboratoire était assez important pour être inclus au sein du répertoire était d'être doté d'un laboratoire comprenant un minimum de 20 (vingt) chercheurs. Le contenu des fiches descriptives des organisations a été rédigé à partir d'informations recueillies sur les pages Internet de celles-ci.

Bonne consultation.

Allemagne

Laboratoires

Applied Informatics / Cooperative Systems, Technical University of Munich
TU München, Institut für Informatik, Boltzmannstr. 3, D-85748 Garching bei München Tél. : (49) 89 289 18656 Téléc. : (49) 89 289 18657 Site Web : http://www11.in.tum.de/index.html.en Personne-ressource : Evelyn Genkow Courriel : gemkow@informatik.tu-muenchen.de Nombre de postes : 26
Projets de recherche
La gestion des connaissances vise à acquérir, à utiliser et à préserver les connaissances au sein des organismes et des collectivités. Nous appliquons ainsi les technologies et les concepts modernes comme le Web sémantique et les ontologies. Notre président organise le Münchner Knowledge Management Kreis depuis 2000, méthode qui favorise l'échange de l'information entre l'industrie et l'université.

Artificial Intelligence Research Koblenz

Universität Koblenz-Landau, Campus Koblenz Institut für Informatik AG Künstliche Intelligenz
Universitätsstr. 1 D-56070 Koblenz Germany

Tél. : (49) 0 261/287 2772 Téléc. : (49) 0 261/287 2731

Site Web : <http://www.uni-koblenz.de/FB4/Institutes/IFI/AGKI>

Personne-ressource : Christoph Wernhard

Courriel : wernhard@uni-koblenz.de

Projets de recherche

La Working group Artificial Intelligence appartient à l'un des groupes de recherche les plus renommés d'Allemagne dans les domaines de la déduction automatisée, de la vérification formelle, de la programmation logique et des agents indépendants. Le principal thème de la recherche porte sur le développement de logiciels intelligents. En bout de ligne, les machines devraient fonctionner librement dans des environnements complexes et aideraient efficacement les utilisateurs dans leur processus décisionnel comme dans leurs actions. Des applications échantillons trouvent de l'information pertinente dans un grand nombre de données et contrôlent des robots intelligents dans des environnements naturels et simulés.

Le traitement actuel des modèles dans la logique des propositions est destiné aux tâches difficiles dans la planification, la vérification et le diagnostic. Quant aux champs d'application plus étendus, il est nécessaire d'élargir l'expressivité langagière (p. ex. logique de prédicat de premier ordre, méthodes non monotones, modèles privilégiés, modèles de soutien, programmation d'une série de réponses et logiques de description. Dans le cadre du projet, les aspects relatifs aux applications de telles techniques sont étudiés et mis en œuvre de façon efficace. L'utilité pratique de cette méthode est confirmée grâce à une étude de cas dans le domaine de la gestion des documents.

Chair of Computer Science VI: Human Language Technology and Pattern Recognition

Chair of Computer Science VI, (Lehrstuhl für Informatik VI), Professor Dr.-Ing. H. Ney,
Ahornstr. 55 , D-52056 Aachen, Germany

Tél. : (49) 241 80 21600 Téléc. : (49) 241 80 22219

Site Web : <http://www-i6.informatik.rwth-aachen.de/>

Personne-ressource : Hermann Ney

Courriel : ney@informatik.rwth-aachen.de

Nombre de postes : 20

Projets de recherche

TC-STAR Technology and Corpora for Speech to Speech Translation? TC-STAR est le fruit de six ans de recherche avancée dans toutes les technologies de base qui concernent la traduction vocale : reconnaissance, traduction et synthèse du langage parlé. Le projet visera une sélection de domaines langagiers conversationnels sans contrainte (c.-à-d. les nouvelles et les discours de diffusion ainsi que quelques langues utiles pour l'économie et la société européennes : chinois, anglais européen et espagnol européen. Les problèmes et les objectifs techniques du projet mettront l'accent sur la conception de nouveaux algorithmes et de nouvelles méthodes, l'intégration de connaissances humaines pertinentes disponibles au moment de la traduction dans un cadre de travail guidé par les données. Parmi des exemples de ces nouvelles méthodes, mentionnons l'intégration des connaissances linguistiques dans l'approche statistique de la traduction du langage parlé, la modélisation statistique relative à la prononciation des conversations normales sans contrainte dans la reconnaissance du langage parlé automatique, et enfin les nouveaux modèles acoustiques et prosodiques pour générer la parole expressive dans la synthèse de la parole. ?TC-STAR est financée par l'*Union européenne*.

Competence Center Semantic Web

DFKI GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany

Tél. : (49) 681 302 5325, Téléc. : (49) 681 302 5338

Site Web : <http://www3.dfki.uni-kl.de/ccsw/>

Personne-ressource : Paul Buitelaar

Courriel : paulb@dfki.de

Nombre de postes :

Projets de recherche

- KSpace Content-based Multimedia Analysis;
- SmartWeb Mobile Broadband Access to the Semantic Web;
- NEWS News Intelligence Technology for the Semantic Web;
- Direct-Info Media Monitoring and Multimodal Analysis;
- VleWs Connecting Information and Knowledge in saarland.de;
- EPOS Evolving Personal to Organizational Knowledge Spaces;
- INKASS Ontology Engineering and Lifecycle Mechanisms for Electronic Knowledge Trading Marketplaces.
- FRODO Framework for Distributed Organizational Memories;
- Clockwork Web-based knowledge management.

Computational Linguistics Lab, University of Potsdam

Computational Linguistics, Department of Linguistics, University of Potsdam, Campus Golm / Haus 24, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam-Golm, Germany

Tél. : (49) 0331 / 977 2016 Téléc. : (49) 0331 / 977 2761

Site Web : <http://www.ling.uni-potsdam.de/cl/cl/>

Personne-ressource : Peter Staudacher

Courriel : staudach@ling.uni-potsdam.de

Nombre de postes : 69

Projets de recherche

Recherche et développement dans le Computational Linguistics Group de Potsdam

Le projet D1 de la **Base de données linguistique pour la structure de l'information** fournit l'infrastructure technique pour créer, conserver et retirer les données linguistiques recueillies grâce à la « structure d'information » de SFB 603. Outre la proposition d'un logiciel et d'un matériel adéquats, D1 favorise les projets individuels dans l'annotation des données.

Parsage des rapports de pathologie Le projet consiste à concevoir un système prototype afin de traiter et de retirer les descriptions d'images de pathologie et les diagnostics, particulièrement dans les maladies pulmonaires.

http://www.ling.uni-potsdam.de/cl/cl/res/forsch_lunge.html/SiCoPS/lang_en/SiCoPS.html

Résumé automatique Le projet consiste à élaborer un système pour le résumé automatique des documents (documents et collections). http://www.ling.uni-potsdam.de/cl/cl/res/forsch_summar.html/SiCoPS/lang_en/SiCoPS.html

Face à l'incertitude Le projet consiste à développer des méthodes en cas d'incertitude quant à l'entrée de l'utilisateur dans les systèmes d'interaction vocale.

Lexique digital Nous développons, en collaboration avec le projet « Digital Lexicon of Current German » de la Berlin-Brandenburg Academy of Sciences and Humanities, un lexique morphème allemand très détaillé.

Analyse des structures rhétoriques Les structures rhétoriques permettent de décrire les phénomènes de nature intercausale dans un texte, qui sont en fonction des propriétés sémantiques et de certaines propriétés syntaxiques. Nous les analysons de façon automatique.

URML – Une syntaxe de balisage sous-spécifiée pour les annotations des structures rhétoriques Pour obtenir des mécanismes d'annotation flexibles et extensibles, nous proposons un format XML pour annoter les arbres sous-spécifiés des structures de rhétorique. Nous discutons les diverses décisions de conception concernées, nous illustrons le format avec un exemple et nous dressons le croquis de certaines applications.

Discourse Marker Lexicon DiMLex Il s'agit de conjonctions et d'autres mots fonctionnels qui indiquent le type de corrélation entre les portées des textes adjacents (p. ex. car --> causal, malgré --> concessif). Notre lexique allemand de marqueurs de discours fournit l'information sur la syntaxe et la sémantique par le biais de XML/XSLT. Il sert également pour l'analyse et la production automatiques de textes.

Computational Linguistics, University of Erlangen-Nuremberg

Friedrich-Alexander-Universität Erlangen-Nürnberg, Abteilung für Computerlinguistik,
Bismarckstr. 6 und Bismarckstr. 12 (Rechnerraum 0.320), 91054 Erlangen

Tél. : (49) 09131 85 29250 Téléc. : (49) 09131 85 29251

Site Web : http://www.linguistik.uni-erlangen.de/en_contents/index.php

Personne-ressource : Jörg Kapfer

Courriel : jmkapfer@linguistik.uni-erlangen.de

Nombre de postes : 116

Projets de recherche

- JSLIM (Java Surface Linear Internal Matching) : Interprète pour des grammaires récursives à gauche (left-associative grammars) en appliquant la sémantique de bases de données.
- Textmill
- Groupe d'étude GLDV pour le passage dans la morphologie, la syntaxe et la sémantique
- Un Malaga-Tutorial interactif (Elke Ertel)
- Malaga, un système de grammaire mis en œuvre par Björn Beutel
- Projet de coopération avec la Corée
- Induction grammaticale de la syntaxe allemande (Markus Schulze)
- Analyse sémantique du texte allemand dans le cadre de travail de Leftassociative Grammar (Manfred Hupfer)
- Entrepôt de corpora supporté par des bases de données (Thomas Künneth)
- Morphologie de la langue coréenne dans Malaga (Prof. Dr. Kiyong Lee, Ph.D.)
- LAPT&DA : Un dictionnaire avec un partitionnement automatique des parties du dictionnaire par domaine et son activation dynamique pendant le taggage des textes inconnus.

Computational Psycholinguistics Group, University of Saarlandes

FR 4.7 Psycholinguistik, Universität des Saarlandes, Gebäude 17.1, Room 1.15, D-66041 Saarbrücken, Germany

Tél. : + (49) 681 302 6555 Téléc. : (49) 681 302 6561

Site Web : <http://www.coli.uni-saarland.de/groups/MC/>

Personne-ressource : Matthew W. Crocker

Courriel : crocker@coli.uni-sb.de

Nombre de postes : 20

Projets de recherche

- ALPHA : Mécanismes adaptatifs pour le traitement des langues humaines
- Parallélisme dans le passage humain
- Assoziationsnormen für Wörter und Bilder im Fernsehen
- AMLaP-2001 Conference : Architectures et mécanismes de traitement des langues
- Collège européen de la technologie des langues et des systèmes cognitifs

Experimental Phonetics Group at the Institute of Language Processing, University of Stuttgart

IMS -Experimentelle Phonetik- Azenbergstrasse 12 D-70174 Stuttgart, Germany

Tél. : (49) 0 711 121 1379 Téléc. : (49) 0 711 121 1366

Site Web : <http://www.ims.uni-stuttgart.de/phonetik/>

Personne-ressource : Sabine Dieterle

Courriel : Sabine.Dieterle@ims.uni-stuttgart.de

Nombre de postes : 20

Projets de recherche

Synthèse de la parole **SmartWeb** *est un système d'accès mobile au Web sémantique*

Acquisition prosodique *Première acquisition par le langage naturel des corrélats acoustiques de l'accent tonique des mots*

Production prosodique *Un modèle computationnel de la production prosodique orientée vers la cible*

SmartKom *Élément de la sortie de la parole pour un système multimodal*

Schwerpunktprogramm Sprachproduktion *La Neuroanatomic Foundation of Language Production : Examen visuel au moyen de l'imagerie par résonance magnétique (IRM)*

Speech Synthesis *Modélisation du contexte linguistique pour la synthèse de la parole*

Alignment Project *Segmentation des données de la parole pour le Mannheim Institute of German Language (IDS)*

German Research Center for Artificial Intelligence

DFKI GmbH Stuhlsatzenhausweg 3 (Building D 3 2) D-66123 Saarbrücken

Tél. : (49) 0 681 302 5151 Téléc. : (49) 0 681 302 5341

Site Web : <http://www.dfki.de/web/>

Personne-ressource : Brigitte Selzer

Courriel : Brigitte.selzer@dfki.de

Nombre de postes : 38

Projets de recherche

Gnowsis

Durée : 2002 - (2005) **Brève description** : Gnowsis est un outil de logiciel qui rassemble, ordonne et structure l'information ainsi que les données diverses susceptibles de se trouver dans les ordinateurs individuels par les principes du Web sémantique.

NEWS

Durée : D'avril 2004 à mars 2006

Bailleur de fonds : NEWS est parrainé par le Information Society Programme (IST) de la Commission européenne, FP6 - 001906

Brèves descriptions : NEWS a pour objectif de créer une technologie d'intelligence News pour le Web sémantique.

SmartWeb

Durée : D'avril 2004 à avril 2007

Bailleur de fonds : Ministère fédéral allemand de l'Enseignement, des Sciences, de la Recherche et de la Technologie (BMBF)

Brève description : Le projet a pour objectif de jeter les bases pour des interfaces-utilisateurs multimodales

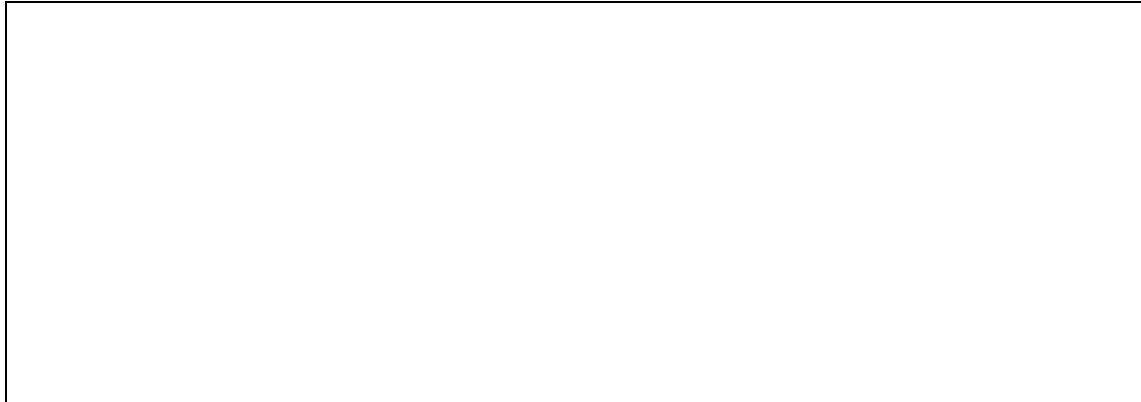
Le projet SmartWeb a pour objectif de jeter les bases d'interfaces multimodales à des services Web distribués et composables sur des appareils mobiles. Le consortium SmartWeb réunit des spécialistes dans divers domaines de recherche : services mobiles, interfaces-utilisateurs intelligentes, technologie de la parole, extraction de l'information et technologies du Web sémantique.

RISE

Durée : De janvier 2004 à décembre 2005

Bailleur de fonds : Ministère fédéral allemand de l'Enseignement, des Sciences, de la Recherche et de la Technologie (BMBF)

Brève description : Rise a pour objectif de concevoir une solution générale pour réutiliser les produits, les méthodes, les modèles et l'ingénierie de logiciels. Le point de départ des travaux de conception est l'exposition avec les défauts et les paramètres caractéristiques (commissions internes) dans la conception de logiciels. De ce projet résulte une approche particulièrement adaptée pour une conception logicielle pratique au profit des petites et moyennes entreprises.



Institute of Communication Acoustics

Institute of Communication Acoustics Building IC 1/132, Ruhr-University Bochum, 44780 Bochum, Germany

Tél. : (49) 0 234 32 22496 Téléc. : (49) 0 234 32 14165

Site Web : http://www.ika.ruhr-uni-bochum.de/index_en.htm

Personne-ressource : Rainer Martin

Courriel : Rainer.Martin@rub.de

Nombre de postes : 28

Projets de recherche

Spoken-dialogue-systems quality : Plusieurs services d'information sont désormais accessibles pour utiliser la technologie de la parole (reconnaissance de la parole, gestion des dialogues et synthèse de la parole) pour un accès convivial à l'information. La qualité de ces systèmes de dialogue vocale dépend de la technologie utilisée, mais également des conditions dans lesquelles ces services sont utilisés (acoustique environnementale, voies de transmission, etc.). En se basant sur les connaissances disponibles, un domaine de travail relativement nouveau est destiné à évaluer la qualité des systèmes de dialogue parlé dans les scénarios d'application réelle. L'évaluation est réalisée à un niveau global (c.-à-d. le système dans l'ensemble) ou diagnostic (évaluation des conséquences des voies de transmission sur la reconnaissance de la parole ou la synthèse de la parole, évaluation de la gestion du dialogue).

Qualité de la transmission de la parole : La qualité des systèmes de transmission de la parole est un autre thème de recherche primordiale du groupe dans le but de recueillir des données auditives pour vérifier et élargir un modèle computationnel pour la qualité de la transmission de la parole. Il est évident que la qualité de la transmission de la parole est liée étroitement à la qualité de la parole-communication en général.

Les dispositifs de communication-parole doivent répondre aux exigences des utilisateurs réels qui comptent sur le but précis des dispositifs, les situations des utilisateurs, leur expérience de dispositifs similaires et plusieurs autres facteurs. En conséquence, la qualité de la communication-parole n'est pas une entité statique; elle se passe plutôt dans une situation précise pour chaque utilisateur. Néanmoins, les planificateurs des dispositifs de la communication-parole comme les systèmes de transmission de la parole doivent généralement satisfaire les besoins des grands groupes d'utilisateurs. À cet effet, il faut être capable de mesurer la qualité de façon auditive ou, mieux encore, de prédire instrumentalement le niveau de qualité qui peut être obtenu avec un dispositif particulier. La mesure et la prédiction sont fondées sur les caractéristiques physiques du dispositif et des signaux de parole.

Éléments qualitatifs de la synthèse de la parole : Les systèmes de synthèse de la parole sont disponibles depuis plusieurs années avec une qualité suffisante pour plusieurs applications simples (par exemple les annonces automatiques ou les lecteurs de courriels pour les non-voyants). En ce qui concerne les autres exigences, la voix naturelle demeure la meilleure sans équivoque. Parmi les objectifs du groupe, mentionnons la création d'une voix synthétique qui semble aussi naturelle, intelligible et flexible qu'une voix humaine. Malgré les nombreux efforts déployés dans l'intelligibilité, l'aspect naturel et la flexibilité demeurent problématiques.

Institute of Human-Machine Communication

Rcisstr. 16, Gebäude S6, 2. Stock, Zimmer S2640

Tél. : (49) 0 89 289 28554 Téléc. : (49) 0 89 289 28535

Site Web : <http://www.mmk.ei.tum.de/>

Personne-ressource : Robert Lieb

Courriel : lieb@ei.tub.de

Nombre de postes : 51

Projets de recherche

La recherche au Human-Machine Communication met l'accent sur les aspects fondamentaux d'une interaction très intuitive, naturelle et donc multimodale entre les hommes et les systèmes de traitement de l'information. Toutes les formes d'interaction, c.-à-d. les modalités, qui sont accessibles aux hommes doivent faire l'objet d'étude à cette fin. La représentation de l'information par la machine ainsi que la technique d'interaction doivent être prises en considération dans ce contexte, entre autres le texte et la parole, le son et la musique, les graphiques et la vision, les gestes et les imitations, et enfin les émotions.

Le traitement des signaux signifie la théorie et l'application dans les domaines suivants : filtrage, codification, transmission, estimation, détection, analyse, reconnaissance, synthèse, enregistrement et reproduction des signaux au moyen de services ou de techniques digitales ou analogues. Le terme signal inclut l'audio, la vidéo, la parole, l'image, la communication, les signaux médicaux et musicaux, ainsi que d'autres signaux dans une forme continue ou discrète (autrement dit échantillonnée). La compétence dans le traitement des signaux est cruciale pour l'élaboration de nouvelles techniques dans la communication homme-machine.

Notre recherche sur le traitement des signaux vise à concevoir des algorithmes et des systèmes capables de reconnaître automatiquement la parole continue dans des conditions de la situation réelle. À cet effet, notre recherche porte sur les **classificateurs statistiques** et les **systèmes** hybrides. La plupart des méthodes sont fondées sur les modèles de Markov cachés stochastiques qui servent de référence pour les sons de la parole (phonèmes) à partir desquels sont construits des mots et des phrases. Les phrases sont analysées par un module de compréhension de la parole, en interprétant le sens. Il faut résoudre les problèmes spéciaux en raison de l'importante variabilité dans la prononciation et de la forte dépendance du conférencier. Nous appliquons à ce moment les variantes de prononciation et les classificateurs adaptatifs.

Intelligent Information and Communication Systems

Fern Universität in Hagen Intelligent Information and Communication Systems

Informatikzentrum Universitätsstraße 1 58084 Hagen Germany

Tél. : 02331 987 374 Téléc. : 02331 987 392

Site Web : <http://pi7.fernuni-hagen.de/homepage/>

Personne-ressource : Hermann Helbig

Courriel : hermann.helbig@fernuni-hagen.de

Nombre de postes : 20

Projets de recherche

Représentation des connaissances avec des réseaux sémantiques multicouches étendus (paradigme multinet)

Le passage du WOrd CIAss based DI[s]ambiguating (WOCADI) – anciennement NatLink – est un programme informatique rédigé dans la langue de programmation Scheme qui transforme un texte allemand en une représentation sémantique formelle grâce au formalisme MultiNet.

Analyse morpho-lexique du texte.

Analyse syntaxique et sémantique du texte.

Jenna University Language and Information Engineering Lab

Fürstengraben, 3007743 Jena, Germany

Tél. : (49) 3641 9 44324 Téléc. : (49) 3641 9 44321

Site Web : http://supreme.coling.uni-jena.de/component/option,com_frontpage/Itemid,1/

Personne-ressource : Joachim Wermter

Courriel : wermter@coling-uni-jena.de

Nombre de postes : 20

Projets de recherche

Le laboratoire JULIE est membre du EU-funded Network of Excellence entitled Semantic Interoperability and Data Mining in Biomedicine (NoE 507505). Sa mission : faire de l'Europe le chef de file scientifique international en informatique médicale et biomédicale. L'objectif à long terme du réseau sera de concevoir des méthodes et des outils génériques qui supportent les tâches critiques dans le domaine, l'exploration des données, la découverte de connaissances, la représentation des connaissances, l'abrégié et l'indexation de l'information, la recherche axée sur la sémantique de l'information dans un espace complexe à grand nombre de dimensions, ainsi que les systèmes adaptateurs basés sur les connaissances, le tout pour un soutien aux décisions destinées à la dissémination d'une médecine fondée sur l'expérience clinique.

Morphosaurus est un outil langagier unique et puissant, qui a été mis au point par le département d'informatique médicale de Freiburg (Allemagne) en collaboration avec le laboratoire JULIE et le Health Technology Master Program de l'Université Parana Catholic de Curitiba (Brésil). Le projet est financé principalement par la Deutsche Forschungsgemeinschaft (DFG), grant KL 640/5-1 and 5-2.

Le laboratoire JULIE est l'un des groupes fondateurs de la BioTem-Initiative. Il s'agit d'un consortium de groupes de recherche allemands de renommée dans le domaine dynamique de l'exploration des textes biomédicaux. Nous avons pour mission de montrer l'importance de l'exploration des textes biomédicaux aux responsables allemands du financement de la recherche, et en bout de ligne, de créer un centre virtuel allemand pour explorer les textes à u niveau institutionnel.

Language Technology Lab, DFKI

DFKI GmbH Forschungsbereich Sprachtechnologie? Stuhlsatzenhausweg 3 / Building 43.1 D-66123 Saarbrücken Germany

Tél. : (49) 681 302 5282 Téléc. : (49) 681 302 5338

Site Web : <http://www.dfki.de/lt/v>

Personne-ressource : Hans Uszkoreit

Courriel : uszkoreit@dfki.de

Nombre de postes : 62

Projets de recherche

- Exploiter – et étendre automatiquement – les ontologies pour le traitement du contenu.
- Intégrer les techniques superficielles et profondes dans le traitement.
- Enrichir le traitement profond avec des méthodes statistiques.
- Combiner la vérification de la langue aux outils de structuration dans la création des documents.
- Indexer les documents en anglais et en allemand.
- Associer automatiquement l'information reconnue à l'information connexe et créer ainsi des connaissances collectives.
- Structurer automatiquement et visualiser l'information extraite.
- Traiter l'information codée dans plusieurs langues, entre autres le chinois et le japonais.

Linguistics Department, University of Konstanz

Wolke, Irene Office, G 213

Tél. : (49) 7531 88 4166 Téléc. : (49) 7531 88 4160

Site Web : <http://ling.uni-konstanz.de/pages/index-en.html>

Personne-ressource : Markus Bader

Courriel : markus.bader@uni-konstanz.de

Nombre de postes : 67

Projets de recherche

Le lexique international : Contours et alignement dans les textes de lecture

L'archive d'universels est accessible sous forme d'archive interrogeable. Il permet à ses utilisateurs en ligne de retirer des mots ou des combinaisons de mots dans leur formulation ou leur documentation. Il suffit également de parcourir toutes les archives des universels

Manfred Pinkal Computational Linguistics Group

Fachrichtung 4.7 Allgemeine Linguistik? Universität des Saarlandes Im Stadtwald - Gebäude C 7.2 Postfach 15 11 50 66041 Saarbrücken Germany.

Tel: (49) 681 302 4344 Téléc. : (49) 681 302 4351

Site Web : <http://www.coli.uni-saarland.de/groups/MP/page.php?id=index>

Personne-ressource : Manfred Pinkal

Courriel : pinkal@coli.uni-sb.de

Nombre de postes : 31

Projets de recherche

CHORUS A trait au traitement sémantique d'énoncés ambigus du langage naturel. Il vise à combiner les préférences et les E de sous-spécification.

CORTE A trait à la reconnaissance automatique, à l'extraction et au traitement des définitions dans les tribunaux allemands en utilisant la technologie linguistique.

Dialog Étudie empiriquement les stratégies flexibles de gestion du dialogue dans les dialogues complexes de tutorat de mathématiques et met au point un système prototype expérimental qui concrétise graduellement les résultats empiriques.

SALSA Crée une ressource de corps annotée sémantiquement et étudie les méthodes de son utilisation.

TALK Vise à généraliser la méthode TRINDI et SIRIDUS « Information State Update ou ISU » à la gestion du dialogue pour mettre au point des systèmes adaptatifs de dialogue multimodal.

InDiGen Vise à concevoir une approche intégrée de planification du discours et de la phrase, qui saisit l'interaction de la sélection du marqueur de discours, les points de suspension et la structure du discours.

COLLATE Est le fruit d'un effort conjoint entre la DFKI, le département de linguistique computationnelle et de phonétique et le département d'informatique de l'Université de Saarland afin de créer un centre de compétence allemand de la technologie linguistique à Saarbruecken.

MILCA est un projet conjoint d'apprentissage en ligne dans lequel diverses positions de mémoire de la linguistique computationnelle imbriquent leur curricula par une variété de cours virtuels. MILCA Saarbrücken conçoit le matériel d'enseignement sur la sémantique computationnelle, les algorithmes de traitement du langage naturel et le traitement du dialogue.

Phonetics Institute, University of Saarlandes

Universität des Saarlandes, FR 4.7 Phonetik, Geb. 17.2, Raum 4.11, Postfach 15 11 50, D-66041 Saarbrücken, Germany

Tél. : (49) 681 302 2926 Téléc. : (49) 681 302 4684

Site Web : <http://www.coli.uni-saarland.de/groups/WB/phonetik/>

Personne-ressource : William Barry

Courriel : wbarry@coli.uni-sb.de

Nombre de postes : 88

Projets de recherche

SecurePhone

Ce projet intègre un dispositif de reconnaissance biométrique dans un PDA 3G/B3G. L'Université de Saarland a pour principale mission la vérification du locuteur de message fixe.

Logox

Synthèse texte-parole de l'allemand et l'anglais à partir de la concaténation de microsegments, conjointement avec G DATA.

Mary

Un système modulaire de synthèse texte-parole de l'allemand, conjointement avec DFKI.

Intonation allemande

Vérification et expansion d'un modèle d'intonation allemande (voir également GToBI).

NaDia

Synthèse de la parole fondée sur les phrases, conjointement avec CLT.

PhonASR

Caractéristiques phonétiques, débit de parole et unités de sous-mots dans le décodage acoustique pour la reconnaissance automatique de la parole.

Speaking lift

Reconnaissance de la parole pour désimposer dans la conception linguistique générale, conjointement avec les départements de linguistique computationnelle et d'informatique.

PrAnK

Analyse prosodique du KielCorpus allemand.

NECA

Un environnement net pour les agents conversationnels émotionnels imbriqués, conjointement avec OEFAI (coordonnateur), DFKI Freeserve, ITRI et Sysis

Phon2

Cours de phonétique en ligne, destiné initialement aux personnes dont la langue maternelle est le bulgare, l'anglais, le français et l'allemand, conjointement avec l'Université de Reading (coordonnateur), l'Université de Sofia et l'Université de Pau.

Profils vocaux

Classification des profils vocaux en santé et pathologiques à partir de microphone et de signaux électroglottographiques, conjointement avec ENT clinic of Caritasklinik St. Theresia, Saarbrücken et les départements de neurologie et de neurochirurgie de l'Université de Saarland de Hamburg. Le projet est financé par le conseil allemand de la recherche (DFG).

Organismes et associations

Bavarian Archive for Speech Signals

Bavarian Archive for Speech Signals c/o Institut für Phonetik, Universität München
Schellingstr. 3 / II 80799 München

Tél. : (49) 89 2180 2758 Téléc. : (49) 89 280 0362

Site Web : <http://www.bas.uni-muenchen.de>

Personne-ressource : Florian Schiel

Courriel : Schiel@phonetik.uni-muenchen.de

Nombre de postes :

Projets de recherche

Le *Bavarian Archive for Speech Signals (BAS)* est un organisme public basé à l'Université de Munich. Il a été fondé pour rendre le corpora de l'allemand parlé actuel accessible à la recherche de base et à la technologie de la parole grâce à une base de données numérique exhaustive du signal de la parole. Le matériel de parole, dont l'évaluation acoustique-phonétique et linguistique-phonétique fait partie intégrante, sera structuré de façon à fournir un accès flexible et précis.

Tâches : Les dernières années ont enregistré une hausse subite de la demande d'importantes collections de données sur le signal de la parole, de la part des chercheurs universitaires qui mènent des recherches de base et des ingénieurs industriels qui travaillent dans le nouveau domaine intégré de la technologie de la parole et de l'information. Les raisons de cette demande sont nombreuses. Il faut d'abord attribuer cette hausse soudaine au rythme effréné des progrès matériels et logiciels qu'a connu le traitement du signal de parole. Le nombre croissant de techniques de traitement de signaux acoustiques-phonétiques et la quantité croissante de données de parole qui peuvent être manipulées et traitées efficacement créent une demande non seulement de matériel de texte intéressant linguistiquement (généralisé bien entendu automatiquement de l'industrie moderne d'impression), mais aussi de matériel de langue parlée qui soit acquis de façon fiable et évalué phonétiquement. Bon nombre d'initiatives nationales et internationales (entre autres BDSON, PHONDAT, LDC, SPEX ou COCODA) ont déjà engendré la collecte et la répartition d'un grand corpora langagier. Toutefois, ces initiatives présentent divers formats conformément aux divers objectifs visés. En ce qui concerne la langue allemande, l'absence d'une institution centrale pour réaliser ces tâches dans une perspective à long terme était évidente. *BAS* aura la responsabilité, en Allemagne, du corpora distribuable de l'allemand, de la collecte, du maintien et de leur accessibilité de façon normalisée.

De plus, *BAS* élaborera ses propres procédures de marquage et de segmentation automatiques pour que les résultats soient accessibles avec le corpora de parole distribué.

Australie

Laboratoires

Centre for Language Technology, Macquarie University
Macquarie University, NSW 2109, Australia Tél. : (61) 2 9850 7111 Site Web : http://www.clt.mq.edu.au/ Personne-ressource : Robert Dale Courriel : ltinfo@ics.mq.edu.au Nombre de postes : 28
Projets de recherche

La recherche du Centre for Language Technology est largement réalisée dans le contexte des projets. Les projets courants et certains autres passés sont ci-énumérés. Voir le site Web des membres du centre pour plus de détails sur les domaines de recherche.

AnswerFinder (coordonnateur : Diego Molla) Ce projet explore comment utiliser les bases de documents existants en tant que source de réponses à des questions relatives au langage naturel. AnswerFinder combine les méthodes traditionnelles de traitement de la langue aux méthodes robustes actuelles. À l'étape d'indexation, AnswerFinder exploite l'information linguistique dans les documents cibles. Il en résulte une représentation simplifiée de la forme logique des phrases d'un document. À l'étape de recherche, la question fait l'objet d'une analyse exhaustive. Une procédure de recherche de secours trouve les phrases dont les formes logiques indiquent l'existence d'une réponse, puis extrait les réponses.

DDD (coordonnateur Robert Dale) Ce projet porte sur la livraison dynamique des documents : l'utilisation des techniques de production du langage naturel pour fournir une information personnalisée à jour via le Web.

Le **Meeting Room Project** (coordonnateur Steve Cassidy) Ce projet vise à créer des applications de technologie de l'information dans une salle de réunion. L'un de nos principaux objectifs consiste à avoir une technologie qui peut être déployée sans instrumentation invasive ou complexe (p. ex. écouteurs, microphone à grande échelle). Nous nous intéressons également à la confluence de la technologie de la parole et du langage et comment les appliquer pour que la technologie de la parole dans la salle de réunion soit une ressource fiable et utile.

PENG est un langage naturel contrôlé et traitable par ordinateur, destinée à la rédaction de cas d'utilisation et de spécifications précises et sans ambiguïté. PENG couvre une sous-série stricte de l'anglais standard. Il est également défini avec précision par une grammaire et un lexique contrôlés. Contrairement aux langues contrôlées, l'auteur n'a pas besoin de connaître les contraintes grammaticales de la langue de façon explicite. ECOLE, éditeur de texte anticipé, souligne les contraintes grammaticales durant la rédaction du texte.

Department of Computer Science and Software Engineering, University of Melbourne

4th Floor, ICT Building, University of Melbourne, 111 Barry St., Carlton Vic 3053, Australia.

Tél. : (61) 3 8344 1301 Téléc. : (61) 3 9348 1184

Site Web : <http://www.cs.mu.oz.au/research/index.html>

Personne-ressource : Ramamohanarao Kotagiri

Nombre de postes : 77

Projets de recherche

Technologie du langage humain

Le langage et la communication humaines sont des phénomènes extrêmement complexes et naturels. L'analyse de la langue, que ce soit la parole, le texte ou des formes multimodales, pose un défi computationnel. À fur et à mesure que l'information sur le Web prend de l'ampleur, le besoin de systèmes d'analyses automatiques de la langue devient plus pressant. Dans la même veine, à mesure que prolifèrent les ordinateurs mobiles et intégrés, le besoin d'interaction linguistique naturelle entre les hommes et les machines devient plus réel. Les technologies de la langue commencent à répondre à ces besoins. Dans le cadre de ce projet, nous approfondissons les questions de recherche ouvertes dans les domaines suivants : modélisation de la langue, annotation linguistique multimodale, traitement haut niveau du langage naturel, documentation électronique des langages menacés de disparaître, modèles structurés pour les données linguistiques et bibliothèques numériques de ressources langagières.

Découverte d'information

Les importantes collections de documents, telles que celles des corporations, posent des défis de différentes façons : le stockage des données efficacement, leur indexation pour une recherche par thème et la façon d'interroger pour retourner les documents susceptibles d'être utiles à l'interrogation de la personne qui fait la recherche. Dans le cadre de ce projet, nous étudions la mise en place efficiente et efficace des systèmes d'extraction de l'information pour les collections de textes très volumineux. Nous nous intéressons particulièrement au stockage et à l'indexation des solutions qui proposent une interrogation rapide et heuristique, et aux techniques qui offrent la perspective de mise à l'échelle pour manipuler les collections volumineuses. Les travaux d'expérimentation incluent la participation à la recherche TREC qui est subventionnée par les États-Unis.

Apprentissage machine et exploration des données

Il est difficile de trouver les formes dans les grandes collections de données et de les utiliser dans le raisonnement. Nous nous intéressons dans ce projet aux algorithmes efficaces pour explorer un nombre de formes et de méthodes fondamentales à utiliser dans la classification.

Knowledge Management Research Group, University of Sydney

School of Information Technologies, Madsen Building, University of Sydney, NSW 2006

Tél. : (61) 2 9351 4291 Téléc. : (61) 2 9351 3838

Site Web : <http://kmrg.it.usyd.edu.au/index.html>

Personne-ressource : Joseph Davis

Courriel : jdavis@it.usyd.edu.au

Nombre de postes : 21

Projets de recherche

La gestion des connaissances (KM) est un nouveau domaine de recherche qui soulève les questions conceptuelles et pratiques relatives à la mise en correspondance, à la codification, à l'extraction et à l'utilisation de diverses formes de connaissances dans les organisations, et la conception ainsi que la mise en œuvre d'outils et de systèmes pour faciliter l'échange des connaissances. Ce domaine de recherche s'appuie sur les développements de plusieurs thèmes et les intègre, entre autres les systèmes de gestion du flux et des documents, le traitement du langage naturel, la découverte de connaissances et l'exploration des textes et du Web, les systèmes organisationnels de mémoire, l'informatique pair-à-pair et l'analyse du réseau social.

Thèmes de recherche :

- Découverte des connaissances;
- Représentation des connaissances;
- Échange et transfert des connaissances;
- Économie des connaissances.

Speech Hearing and Language Research Centre, Macquarie University

Department of Linguistics, Macquarie University, NSW Australia 210

Tél. : (61) 2 9850 8788 Téléc. : (61) 2 9850 9199

Site Web : <http://www.ling.mq.edu.au/shlrc/index.htm>

Personne-ressource : Linda Cupples

Courriel : Linda.cupples@ling.mq.deu.au

Nombre de postes : 67

Projets de recherche

Les membres du Speech Hearing and Language Research Centre (SHLRC) étudient les processus perceptuels acoustiques, verbaux, psycholinguistiques et sociolinguistiques de production et de compréhension de la parole et de la langue. Nous visons ainsi à examiner les processus de la parole, de l'écoute et de la langue chez les gens qui ont des capacités typiques en matière de parole et de langue ainsi que les personnes avec des déficiences de parole, d'écoute ou de langue. Aussi, comprendrons-nous davantage les processus à la base de l'utilisation de la parole, de l'écoute et de la langue ainsi que les façons dont ces processus peuvent être influés par diverses formes de déficience, notamment le traumatisme crânien, la déficience auditive et la déficience développementale.

Le centre compte des spécialistes en parole, en ouïe et en langage, avec des antécédents de travail diversifiés dans des domaines tels que l'audiologie, la linguistique de corpus, la phonétique expérimentale et la phonologie, la psychologie, la pathologie de la parole et de la langue et la sociolinguistique. Les membres du centre adaptent un éventail de perspectives théoriques de plusieurs disciplines scientifiques, y compris la neuropsychologie cognitive, la linguistique clinique, la psycholinguistique, la sociolinguistique, la science et la technologie de la parole et la physiologie. Notre recherche s'appuie sur différentes méthodes méthodologiques, comme la conception expérimentale et les méthodes quantitatives, les études de cas, l'analyse du discours, les méthodes qualitatives, l'épidémiologie, la recherche longitudinale et les études d'intervention. L'utilisation de ces différentes méthodologies permet aux membres de soulever les questions interdépendantes d'une importance théorique et pratique de façon complémentaire et souvent intégrée, renforçant ainsi mieux la force de recherche du centre.

Sydney Language Technology Research Group, University of Sydney

Sydney Language Technology Research Group, School of Information Technologies,
University of Sydney, NSW 2006

Tél. : (61) 2 9351 4174 Téléc. : (61) 2 9351 3838

Site Web : <http://www.cs.usyd.edu.au/~rcdmnl/>

Personne-ressource : Joseph Davis

Courriel: jdavis@it.usyd.edu.au

Nombre de postes : 20

Projets de recherche

- Exploration de données semi-structurées;
- Ontologie de base de la langue basque;
- Ontologie de base de la langue basque (certains résultats immédiats ;
- Phonologie diachronique;
- Passage-éditeur intelligent d'auto-apprentissage pour les bases de données des dictionnaires (description de project) ;
- Matériel de référence électronique pour la traduction et l'apprentissage de la langue – validation de principe;
- Méthodologie de flux pour la régénération automatique de publications linguistiques multimédias (description de projet);
- Architecture et ingénierie logicielles pour les systèmes TL;
- Analyseurs syntaxiques fondés sur des éléments probants;
- Ensemble d'utilitaires pour la classification de documents;
- Reconnaissance de l'entité nommée;
- Traitement d'un ontolexique;
- Un ensemble d'utilitaires pour le balisage de la linguistique fonctionnelle systémique;
- Méthodes d'extraction des connaissances tacites pour la gestion des connaissances.

Organismes et associations

Australian Computer Science Society

Level 3 160 Clarence Street, SYDNEY NSW 2000

Tél. : (61) 2 9299 3666 Téléc. : (61) 2 9299 3997

Site Web : <http://www.acs.org.au/>

Personne-ressource : Dennis Furini

Courriel : dennis.furini@acs.org.au

Projets de recherche

L'Australian Computer Society (ACS) est une association reconnue de professionnels dans le domaine de la technologie d'information et de communication (ICT). Elle attire un grand nombre de membres actifs de tous les niveaux de l'industrie de technologie d'information et de communication. Membre du Australian Council of Professions, l'ACS est la voix publique des professionnels de la technologie d'information et des communications et le gardien de l'éthique et des normes de cette industrie. L'ACS a pris un engagement envers la grande communauté pour assurer l'usage bénéficiaire de la technologie d'information et de communications.

Fondée en 1966, l'association a pour objectif d'approfondir l'étude, la science et l'application de la technologie d'information; de promouvoir, de développer et de suivre les capacités en termes de pratique de la technologie d'information et de communications par les personnes et les organisations; concevoir et promouvoir un code d'éthique à l'intention des membres de l'association; définir et promouvoir les normes de connaissances de la technologie d'information et de communications; promouvoir la formulation de stratégies efficaces sur la technologie d'information et de communications et d'autres thèmes connexes; développer les connaissances et élargir la vision de la technologie d'information et de communications dans la communauté; promouvoir les avantages de l'adhésion à l'association et promouvoir les avantages de recruter des membres de l'association.

Les membres de l'ACS travaillent dans tous les domaines des affaires et de l'industrie, du gouvernement et du milieu universitaire. Ils possèdent les qualifications et l'expérience dans le domaine de la technologie d'information et de communications, et s'engagent à respecter le code d'éthique de l'association ainsi que le code d'éthique professionnel et d'exercice professionnel. Les membres de l'ACS sont des professionnels.

Principal objet

Promouvoir le développement de ressources australiennes de technologie d'information et de communications.

Australian Speech Science and Technology Association

Site Web : <http://www.assta.org/>

Personne-ressource : Dennis Burnham

Courriel : secretary@assta.org

Mission

L'Australian Speech Science and Technology Association (ASSTA) est une association scientifique qui vise à améliorer les connaissances de la science de la parole et son application à la technologie de la parole de façon à les adapter à la Nouvelle-Zélande et à l'Australie. Pour ce faire, l'association veut assurer un échange considérable entre ceux qui sont engagés activement dans la science de la parole et la recherche ainsi que la pratique de la recherche, entre ces professionnels pluridisciplinaires et la communauté dans l'ensemble, et entre les organismes nationaux et internationaux qui ont des objectifs inhérents. L'association met à jour régulièrement sa constitution pour refléter ses besoins changeants. Elle est incorporée dans le Australian Capital Territory en tant qu'association scientifique sans but lucratif qui a des objectifs précis pour promouvoir sa mission.

KnowledgeNet (ARC Research Network for Next Generation Web Technologies)

ARC, Macquarie University, NSW Australia 210

Site Web : <http://www.comp.mq.edu.au/research/semantic/index.html>

Personne-ressource : Abhaya Nayak

Courriel : abhaya@ics.mq.edu.au

Nombre de postes : 100

Mission

Le World Wide Web (WWW) a été conçu en tant qu'espace d'information pour qu'il soit utile non seulement pour la communication entre hommes, mais également pour que les machines soient capables de participer et d'aider. Pour tirer pleinement parti de son potentiel, le Web sémantique (connu par beaucoup comme étant la prochaine génération du Web) augmente le WWW en ajoutant un contenu compris par une machine aux ressources Web, en fournissant plus de données sur le contenu et en permettant aux agents de remplacer les hommes par le raisonnement avec l'information obtenue. À l'échelle internationale, le développement et l'exploitation du Web sémantique sont précoces, mais très actifs. Pour exploiter au maximum le pouvoir du Web sémantique, nous devons mettre en valeur les diverses activités de recherche menées dans des domaines des mots apparentés, notamment les technologies des agents, les technologies Web, la sécurité Web, l'ingénierie des connaissances, la visualisation de l'information, le traitement d'images et les technologies de collaboration. Les activités ciblées dans KnowledgeNet, réseau de recherche pour les technologies Web de la prochaine génération, permettront aux chercheurs australiens de collaborer dans des programmes de recherche internationaux qui permettront de façonner les principales technologies à la base de la prochaine génération du WWW et de réaliser la vision de l'eAustralie.

KnowledgeNet a évolué d'une tentative de prise de financement initial intitulé « Intelligent Applications Through The Semantic Web » (SR0354476). Il a attiré de nouveaux membres et de nouvelles activités du « Transforming Knowledge Spaces: Open Technologies for Research Collaboration and Research Communication » (SR0354839). Il a également attiré un nombre d'autres chercheurs avec des forces complémentaires.

KnowledgeNet est basé à l'Université Macquarie qui a un dossier exceptionnel dans la promotion de l'excellence de recherche, l'attrait de l'ARC et d'autres fonds compétitifs de recherche, la commercialisation de la recherche et la sensibilisation du public.

National Information and Communication Technology Association of Australia

Australian Technology Park Bay 15 Locomotive Workshop Australian Technology Park
Eveleigh NSW 1430 Australia

Tél. : : (61) 2 9209 4750 Téléc. : (61) 2 9209 4748

Site Web : <http://nicta.com.au/>

Personne-ressource : Clare Gill

Courriel : clare.gill@nicta.com.au

Mission

La technologie d'information et des communications (ICT) a énormément changé le portrait social et économique du monde.

L'Australie possède un talent énorme inexploité en technologie d'information et de communications, selon les normes internationales. Créée dans le cadre de la *Backing Australia's Ability* – une initiative du gouvernement australien visant à promouvoir la science et la technologie – la National ICT Australia (NICTA) vise à exploiter ce potentiel inexploité.

NICTA a été formée par le ministère des Communications, de la Technologie d'information et des Arts du gouvernement australien et le conseil australien de recherche (Australian Research Council). Les partenaires de consortium de la NICTA se composent du Australian Capital Territory Government, du New South Wales Government, de l'University of New South Wales, et l'Australian National University. Cette union puissante a donné lieu à un laboratoire national en se fondant sur la recherche, la commercialisation, l'enseignement et la collaboration. En mettant l'accent sur la recherche de base inspirée par l'utilisation, la NICTA attire, développe et réseaute un talent exceptionnel pour la prospérité future de l'Australie.

La recherche de la NICTA met l'accent sur les problèmes de technologie qui confrontent l'industrie, la communauté et l'intérêt national. La recherche de la NICTA est principalement motivée par les *enjeux prioritaires*.

La création de la technologie pour servir à la société, au marché et à l'environnement inspire les enjeux prioritaires dont les deux premiers sont *Base de confiance des réseaux informatiques sans fil* et *Des données aux connaissances*.

Smart Internet Technology CRC

Smart Internet Technology CRC Bay 8 Suite 9/G12 Australian Technology Park Eveleigh
NSW 1430

Tél. : (61) 2 8374 5080 Téléc. : (61) 2 8374 5090

Site Web : <http://www.smartinternet.com.au/SITWEB/index.jsp>

Personne-ressource: Darrel Williamson

Courriel : darrell.williamson@smartinternet.com.au

Mission

Garantir une recherche efficiente et efficace de calibre mondial avec d'importants résultats technologiques et commerciaux.

Intégrer l'innovation technologique et la recherche sociale et comportementale

Proposer des technologies, des services et des solutions Internet de grande importance qui sont compétitifs sur le marché international.

Offrir des avantages réels et futurs qui dépassent cinq fois la valeur des ressources utilisées.

Produire de professionnels demandés en Australie qui soient informés, spécialisés et dynamiques dans le domaine de la technologie d'information et de communications.

Développer des technologies clés, des solutions et des services visant à accroître la productivité et à réaliser des gains de rendement pour l'Australie grâce à l'utilisation intense de l'Internet.

Prendre part, au moyen du transfert de technologie, au progrès commercial de l'industrie australienne, notamment les entreprises petites à moyennes.

Belgique

Laboratoires

Center for Computational Linguistics, Université Catholique de Louvain

Katholieke Universiteit Leuven, Center for Computational Linguistics
Maria-Theresiastraat 21, B-3000 Leuven

Tél. : (32) 16 32 5088 Téléc. : (32) 16 32 5098

Site Web : <http://www.ccl.kuleuven.be/about.en.php>

Personne-ressource : Geert Adriaens

Courriel : geert.adriaens@ccl.kleuven.ac.be

Nombre de postes : 12, nombre d'étudiants inconnu

Projets de recherche

Traduction automatique : EUROTRA, METAL, METIS, METIS-II.

Syntaxe et sémantique computationnelles : La sémantique et la traduction automatique, Sémantique formelle du discours, Caractéristiques linguistiques pour les futures normes industrielles, MODUL : Spécifications linguistiques monolingues du néerlandais, LS-GRAM+ (néerlandais), LINGUADUCT : interprétation contextuelle des textes de langage naturel par le raisonnement abductif et les connaissances acquises de façon inductive, MultiMeteo (composants du néerlandais), sémantique formelle, Universals et la traduction automatique, Computacionele syntaxis en semantiek, traitement parallèle du langage naturel : PARLEX, une grammaire probabilistique de l'espagnol, SEMADUCT : combiner les techniques déductives et inductives à la sémantique lexicale.

Lexicographie computationnelle : PROTON, représentation des connaissances linguistiques ancrées lexiquement pour les applications de traitement du langage naturel, la construction de dictionnaires bilingues.

Phonologie computationnelle : FONILEX : Une base de données phonologique avec la prononciation du néerlandais en flamand.

Linguistique de corpus : ANNO : Base de données linguistique annotée pour le néerlandais, CGN : Corpus Gesproken Nederlands, D-coi : Initiative de corpus de la langue.

Transcription automatique : ATraNoS : Transcription automatique et normalisation de la parole.

Manipulation des documents : ALCOGRAM, MENELAS, SALOMON : Vers une catégorisation automatique du texte juridique et l'extraction de données juridiques, SECC : Un correcteur simplifié d'anglais.

Développement d'un cours de langue assisté par ordinateur : A.D.A.M. & E.V.E. : Analyse et manipulation automatiques des documents et grande variété d'exercices, CTP : Matériel

exhaustif d'apprentissage, Word Drum, ZÉRO FAUTE, DECIDE : Développement de cours européens sur l'ingénierie de l'information et Datacom, E.E.T. : Liste-vocabulaire exhaustive pour l'enseignement de l'anglais, MYTHE : Thesaurus multimédia éducatif à l'intention des jeunes et des enfants.

Center for Processing Speech and Images (PSI)

U.Leuven - ESAT/PSI Kasteelpark Arenberg 10 3001, Heverlee BELGIUM

Tél. : (32) 16 32 1057 Téléc. : (32) 16 32 1723

Site Web : <http://www.esat.kuleuven.be/psi/spraak/>

Personne-ressource : Patrick Wambacq

Courriel : patrick.wambacq@esat.kuleuven.be

Nombre de postes : 70

Projets de recherche

Le Center for Processing Speech and Images (PSI) compte 70 personnes. Il est l'un des grands centres de recherche du département de génie électrique (ESAT). La recherche couvre une grande partie du traitement audiovisuel de l'information.

Projets de recherche sur la parole La combinaison du traitement de la parole et de l'image au sein du même groupe est un important avantage pour l'avenir, évènement assez unique dans le contexte européen. Nos chercheurs sont issus de divers milieux : ingénieurs, physiciens, mathématiciens, linguistes, etc., une multidisciplinarité qui permet au centre de jumeler l'innovation fondamentale avec des solutions pratiques à l'échelle du grand spectre d'applications. Le grand nombre de projets européens marque notre réputation internationale. Le centre a été le point de départ d'essaimages réussis et un partenaire fiable pour l'industrie.

MIDAS : Solutions de données manquantes (Missing Data Solutions).

ESSCorTeSS : Compression exponentielle sinusoidale de la parole pour la synthèse basée sur le corpus de la parole à partir du texte.

JASMIN-CGN : Jongeren, Andertaligen, Senioren en Machine Interactie voor het Nederlands.

SPACE : Algorithmes de la parole pour les applications cliniques et pédagogiques.

Akoestische Adaptatie van HMMs voor Computerondersteund Leren.

TEMPLATE : Reconnaissance de la parole à partir de données.

FLaVoR – Reconnaissance flexible du vocabulaire très étendu : incorporation de connaissances linguistiques.

**Department of Electronics and Information Processing, Vrije
Universiteit Brussels**

Vrije Universiteit Brussel (VUB) Dept. of Electronics and Information Processing, (ETRO)
Pleinlaan 2 - B-1050 Brussel – Belgium Room 5K321

Tél. : (32) 2 629 29 30 Téléc. : (32) 2 629 28 83

Site Web : <http://www.etro.vub.ac.be/>

Personne-ressource : jpcornel@etro.vub.ac.be

Courriel : Jan Cornelis

Nombre de postes : 107

Projets de recherche

- Intelligibilité de la parole.
- Modification et conversion vocale.
- Transplantation des caractéristiques vocales.
- Systèmes de données à partir de la parole.
- Échelonnement temporel et décalage du pas de la parole et de signaux audio.
- Modélisation et codification paramétrique de signaux audio.
- Requantification psychoacoustique de signaux audio.

Digital Speech and Signal Processing, Ghent University

Universiteit Gent, Vakgroep Elektronica en Informatiesystemen St. Pietersnieuwstraat 41, B-9000 Gent, Belgium

Tél. : (32) 9 264 33 66 Téléc. : (32) 9 264 35 94

Site Web : <http://www.elis.ugent.be/>

Personne-ressource : Jan Van Campenhout

Courriel : Jan.VanCampenhout@elis.ugent.be

Nombre de postes : 16, nombre d'étudiants inconnu

Projets de recherche

Le groupe de chercheurs Digital Speech and Signal Processing (DSSP) du département des systèmes électroniques et d'information (ELIS) de l'Université Ghent concentre ses études sur l'analyse, la synthèse et la reconnaissance de la parole et sur l'analyse automatique du contenu musical.

AUTONOMATA (Automata pour dériver les transcriptions phonémiques des noms néerlandais et flamands). Ce projet porte sur la création d'un convertisseur graphème à phonème (g2p) qui donne de meilleurs résultats dans les noms néerlandais et flamands que les produits actuellement disponibles sur le marché.

SPACE (algorithmes de la parole pour les applications cliniques et de la parole) Ce projet porte sur l'application de la reconnaissance automatique de la parole et les techniques de synthèse pour l'évaluation des voix pathologiques (dysarthrie, parole exprimée par les malentendants) et pour l'évaluation de la prononciation et de la lecture chez les étudiants (dyslexie, non-autochtones).

Théorie des Circuits et Traitement du Signal, Faculté Polytechnique de Mons

MULTITEL-TCTS Lab, Faculté Polytechnique de Mons, Parc Initialis, Copernic Avenue, B-7000 MONS, Belgium

Tél. : (32) 65 37 4730 Téléc. : (32) 65 37 4729

Site Web : <http://tcts.fpms.ac.be/index.php>

Personne-ressource : Joel Hancq

Courriel : joel.hancq@fpms.ac.be

Nombre de postes : 28

Projets de recherche

La synthèse de la parole porte sur la production sémantique de la parole à partir du texte. Nous avons conçu et breveté MBROLA, un algorithme interne de synthèse de la parole. Actuellement, nous coordonnons le projet MBROLA pour la synthèse multilingue de la parole (34 langues grâce à la collaboration internationale). La recherche actuelle met l'accent sur de nouveaux modèles de parole par le traitement de signaux, sur la conversion vocale pour les systèmes de synthèse de la parole ainsi que sur les interfaces en temps réel pour la synthèse audio. Applications : applications multimédias, aides techniques pour les personnes invalides, communication homme-machine, lecteurs de courriels, produits à grande consommation.

Grâce à plusieurs projets européens et nationaux, et en collaboration avec le centre de recherche MULTITEL ASBL, l'équipe ASR soulève plusieurs aspects du problème de reconnaissance de la parole.

Reconnaissance robuste de la parole : Nous avons, grâce à plusieurs projets (RESPITE, MODIVOC, AURORA), étudié minutieusement le problème de la reconnaissance robuste de la parole : soustraction spectrale, filtrage de Wiener, estimation du bruit, tableaux microphones.

Détection de mots clés : Le problème de détection de mots clés est crucial dans l'application en temps réel afin de manipuler partiellement la parole spontanée dans les dialogues homme-machine. Vient ensuite le problème de rejet des mots hors vocabulaire ou des mots peu reconnus, qui a également fait l'objet d'étude par l'estimation des niveaux de confiance pertinents.

Adaptation de modèles : L'adaptation rapide de modèles permet de mettre à jour les modèles statistiques avec très peu de données. L'adaptation au bruit et l'adaptation au locuteur, plus particulièrement dans le cadre des systèmes hybrides HMM/ANN.

Organismes et associations

Belgian Acoustical Society (ABAV)

ABAV General Secretariat CSTC – WCTB avenue Pierre Holoffe 21? B-1342 Limelette
Belgium

Tél : (32) 16 32 7128 Téléc. : (32) 16 32 7984

Site Web : <http://www.abav.be/>

Personne-ressource : Gerrit Vermeir

Courriel : gerrit.vermeir@bwk.kuleuven.ac.be

Mission

L'Association belge des Acousticiens (ABAV) a été fondée en 1966 en tant que société nationale sans profit financier. L'association comprend des personnes œuvrant dans le domaine de l'acoustique, notamment des universitaires, des laboratoires de recherche, des organisations gouvernementales, des bureaux d'études ou des sociétés industrielles dont l'intérêt ou les activités sont en relation avec l'acoustique physique, le contrôle du bruit industriel, l'acoustique des bâtiments, la protection de l'environnement, l'électro-acoustique, la protection contre le bruit, etc.

L'association compte actuellement environ 125 membres (efficaces, associés, étudiants et honoraires) et 40 membres donateurs.

Voici les objectifs de l'ABAV :

Créer un cadre de travail permanent pour l'échange d'information entre les membres relativement à toutes les disciplines acoustiques; recueillir et échanger l'information sur les thèmes d'intérêt pour la société; créer un forum de discussion et d'échange sur la gestion, la protection et le contrôle du bruit dans le but de créer un environnement acoustique acceptable; contribuer au développement et à la promotion de l'acoustique en tant que discipline scientifique; établir des liens solides avec des associations nationales et internationales similaires.

Chine et Taiwan

Laboratoires

Institute of Information Science, Academia Sinica
<p>INSTITUTE OF INFORMATION SCIENCE, ACADEMIA SINICA, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan Tél. : (886) 2 2788 3799 Téléc. : (886) 2 2782 4814 Site Web : http://www.iis.sinica.edu.tw/ Personne-ressource : Jane Win-Shih Liu Courriel : janeliu@iis.sinica.edu.tw Nombre de postes : 54</p>
Projets de recherche
<ul style="list-style-type: none">- Laboratoire de recherche sur des agents Internet intelligents adaptatifs Adaptive.- Technologies avancées de réseautage et Groupe de services.- Laboratoire d'automatisation.- Laboratoire de communication et de systèmes informatiques.- Laboratoire de systèmes d'agents intelligents.- Systèmes de transport intelligents.- Laboratoire de classification et d'apprentissage par machine.- Laboratoire d'automatisation de vérification.- Laboratoire de collaboration Web.

National Taiwan University , Department of Computer Science and Information Engineering

#1 Roosevelt Rd. Sec. 4, Taipei, Taiwan 106, ROC Dept. of CSIE, NTU

Tél. : (886) 2 3366 4888 poste 203 Téléc. : (886) 2 2362 8167

Site Web : <http://www.csie.ntu.edu.tw/>

Personne-ressource : Tei-Wei Kuo

Courriel : chairman@csie.ntu.edu.tw

Nombre de postes : 32

Projets de recherche

- Traitement acoustique : variation des caractéristiques, de la modélisation, de la formation et de la prononciation;
- Traitement linguistique et modélisation langagière;
- Algorithme d'interrogation/de codification et détection de mots clés;
- Gestion multilingue de l'information;
- Ingénierie de langue.

National Tsing Hua University, National Tsing Hua University

Department of Computer Science, National Tsing Hua University, 101, Kuang Fu Rd, Sec.2, HsingChu, Taiwan 300 R.O.C

Tél. : (886) 3 571 4787 Téléc. : (886) 3 572 3694

Site Web : www.cs.nthu.edu.tw

Personne-ressource : Biing-Feng Wang

Courriel : bfwang@cs.nthu.edu.tw

Nombre de postes : 41

Projets de recherche

Recherche sur les techniques d'indexation et d'interrogation pour les multimédias, accent sur le rythme, reconnaissance de l'accord, la synthèse sonore et la reconnaissance vocale;

Conception parallèle du langage, optimisation de compilation pour les systèmes imbriqués.

Spoken Language Group, Academia Sinica

INSTITUTE OF INFORMATION SCIENCE, ACADEMIA SINICA, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

Tél. : (886) 2 2788 3799 Téléc. : (886) 2 2782 4814

Site Web : <http://sovideo.iis.sinica.edu.tw/SLG/index.htm>

Personne-ressource : Hsin-Min Wang

Courriel : whm@iis.sinica.edu.tw

Nombre de postes : 29

Projets de recherche

- Intégration de la reconnaissance de la parole et du locuteur pour l'authentification biométrique
- Développement d'une base de connaissances linguistiques et phonologiques pour la prochaine génération.
- Reconnaissance automatique de la parole.
- Une étude sur la technologie de bases de données pour l'archives audio électronique/vidéo numérique.

Corée

Laboratoires

Postech Intelligent Software Lab.
San 31, Hyoja-Dong, Pohang, 790-784, Korea Tél. : (82) 54 279 5581 Téléc. : (82) 54 279 2299 Site Web : isoft.postech.ac.kr Personne-ressource : Prof. Gary Geunbae Lee Courriel : gblee@postech.ac.kr Nombre de postes : 39
Projets de recherche
<ul style="list-style-type: none">- Reconnaissance de la parole et TTS.- Recherche d'informations et traitement de la langue.- Traitement multilingue de l'information.- Bio-Informatique et exploration de texte.- Gestion du dialogue parlé.- Compréhension robuste de la langue parlée.

Danemark

Laboratoires

Center for Language Technology, Faculty of Humanities, University of Copenhagen

Njalsgade 80, 2300 Copenhagen S.

Tél. : (45) 3532 9090 Téléc. : (45) 3532 9089

Site Web : <http://www.cst.dk/uk/index.html>

Personne-ressource : Bente Maegaard

Courriel : bente@cst.dk

Nombre de postes : 20

Projets de recherche

- Infost – structure d'information dans la langue danoise. Le projet prend fin en mars 2006.
- LEXADV adverbos danois, norvégiens et suédois.
- MULINCO MULTIlingual Corpus de l'université de Copenhague.
- OntoQuery Interrogation basée sur l'ontologie.
- Système MT basé sur l'unification PaTrans pour traduire les brevets de l'anglais au danois.
- Ressources du langage écrit du centre de validation
- Centre danois de documentation DanDokCenter pour les résultats de recherche de HLT.
- Outil DefSum pour résumer les textes danois.
- Système DIALOG PC pour le dialogue parlé en danois aux fins de réservation de vols.
- Évaluation EAGLES des systèmes de traitement du langage naturel.
- Enabler (European National Activities for Basic Language Resources). Le projet a pris fin le 31 août 2003.
- IATE - Services de développement d'un système interactif de bases de données de terminologie.
- Caractéristiques linguistiques LINDA pour le danois.

Department of Computational Linguistics, Copenhagen Business School

CBS, Solbjerg Plads 3 DK-2000 Frederiksberg, Copenhagen

Tél. : (45) 3815 3123

Site Web : http://uk.cbs.dk/forskning_viden/fakulteter_institutter_centre/institutter/sprog/id

Personne-ressource : Sabine Kirchmeier-Andersen

Courriel : ska.id@cbs.dk

Nombre de postes : 23

Projets de recherche

La recherche est réalisée dans le département de linguistique computationnelle dans trois domaines de base :

- **Descriptions formelles de la syntaxe et de la sémantique du langage naturel.**
- **Connaissance de la langue de spécialité (LSP) et modélisation des données.**
- **Traitement du langage statistique.**

Une attention particulière est accordée aux domaines suivants :

Bases de données avec contenu linguistique : La terminologie dynamique et les bases de connaissances avec des interfaces définies par les utilisateurs facilitent la systématisation des concepts dans les systèmes de concept.

Recherche d'information : Les systèmes basés sur l'ontologie offrent une recherche plus intelligente et conviviale que la recherche de caractères d'aujourd'hui.

Interfaçage du langage naturel : Il propose une communication flexible avec l'ordinateur par la parole et l'écrit en danois.

Traduction automatique : La mémoire de traduction, ou la traduction automatique (les systèmes à base de règles sont combinés avec des méthodes statistiques) est une solution aux difficultés linguistiques globales que confrontent les compagnies aujourd'hui.

Linguistique de corpus : Une recherche linguistique moderne à l'aide de corpora – les collections de textes annotés avec une information grammaticale (treebanks) constituent la base empirique de nos travaux.

Department of Nordic Studies and Linguistics, University of Copenhagen

Njalsgade 80, DK-2300 Koebenhavn S, Denmark

Tél. : (45) 35 32 86 40 Téléc. : (45) 35 32 86 35

Site Web : <http://www.cphling.dk/>

Personne-ressource : Nina Grønnum

Courriel : ng@cphling.dk

Nombre de postes : 40

Projets de recherche

Le département de linguistique générale et appliquée fait partie de la faculté ès lettres. Il se trouve dans le complexe universitaire d'Amagar (KUA). Le département a été fondé en 1988 à la suite d'une fusion de Institute of Linguistics, Institute of Phonetics, Institute of Applied and Mathematical Linguistics et le Center for Audiologopedics.

La fusion de ces unités auparavant indépendantes a donné lieu à un département d'un vaste éventail d'activités plus ou moins reliées. Ces unités sont étroitement liées à la théorie et la méthodologie linguistiques générales (linguistique dans son sens global). Il est donc important que la recherche sur la langue danoise ou autre contribue aux descriptions générales théoriques et typologiques. La théorie et la méthodologie linguistiques générales font également l'objet d'étude en tant qu'outil naturel dans la recherche linguistique analytique et descriptive. Mis à part les disciplines plus théoriques, le département s'occupe également d'un nombre de disciplines appliquées.

De septembre 1999 jusqu'à décembre 2001, IAAS a collaboré avec le Center for Person Communication, l'Université Aalborg et TDC (Danish TeleCom) pour améliorer et développer la synthèse texte-parole concatenative du danois. Le projet a été financé par le ministère danois de la Recherche, Ide la Technologie et du Danois. Le système est commercialisé par Speech-Ware.

Espagne

Laboratoires

Artificial Intelligence section, Software Dept., Technical University of Catalonia

Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya.
C/Jordi Girona Salgado, 1-3, Mòdul C6. 08034 Barcelona - Spain

Tél. : (34) 93 401 56 42 Téléc. : (34) 93 401 70 14

Site Web : <http://www.lsi.upc.es/dept/investigacion/sectia/home.html>

Personne-ressource : René Alquézar

Courriel : alquezar@lsi.upc.es

Nombre de postes : 45

Projets de recherche

EUROWORDNET : Concevoir un wordnet multilingue avec des relations sémantiques entre les mots.

Application de systèmes à base de connaissances à la gestion des installations d'épuration des eaux usées et des processus biotechnologiques. Élimination simultanée du dioxyde de carbone, du nitrogène et du phosphore.

Construction et intégration des bases de connaissances pour les systèmes à base d'information.

ITEM : Recherche d'information textuelle dans un environnement multilingue en utilisant les techniques du langage naturel.

Compilers and Languages, Escuela Superior de Ingeniería Informática, Orense

Escuela Superior de Ingeniería Informática, Departamento de Informática, Edificio Politécnico, Campus As Lagoas s/n, 32004 Orense (SPAIN)

Tél. : (34) 988 387 280 Téléc. : (34) 988 387 001

Site Web : <http://coleweb.dc.fi.udc.es/>

Personne-ressource : Manuel Vilares Ferro

Courriel : vilares@uvigo.es

Nombre de postes : 20

Projets de recherche

- Qui, quoi, où, combien, comment, pourquoi? Outils de traitement de langues naturelles, apprentissage machine et statistiques bayésiennes pour concevoir des robots d'interrogation de grande précision dans les domaines financiers.
- Extraction d'information économique multilingue (ETIMON)
- Traitement automatique de l'information économique multilingue (TIEMPO) coordonné par
- Recherche d'information pour la réponse aux questions dans les textes économiques (RIBETE)
- Traitement automatique de l'information économique multilingue (TIEMPO)
- Application de l'ingénierie de langage aux systèmes collaboratifs et à la microédition.
- Extraction d'information des nouvelles de la bourse pour évaluer l'attitude du marché.
- Systèmes de recherche d'information sur le langage naturel pour les évaluations cognitives de l'information.
- Groupement de 30 nœuds avec une architecture x86.
- Interrogation de bases de données textuelles structurées.
- Recherche et extraction d'information en appliquant les connaissances linguistiques.
- Analyse automatique des constructions verbales en espagnol (niveau syntaxique)
- Unités de traduction : analyse empirique.
- Conception d'analyseurs tabulaires pour les langages naturels.

Database Lab, University of A Coruña

Laboratorio de Bases de Datos Facultad de Informática Campus de Elviña s/n 15071 A Coruña (España)

Tel: (34) 981 16 70 00 poste 1306 Téléc. : (34) 981 16 71 60

Site Web : <http://cedeira.dc.fi.udc.es/PlainConsulta/MainPage.do?forward=MainPage.do>

Personne-ressource : Nieves Rodríguez Brisaboa

Courriel : brisaboa@udc.es

Nombre de postes : 20

Projets de recherche

Le laboratoire de bases de données (LBD) de l'Université de A Coruña a été créé en 1994. Depuis, il a été présidé par Nieves R. Brisaboa (brisaboa@udc.es). Actuellement, le LBD regroupe des chercheurs et des enseignants du département d'informatique de l'université A Coruña, qui travaillent dans les domaines de recherche suivants :

- Ingénierie logicielle
- Bases de données déductives
- Bibliothèques numériques et fédération des bibliothèques numériques
- Systèmes d'information géographique
- Compression, Indexation et recherche de textes
- Traitement du langage naturel

Department of Artificial Intelligence, Technical University of Madrid

Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, España

Tél. : (34) 91 352 4819 Téléc. : (34) 91 336 7444

Site Web : <http://www.dia.fi.upm.es/indexeng.htm>

Personne-ressource : Ana Garcia Serrano

Courriel : agarcia@dia.fi.upm.es

Nombre de postes : 41

Projets de recherche

Le département d'intelligence artificielle compte plusieurs groupes qui travaillent sur les questions relatives à la technologie du langage :

Grupo de Investigación de Sistemas Inteligentes (ISYS) (Intelligent Systems Research group)

Principaux chercheurs : professeur associé García Serrano, Ana Ma. (agarcia@dia.fi.upm.es), professeur associé Molina González, Martín (mmolina@fi.upm.es)

Membres : 3 maîtres de conférence, 8 adjoints à la recherche, 3 étudiants

Thèmes de recherche : Systèmes axés sur la connaissance (architectures de deuxième génération, applications réelles), intelligence artificielle répartie, télétravail collectif, langage naturel.

Grupo de Investigación en Validación y Aplicaciones Industriales

Principaux chercheurs : professeur associé Cardeñosa Lera, Jesús (carde@fi.upm.es)

Membres : 2 maîtres de conférence, 4 adjoints à la recherche, 5 étudiants

Thèmes de recherche : Intelligence artificielle appliquée, systèmes axés sur la connaissance, processus des systèmes de qualité, Réingénierie (BPR), possibilité de réutilisation, analyse du domaine, télétravail collectif (CSCW), systèmes automatiques de négociation, langage naturel et traduction automatique.

Grupo de Ontologías (Ontologies)

Principaux chercheurs : professeur associé Gómez Pérez, Asunción (asun@dia.fi.upm.es)

Membres : 1 maître de conférence, 3 adjoints à la recherche, 7 étudiants

Thèmes de recherche : Méthodologies de conception des ontologies, évaluation formelle des ontologies, Outils de conception des outils, applications axées sur l'ontologie dans des domaines tels que : le commerce en ligne, gestion des connaissances, information d'intégration intelligente, interrogation intelligente sur Internet.

Laboratorio de Inteligencia Artificial (LIA) (Artificial Intelligence Laboratory)

Principaux chercheurs : professeur titulaire Pazos Sierra, Juan (jpazos@fi.upm.es), professeur associé Gómez Pérez, Asunción (asun@dia.fi.upm.es)

Membres : 4 maîtres de conférence, 3 maîtres de conférence, 4 adjoints à la recherche, 7 étudiants en doctorat, 3 étudiants en maîtrise, étudiants de 30 ou plus

Thèmes de recherche : Systèmes d'expert, langage naturel, validation de systèmes axés sur les connaissances, intelligence artificielle, partage et réutilisation des connaissances, informatique moléculaire.

IXA Group

Ixa Group, 649 Posta kutxa, 20080 Donostia (Basque Country)

Tél. : (34) 943 015 019 Téléc. : (34) 943 015 590

Site Web : <http://ixa.si.ehu.es/Ixa>

Personne-ressource : Eneko Agirre

Courriel : e.agirre@ehu.es

Nombre de postes : 48

Projets de recherche

Le groupe IXA a été créé afin de promouvoir la modernisation de la langue basque en développant les ressources informatiques de base nécessaires. À la suite de nos travaux, quatre applications sont actuellement accessibles aux utilisateurs communs : un correcteur d'orthographe, un inforobot de recherche Web axé sur la lemmatisation, une lemmatisation axée sur le dictionnaire bilingue en ligne et un générateur de bulletins d'information météorologique.

Le développement de la technologie des langues de la minorité varie dans divers aspects de la technologie des langues courantes. La grande puissance des ordinateurs actuels, combinée avec la rareté des ressources humaines et linguistiques, incite à la conception de stratégies nouvelles et différentes.

Les technologies du langage humain apporteront une contribution essentielle au succès de la société de l'information. Toutefois, la plupart des applications de travail ne seront disponibles qu'en anglais. Beaucoup d'efforts sont requis pour résoudre les problèmes que confrontent ceux et celles qui travaillent sur les langues de la minorité.

MEANING : Développer des technologies langagières multilingues sur le Web.

HERMES : nouvelles bases de données : Extraction sémantique et recherche d'information translinguistique.

XUXENG : Conception et application d'un prototype de correcteur d'orthographe de la langue basque.

Laboratory for Research and Development in Artificial Intelligence

LIDIA. Department of Computer Science Faculty of Informatics University of A Coruña, 15071.

A Coruña, Spain

Tél. : (34) 81 16 70 00 Téléc. : (34) 981 16 71 60

Site Web : <http://www.dc.fi.udc.es/lidia/lidia.html>

Personne-ressource : Amporo Alonzo Betanzos

Courriel : ciamparo@udc.es

Nombre de postes : 20

Projets de recherche

- Les projets sont réalisés dans les domaines suivants :
- Systèmes de connaissances et intelligence artificielle;
- Modèles symboliques de traitement, représentation des connaissances et incertitude;
- Exploration de données.

Natural Language Processing and Information Retrieval Group at National Distance Learning University (UNED)

Dpto. De Lenguajes y Sistemas Informaticos, ETSI Informatica, UNED, C/ Juan del Rosal, 16
Ciudad Universitaria, 28040 Madrid

Tél. : (34) 91 398 7983 Téléc. : (34) 91 398 6535

Site Web : <http://nlp.uned.es/index.php>

Personne-ressource : Felisa Verdejo Maillo

Courriel : felisa@lsi.uned.es

Projets de recherche

Accès intelligent à l'information

Accès multilingue à l'information.

- Assistants de recherche des langues étrangères pour la recherche de documents, la réponse aux questions et la synthèse d'information.
- Multilingual website search engines.
- Parcours multilingue de phrases et recherche de terminologie.
- Traitement superficiel et efficace du traitement du langage naturel pour l'accès à l'information.
- Organisation, visualisation et parcours intelligents des résultats de recherche.
- Évaluation de systèmes d'accès à l'information.
- Organismes conjoints du Cross-Language Evaluation Forum (CLEF).
- Organismes conjoints du Interactive Cross-Language Evaluation Forum (iCLEF).
- Organismes conjoints du CLEF Cross-Language Question Answering track.

Natural Language processing Group, Technical university of Catalonia

C/ Jordi Girona, 1-3, edif. Omega, despatx 315, Barcelona, Spain 08034

Tél. : (34) 93 413 7856 Téléc. : (34) 93 413 7833

Site Web : <http://www.lsi.upc.edu/~nlp/>

Personne-ressource : Núria Castell

Courriel : castell@lsi.upc.edu

Nombre de postes : 31

Projets de recherche

CHIL : Computers in the Human Interaction Loop

HOPS : Valider un concentrateur intelligent axé sur le langage naturel pour le déploiement à grande échelle de services publics en ligne avancés et complétés par des technologies sémantiques. Ces services sont également supportés par un système multicanal.

Meaning : Développer des technologies multilingues web.

LC-STAR : Lexique et corpora pour les composants de la traduction vocale.

FAME : Agent de facilitation dans l'échange multiculturelle.

NAMIC : Catégorisation d'information multilingue des agences de presse.

EUROWORDNET : Créer un wordnet multilingue avec des relations sémantiques entre les mots.

OpenTrad : Traducción Automática de Código Abierto para las lenguas del Estado Español

HERMES : Bibliothèques électroniques avec recherche d'information multilingue et traitement sémantique.

PETRA : Interfaces de la parole pour des applications avancées unifiées de messagerie.

BASURDE : Système de dialogue spontané dans les domaines limités.

ITEM : Recherche d'information textuelle dans un environnement multilingue en utilisant des techniques du langage naturel.

RILE : Serveur de ressources linguistiques en espagnol.

Speech Processing Group (VEU), Technical University of Catalonia

Department of Signal Theory and Communications Building D5 - Campus Nord C/Jordi Girona, 1-3 08034 Barcelona

Tél. : (34) 93 401 64 40 Téléc. : (34) 93 401 64 47

Site Web : <http://gps-tsc.upc.es/veu/>

Personne-ressource : José B. Mariño Acebal

Courriel : canton@gps.tsc.upc.edu

Nombre de postes : 28

Projets de recherche

Le Groupe de traitement de la parole (VEU) travaille sur l'ingénierie linguistique et le traitement de la parole. Il a pour objectif principal de concevoir des systèmes de communication orale et leurs divers domaines de recherche : codification de la parole, extraction robuste des caractéristiques de la parole, conversion du texte à la parole, reconnaissance de la parole, traduction vocale et dialogue oral. Le principal intérêt de la recherche consiste à concevoir des systèmes qui franchiront les obstacles linguistiques et amélioreront l'accessibilité des systèmes d'information.

The Research Group on Signals, Telematic and Communications, University of Granada

Dpto. Teoría de la Señal, Telemática y Comunicaciones, ETSI Informática, Universidad de Granada, 18071 – Granada

Tél. : (34) 958 243 283 Téléc. : (34) 958 243 230

Site Web : http://ceres.ugr.es/info_eng/

Personne-ressource : José Carlos Segura Luna

Courriel : segura@ugr.es

Nombre de postes : 15, nombre inconnu d'étudiants assistants

Projets de recherche

Le groupe de recherche sur les signaux, de la télématique et des communications (de Investigación en Señales, Telemática y Comunicaciones, GSTC) compte 14 chercheurs du département de l'électronique et de technologie informatique (DETC) de l'Université de Grenade (UGR).

Les membres du groupe sont bacheliers en physique (spécialité en électronique) ou en informatique. Douze d'entre eux ont un doctorat. Les autres membres sont des étudiants de doctorat qui prennent part au programme de doctorat *Multimedia Technologies*. Ils enseignent également dans les domaines du traitement et de la codification des signaux, les communications et la transmission des données, les réseaux informatiques, la robotique et les communications électroniques.

La recherche actuelle couvre particulièrement les domaines suivants :

- Traitement des signaux;
- Reconnaissance continue de la parole dans une ambiance bruyante;
- Conversion du texte à la parole;
- Compression et codification des signaux.

Organismes et associations

Catalan Association of Artificial Intelligence
<p>Asociació Catalana d'Intel·ligència Artificial Institut d'Investigació en Intel·ligència Artificial (IIIA) Campus UAB 08193 Bellaterra, Catalonia</p> <p>Tél. : (34) 93 580 9570 Téléc. : (34) 93 580 9661</p> <p>Site Web : http://www.acia.org</p> <p>Personne-ressource : Carles Sierra</p> <p>Courriel : sierra@iiia.csic.es</p>
Mission
<p>L'association catalane de l'intelligence artificielle (Associació Catalana d'Intel·ligència Artificial) est une association sans but lucratif destinée à faire avancer l'intelligence artificielle dans la société catalane. L'association a pour principal objectif de favoriser la communication entre les personnes et les organismes qui concernés et de promouvoir la sensibilisation sociale, scientifique, économique et gouvernementale à l'intelligence artificielle.</p>

États-Unis

Laboratoires

Brain and cognitive sciences, Rochester University

Meliora Hall, Box 270268, University of Rochester, Rochester, NY 14627-0268

Tél. : (585) 275 1844 Téléc. : (585) 442 9216

Site Web : <http://www.bcs.rochester.edu/>

Personne-ressource : Kathy Corser

Courriel : kcorser@bcs.rochester.edu

Nombre de postes : 58

Projets de recherche

Une constellation de la faculté et d'étudiants se penchent sur la *nature du traitement et de l'acquisition du langage*. Cette recherche couvre un vaste éventail de thèmes : perception des sons de la parole et des mots, production et compréhension du langage écrit et oral, acquisition des langages parlés par les jeune enfants et les adultes, études de la structure, du traitement et d'acquisition des langages gestuels. Le centre des sciences du langage, une unité de recherche interdisciplinaire associée au département des sciences cognitives et du cerveau, rassemble les chargés de cours et les étudiants de notre faculté ainsi que les scientifiques spécialisés dans le langage des départements de linguistique, d'informatique et de philosophie.

Cambridge research Lab

Hewlett-Packard, HP Labs, one Comabridge Center, Cambridge, Massachusetts 02142-1612

Tél. : (617) 551 7600 Téléc. : (617) 551 7650

Site Web : <http://www.hpl.hp.com/research/crl/>

Personne-ressource : Richard Zippel

Courriel : webmaster@hpl.hp.com

Nombre de postes : 35

Projets de recherche

Le laboratoire de recherche Cambridge (CRL) a été créé en 1987 afin de faire progresser les connaissances en matière de traitement de base et d'interaction homme-machine. Aujourd'hui, notre recherche comprend des activités en saisie, codification, stockage, indexation, recherche, décodage et rendu des données des multimédias.

Indexation multimédia à grande échelle – concevoir des technologies permettant un accès, une navigation et une organisation faciles relativement aux grands dépôts multimédias (audio, images et vidéo). Ces travaux comprennent SpeechBot, qui intègre la puissance de la recherche Web au contenu audio et vidéo.

Center for advanced information processing

CAIP Center at Rutgers University, CORE Building 706, Busch Campus, 96 Frelinghuysen Rd. Piscataway, NJ 08855-1390

Tél. : (732) 445 3443

Site Web : <http://www.caip.rutgers.edu/>

Personne-ressource : David Gaut

Courriel : daut@ece.rutgers.edu

Nombre de postes : 20

Projets de recherche

CAIP Esu un centre e technologie avancée (ATC) qui a été fondé au début de 1985. Il est conjointement financé par la New Jersey Commission on Science and Technology, l'Université Rutgers et des membres du secteur industriel.

Recherches actuelles

- VRLAB : Laboratoires d'interface homme-machine (Burdea);
- Laboratoire de vision computationnelle (Dana);
- Laboratoire de traitement de la parole et du langage (Dusan, Flanagan, Rabiner);
- Mécanique fluide de la production de sons de la parole (Krane, Wei);
- Technologie de fabrication par étapes (Langrana);
- Laboratoire d'ingénierie de systèmes d'avertissement et d'indicateurs (Mammone);
- Laboratoire de système d'information multimédia (Marsic, Flanagan);
- RIUL : Laboratoire de compréhension robuste des images (Meer);
- Laboratoire de biorobotique (Bouzit);
- TASSL : Laboratoire des systèmes logiciels appliqués (Parashar);
- Interaction homme-machine (Tremaine).

Center for automated learning and discovery

Carnegie Mellon University, School of Computer Science, 5000 Forbes Avenue, Wean Hall
4609, Pittsburgh, PA 15213-3891

Tél. : (412) 268 1299 Téléc. : (412) 268 3431

Site Web : <http://www.cald.cs.cmu.edu/about/index.html>

Personne-ressource : Tom Mitchell

Courriel : Tom.Mitchell@cmu.edu

Nombre de postes : 50

Projets de recherche

Le Center for Automated Learning and Discovery (CALD) est un département universitaire de l'école d'informatique de l'Université Carnegie Mellon. Nous mettons l'accent sur la recherche et la formation dans tous les domaines d'apprentissage machine statistique.

L'apprentissage machine est un domaine scientifique qui soulève la question « Comment pouvons-nous programmer des systèmes pour apprendre automatiquement et nous améliorer avec l'expérience. Nous étudions l'apprentissage de plusieurs expériences, entre autres apprendre à prévoir les patients médicaux qui répondront aux traitements, en analysant l'expérience enregistrée dans les bases de données à partir des enregistrements médicaux en ligne. Nous étudions également les robots mobiles qui apprennent à bien naviguer selon l'expérience qu'ils acquièrent des capteurs en errant à travers leur environnement et des outils d'aide à la découverte scientifique qui combinent les premières hypothèses scientifiques avec des nouvelles données expérimentales afin de produire automatiquement des hypothèses scientifiques évoluées qui s'adaptent mieux aux données observées.

Afin de s'attaquer à ces problèmes, nous mettons au point des algorithmes qui découvrent des conjectures générales et des connaissances à partir de données et d'une expérience précises en se basant sur des principes statistiques et computationnels avisés. Nous concevons également des théories de processus d'apprentissage qui caractérisent la nature fondamentale du calcul et de l'expérience suffisante pour un apprentissage efficace des machines et des hommes.

CALD Research CALD vise à concevoir de nouveaux algorithmes statistiques d'apprentissage à usage généra., de nouvelles théories fondamentales d'apprentissage et de nouvelles applications dans les domaines de la robotique, l'imagerie cérébrale et l'interprétation du texte dans le langage naturel.

Human Language Technology Research institute

Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688

Tél. : (972) 883 4654 Téléc. : (972) 883 2349

Site Web : <http://www.hlt.utdallas.edu/page.php?p=index>

Personne-ressource : Sandra Harabagiu

Courriel : sanda@hlt.utdallas.edu

Nombre de postes : 41

Projets de recherche

Le Human Language Technology Research Institute (HLTRI) a pour objectif principal de stimuler et promouvoir la recherche en technologie du langage humain, notamment le traitement du langage naturel et la synthèse ainsi que la reconnaissance automatique de la parole; et à accroître la visibilité et la réputation de l'Université de Texas de Dallas en tant que lieu d'excellence pour la recherche en technologie du langage humain. Nous visons à mettre sur pied un groupe de recherche de première classe en nous attaquant aux problèmes de recherche les plus importants en matière de technologie du langage humain. Le HLTRI fait partie du Erik Jonsson School of Engineering and Computer Science. Il fournit un environnement riche qui permet aux étudiants de troisième cycle d'apprendre et de travailler avec le personnel de la faculté et le personnel technique sur des recherches pour le gouvernement et les entreprises de financement. L'institut compte cinq centres de recherche spécialisés dans la technologie du langage.

Le Center for Basic Research in Natural Language Processing a pour objectif de concevoir des outils et de développer la recherche de base en lexico-sémantique, en traitement syntaxique et sémantique des textes et des dialogues, en traitement des discours et en mise au point de modèles cognitifs de traitement du langage.

Le Center for Emerging Natural Language Applications a pour objectif d'élaborer des techniques et d'étudier les nouveaux domaines des applications du langage naturel (p. ex. réponse aux questions textuelles, résumé multi-documents, exploration de textes et extraction d'information dans un domaine ouvert.

Le Center for Search Engines and Web Technologies a pour objectif de concevoir et d'appliquer des méthodes pour localiser l'information sur le WWW.

Le Center for Text Mining prend part à l'analyse de sources de textes humains naturels telles que les articles de journaux, les pages Web, les articles usenet et d'autres publications. Il a pour objectif de créer de meilleures bases de connaissances automatiques à partir de textes humains.

Le InterVoice Center est le fruit d'un don généreux de InterVoice, Inc. Le centre met l'accent sur la recherche en matière d'intégration de connaissances syntaxiques et sémantiques avec les techniques actuelles de reconnaissance automatique de la parole.

Center for Human-Computer Communication

CHCC, OGI School of Science & Engineering, 20000 NW Walker Rd., Beaverton, OR 97006

Tél. : (503) 748 1248 Téléc. : (503) 748 1875

Site Web : <http://www.cse.ogi.edu/CHCC/>

Personne-ressource : Thad Davidson

Courriel : davisth@cse.ogi.edu

Nombre de postes : 22

Projets de recherche

- Conception d'interfaces mobiles de la prochaine génération pour la parole conversationnelle dynamique.
- Nouvelles orientations dans le traitement adaptatif de systèmes multimodaux.
- Interaction multimodale pour les environnements mobiles et sur le terrain.
- Interaction multimodale pour le Wearable Augmented Reality.
- Interaction multimodale pour les environnements virtuels.
- Systèmes robustes en mode agent qui intègrent des équipes d'agents de communication (AAA, Agent-Talk, STAPLE).
- Interfaces interactives multimodales : Concevoir pour la performance humaine multimodale.

Center for intelligent information retrieval

CIIR, Computer Science Department, University of Massachusetts, 140 Governors Drive, Amherst, MA 01003-9264

Tél. : (413) 545 0463 Téléc. : (413) 545 1789

Site Web : <http://ciir.cs.umass.edu/>

Personne-ressource : W. Bruce Croft

Courriel : croft@cs.umass.edu

Projets de recherche

Le Center for Intelligent Information Retrieval (CIIR) est un centre national des sciences qui fait partie des plus grands laboratoires de recherche d'informations dans le monde. Le centre conçoit des outils qui offre un accès efficace et efficace à des grandes bases de données de textes et de multimédia hétérogènes distribuées. Parmi ses réalisations, mentionnons d'importants progrès en recherche dans les domaines de la recherche d'informations, le filtrage d'informations, l'identification de sujets, la recherche et l'indexation multimédia, le traitement d'images documentaires, collections en téraoctets, exploration de données, résumage, exploration des ressources, interfaces et visualisation et recherche d'informations translinguistiques. Le Center for Intelligent Information Retrieval continue d'appuyer la nouvelle infrastructure par la recherche et le transfert de la technologie. Le centre a pour but de concevoir des outils qui fournissent un accès efficace et efficace à des grandes bases de données de textes et de multimédia hétérogènes distribuées.

AQUAINT Project : Ce projet de recherche, Relevance Models and Answer Granularity for Question Answering, est une initiative d'ARDA dans le cadre du programme Advanced Question and Answering for Intelligence (AQUAINT).

NSF Digital Government Project : Ce projet de recherche, A Language-Modeling Approach to Metadata for Cross-Database Linkage and Search, est une initiative financée par la Fondation nationale des sciences. Le CIR travaille en collaboration avec l'Université Carnegie Mellon, la bibliothèque du Congrès, le ministère du Commerce, le Geological Survey et R.S.S.C.

NSF IDM Mongrel Project: « Supporting Effective Access through User- and Topic-Based Language Models » est un projet de recherche conjoint avec l'Université Rutgers. Il est financé par le programme NSF IDM.

NSF NSDL - Search and Browsing Support for NSDL Le CIR a travaillé sur On an NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) avec un groupe d'institutions qui sont en train de développer les capacités techniques et d'exécuter les responsabilités organisationnelles d'intégration de base du programme NSDL.

Online Music Retrieval and Search Project Le projet de recherche d'OMRAS est un programme subventionné par Digital Libraries II dans le cadre de l'initiative internationale de bibliothèques numériques. Le CIR travaille sur ce projet conjointement avec le Kings College of London et l'Université d'Indiana.

Center for language and speech processing

Barton Hall, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218

Tél. : (410) 516 4237 Téléc. : (410) 516 5050

Site Web : <http://www.clsp.jhu.edu/index.shtml>

Personne-ressource : Frederick Jelinek

Courriel : jelinek@jhu.edu

Nombre de postes : 34

Projets de recherche

Le centre a pour objectif de promouvoir la recherche et la formation dans la science et la technologie du langage et de la parole. La recherche est menée par des universitaires, des chercheurs scientifiques et des étudiants de troisième cycle de six départements universitaires associés : génie biomédical, sciences cognitives, informatique, génie électrique et informatique, mathématiques et psychologie. La recherche inclut des travaux sur tous les aspects de la science et la technologie du langage et de la parole. Il inclut également des études fondamentales en cours dans des domaines tels que la modélisation du langage, le traitement du langage naturel, le traitement neuronal des informations auditives, le traitement acoustique, la théorie d'optimalité et l'acquisition du langage.

Projets

- PIRE : Étude des représentations du sens dans la compréhension du langage naturel pour les systèmes de traduction automatique, Frederick Jelinek, Eugene Charniak, Jason Eisner, Mark Johnson et Sanjeev Khudanpur.
- MALACH : Accès multilingue à de grandes archives vocales William Byrne et Frederick Jelinek.
- Groupement de mots et modèles de langage type, Ahmad Emami et Frederick Jelinek.
- Groupement de mots par l'analyse sémantique de structure latente, Yan Huang et Sanjeev Khudanpur.
- Exploitation de relations syntaxiques non locales dans les modèles de langage pour la reconnaissance de la parole conversationnelle (STIMULATE)? Peng Xu et Frederick Jelinek.
- Modélisation acoustique pour la parole conversationnelle, Shankar Kumar, Asela Gunawardana, Vaibhava Goel, Veera Venkataramani, William Byrne.
- Groupement acoustique à risque minimal pour la combinaison type multilingue acoustique et estimation transinformation maximale des HMM, Stavros Tsakalidis et William Byrne.
- Modélisation de la prononciation de la parole causale du mandarin, Veera Venkataramani et William Byrne.

Center for machine translation

LTI Research, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

Tél. : (412) 268 5480 Téléc. : (412) 268 5578

Site Web : <http://www.lti.cs.cmu.edu/Research/CMT-home.html>

Personne-ressource : Céline Carraux

Courriel : celine@cs.cmu.edu

Nombre de postes : 20

Projets de recherche

Le LTI compte actuellement de nombreux projets actifs dans les domaines de la traduction automatique, la parole, la recherche d'information, l'acquisition de connaissances, les applications pédagogiques et d'autres projets connexes en SCS/CMU.

- Babylon : Traduction vocale bilatérale sur un ordinateur portable.
- JANUS : Traduction vocale de dialogues conversationnels spontanés dans plusieurs langues en utilisant principalement une méthode interlingue.
- KANT : Le projet de traduction automatique basée sur les connaissances; traduction exacte, basée sur les connaissances, de documents techniques.
- LingWear vise à concevoir un support vestimentaire.
- RADD MT : Traduction automatique basée sur des données et adaptable rapidement.
- CAMMIA : Système à fonctionnement multitâche de gestion du dialogue qui peut être utilisé avec la reconnaissance automatique de la parole et VoiceXml pour fournir un accès mobile à l'information.
- Correction des accents dans les langues étrangères.
- FestVox : Créer des voix synthétiques.
- LET'S GO! : Un système de dialogue vocal pour le public général.
- Towards Communicating with Dolphins : Ce projet applique les aspects de la technologie de la parole et de l'apprentissage automatique pour améliorer la communication avec les dauphins.
- SUBLIME : Environnement de gestion d'information basée sur la parole et le langage.
- Speech Graffiti : Interfaces universelles de la parole.
- Adaptive Information Filtering : Suivi automatique d'un ensemble de documents (p. ex. reportages, forums) pour trouver les histoires qui vous intéressent.

Center for spoken language research

CSLR, university of Colorado, 1777 Exposition Drive, Boulder, Colorado 80301

Tél. : (303) 735 5150 Téléc. : (303) 735 5072

Site Web : <http://cslr.colorado.edu/index.html>

Personne-ressource : Catherine Boyer

Courriel : boyer@cslr.colorado.edu

Nombre de postes : 23

Projets de recherche

Diverse Suite of NLP Demos : Dans le cadre de projets de recherche financés par ARDA et NSF, le CSLR a mis au point plusieurs outils de traitement du langage naturel dont ceux qui s'inscrivent sous le projet « Diverse Suite of NLP Demos ». L'objectif étant d'illustrer le fonctionnement des diverses technologies de traitement du langage naturel.

SONIC : Large Vocabulary Speech Recognition System SONIC est un grand système de reconnaissance de la parole disposant d'un très vaste vocabulaire. Ce système est utilisé dans plusieurs projets de recherche du CSLR. Le logiciel est accessible à des fins non commerciales. Les exécutable sont disponibles pour Linux (kernel 2.2/2.4), MS Windows, Sun Solaris et Mac OS X.

NSF-ITR Reading Tutor Project Le projet NSF-ITR Reading Tutor Project est un effort de recherche conjoint entre plusieurs universités afin de créer des caractères informatiques conversationnels intelligents et animés capables de tenir une conversation naturelle face à face dans certains domaines. La mise au point de ces agents nécessite une recherche de base pour créer et améliorer les technologies dans les domaines de la reconnaissance de la parole sur ordinateur, la compréhension du langage naturel, la modélisation de dialogues, la reconnaissance visuelle des mouvements du visage, les expressions et les gestes et la production d'une parole naturelle et expressive sur le plan auditif et visuel grâce à des caractères animés tridimensionnels.

CU Communicator Project : Le communicateur CU est un système de dialogue conversationnel qui interagit avec les utilisateurs par téléphone afin de fournir des renseignements sur les vols et les réservations d'hôtel et de voiture. Ce système sert de test d'évaluation des performances pour notre recherche afin d'améliorer l'état actuel des connaissances relatives aux systèmes de dialogue vocal. Le projet est financé par la Defense Advanced Research Project Agency (DARPA).

Center for Study of Language and Information

Stanford University, CSLI, Ventura Hall, 220 Panama Street, Stanford, CA, 94305-4101

Tél. : (650) 723 0758 Téléc. : (650) 723 0758

Site Web : <http://www-csli.stanford.edu/>

Personne-ressource : Byron Reeves

Courriel : reeves@stanford.edu

Nombre de postes : 72

Projets de recherche

CSLI est un laboratoire interinstitutionnel qui a été fondé en 1983 par des chercheurs de Stanford, SRI International et Xerox PARC. Depuis, le CSLI n'a cessé de promouvoir la collaboration entre les laboratoires industriels et les départements universitaires. Au cours des dernières années, cette collaboration s'est étendue pour inclure des chercheurs d'autres universités, laboratoires et compagnies à l'échelle nationale et internationale. Le Industrial Affiliates Program du CSLI comprend vingt associations membres dont plusieurs envoient leurs chercheurs pour participer à des projets de recherche sur le terrain.

Interface Laboratory : Ces projets sont le fruit d'un effort concerté du CSLI pour résoudre les problèmes d'interface homme-machine dans le monde des télécommunications, du traitement de l'information et de l'électronique grand public.

Research in the Cognitive Sciences : Ces projets interdisciplinaires couvrent un large éventail de thèmes relatifs à l'informatique, la linguistique, la logique et la sémantique, la philosophie, la psychologie et la formation.

Center of Excellence for Document Analysis and Recognition

CEDAR, State University of New York at Buffalo, 520 Lee Entrance, Suite 202, Amherst, NY 14228

Tél. : (716) 645 6162 Téléc. : (716) 645 6176

Site Web : <http://www.cedar.buffalo.edu/>

Personne-ressource : Sargur N. Srihari

Courriel : srihari@cedar.buffalo.edu

Nombre de postes : 55

Projets de recherche

Les travaux couvrent les domaines de la reconnaissance de formes, l'apprentissage automatique, l'exploration de données, la recherche d'information et la linguistique computationnelle dans le cadre de la recherche en informatique et la formation. Il existe un continuum de l'analyse des images documentaires balayées à des domaines tels que l'exploration de textes et la recherche d'information. Il existe également un continuum de l'analyse des documents à des fins judiciaires à l'analyse d'empreintes, en passant par d'autres domaines de la biométrie.

La recherche en matière de reconnaissance et d'analyse de documents est menée à l'Université de Buffalo depuis 1978. Le programme a commencé par une subvention de recherche attribuée en 1980 à Sargur Srihari de la Fondation nationale de sciences dans le cadre de la recherche intitulée « Contextual Algorithms for Text Recognition ». Cette subvention a abouti en 1991 à CEDAR. Parmi les domaines de recherche, mentionnons l'apprentissage automatique, la recherche d'information, l'extraction d'information, la reconnaissance optique de l'écriture manuelle, identification des empreintes, exploration de données et plus.

Computational Linguistics and Information Processing Laboratory

CLIP, University of Maryland, UMIACS, 3357 AV Williams Building, College Park, MD 20742

Tél. : (301) 405 6444 Téléc. : (301) 314 2644

Site Web : <http://www.umiacs.umd.edu/research/CLIP/clip.htm>

Personne-ressource : Bonnie J. Dorr

Courriel : bonnie@cs.umd.edu

Nombre de postes : 32

Projets de recherche

CLIP se compose de deux entités : le groupe des langages naturels et le groupe des bases de données. Le groupe des langages naturels met l'accent sur plusieurs domaines du traitement multilingue à grande échelle (p. ex., la traduction automatique, la détection extensible translingue de documents et la recherche d'information interlingue. Le groupe des bases de données met l'accent sur les architectures pour le traitement étendu avec des serveurs d'information hétérogènes (p. ex., découverte scientifique à partir de sources de données biomoléculaires).

- Méthodes computationnelles en linguistique
- Moteur de recherche de psycholinguistique computationnelle pour modéliser le changement historique.
- Technologie du langage humain.
- Divergence Unraveling (DUSTer) GALE Generation Heavy Hybrid MT (GHMT) Headline Generation Interlingual Annotation MALACH Statistical MT TIDES.
- Traitement d'information, optimisation des interrogations et sémantique Web.
- Projets connexes dans les laboratoires où travaillent les chercheurs en partenariat.
- Métadonnées de la qualité des données des dialogues homme-machine pour le Web sémantique Websources.

Computational Linguistics at Stanford University

CLSU, Linguistics Dept, Bldg 460, 450 Serra Mall, Stanford University, Stanford, CA 94305-2150.

Tél. : (650) 723 0924 Téléc. : (650) 723 5666

Site Web : <http://www.stanford.edu/dept/linguistics/compling/>

Personne-ressource : Dan Jurafsky

Courriel : jurafsky@stanford.edu

Nombre de postes : 21

Projets de recherche

Le programme de linguistique computationnelle de Stanford est l'un des plus vieux programmes au pays. Il considère la linguistique computationnelle sous un angle très large qui couvre les études théoriques et appliquées ainsi que divers domaines de linguistique, de la phonologie computationnelle à la sémantique et la pragmatique computationnelles, en passant par la syntaxe et la morphologie.

Parmi les priorités théoriques, mentionnons les modèles de parsing et d'apprentissage de la structure grammaticale, les modèles de communication, la conversation et le dialogue ainsi que les modèles psycholinguistiques computationnels de la compréhension et la production des langages humains. CLSU travaille actuellement sur les principales applications dans la technologie du langage humain telles que les systèmes de dialogue multipartites, la traduction automatique, la compréhension des phrases, le parsing et le tagging probabilistiques, la reconnaissance des dialogues, la synthèse de la parole, l'extraction d'information biomédicale, induction de grammaire, désambiguïsation du sens des mots et réponse automatique aux questions.

Bien que le département compte divers projets et de recherche et équipes qui étudient différents thèmes, il partage l'objectif ambigu de faire face à la complexité et l'incertitude du langage humain en intégrant les riches modèles de la structure linguistique à chaque niveau linguistique avec des techniques probabilistiques et statistiques modernes et sophistiquées.

Conjointement avec le département d'informatique, le département compte divers laboratoires de recherche, groupes de lecture et ateliers de travail informels sur la linguistique computationnelle. Il entretient également des relations étroites avec plusieurs compagnies qui oeuvrent dans les applications de linguistique computationnelle dans la région de la Silicon Valley.

Columbia Natural Language Processing Group

450 Computer Science Building, 1214 Amsterdam Avenue, 0401 / New York, New York
10027-7003

Tél. : (212) 939 7000 Téléc. : (212) 666 0140

Site Web : <http://www1.cs.columbia.edu/nlp/index.htm>

Personne-ressource : Sandra Morris

Courriel : swm16@columbia.edu

Nombre de postes : 48

Projets de recherche

AQUAINT est un projet de réponse aux questions qui met l'accent sur la réponse à des questions complexes.

CLiMB (Computational Linguistics for Metadata Building) a pour objectif d'explorer et de développer les stratégies assistées par ordinateur pour extraire des métadonnées à la pièce (c.-à-d. catalogage qui peut être utilisé dans les grands projets de collections de librairie numérique) du corps de la littérature érudite existante monographique.

MAGIC est un système intelligent de présentation multimédia pour le domaine médical. Après qu'un patient ait subi une chirurgie cardiaque, le médecin du bloc opératoire doit informer les soignants dans l'unité des soins intensifs (USI) de ce qui s'est passé durant la chirurgie pour préparer le patient lorsqu'il arrive à l'USI. MAGIC remplace les médecins du bloc opératoire dans ce scénario, puisqu'il présente une information similaire au moyen de parole et de graphiques coordonnés.

Newsblaster est un système qui aide les utilisateurs à trouver les nouvelles qui les intéressent. Chaque jour, il recueille, rassemble, catégorise et résume automatiquement les nouvelles de plusieurs sites sur le Web (CNN, Reuters, Fox News, etc.) et propose aux utilisateurs une interface conviviale pour parcourir les résultats.

PERSIVAL (PErsonalized Retrieval and Summarization of Image, Video And Language resources) vise à fournir un accès personnalisé à une bibliothèque numérique distribuée des soins aux patients. PERSIVAL est une initiative de recherche conjointe entre les domaines de NLP, l'interaction machine-homme, l'informatique médicale, le traitement vidéo, la bibliothéconomie et la science cognitive.

TIDES est un projet de détection, d'extraction et de résumé d'information translingue de DARPA. Il consiste en un système pratique de suivi et de résumé d'information multidocuments. Le système est destiné à être transférable en matière de langues. Nous sommes d'ailleurs en voie d'étendre le système pour inclure l'arabe, le japonais et le chinois. La technologie conçue en TIDES est utilisée tous les jours dans Newsblaster.

Computing research laboratory

Science Hall Building, University of New Mexico, NM 88003-8001

Tél. : (505) 646 5466 Téléc. : (505) 646 6218

Site Web : <http://crl.nmsu.edu/index.html>

Personne-ressource : James R. Cowie

Courriel : jcowie@crl.nmsu.edu

Nombre de postes : 24

Projets de recherche

AWARE - Discovering characteristics of habitable answering systems Le projet vise à étudier la meilleure façon pour que les systèmes de réponse présentent les résultats de façon à ce que l'utilisateur sache quelle est la question à laquelle le système a répondu et comment a-t-il procédé pour y arriver. Nous étudions dans le cadre de ce projet la façon dont les utilisateurs interagissent avec le système de réponses MIT's START et les systèmes de réponse Language Computer Corporation's Internet Answer. Nous examinons en outre comment les personnes utilisent les expressions de référence lorsqu'ils interagissent avec des systèmes de langage naturel.

Computer-Mediated Multilingual Collaboration Ce projet comprend l'évaluation d'un système multilingue de conversation en ligne à deux dimensions : le degré d'efficacité de l'interaction utilisateur-système pour accomplir les tâches; la qualité des agents de traduction produits dans les conversations en ligne multilingues. Pour tester le dernier cas, nous avons étudié l'efficacité du système relativement aux langues anglaise, chinoise, japonaise, espagnole et coréenne.

Gene Literature Exploration Environment Mené en collaboration avec le Cancer Research and Treatment Center de l'Université de New Mexico, ce projet vise à accélérer le processus de recherche des volumes de données pour contribuer à l'analyse de la nouvelle recherche en génétique.

Interlingual Annotation of Multilingual Text Corpora CRL a eu un projet NSF en collaboration avec UMIACS (Université de Maryland, College Park), ISI (Université de Southern California), LTI (Université de Carnegie-Mellon), Université de Columbia Mitre Corporation.

Meaning Oriented Question Answering Il s'agit d'un projet de CRL conjointement avec CoGenTex de Ithaca, NY et ILIT (Institute for Language and Information Technologies) de UMBC (Université de Maryland, Baltimore County). Le projet vise à appliquer l'ontologie ONTOSEM de CRL et la représentation du sens du texte (TMR) pour améliorer la précision et la convivialité des systèmes de réponse aux questions.

Biography Generator Ce projet consiste à élaborer un système multilingue pour générer des biographies personnelles. Le système est actuellement testé en anglais, en espagnol et en russe.

Department of Cognitive and Linguistic Sciences, Brown University

Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912-1978

Tél. : (401) 863 2616 Téléc. : (401) 863 2255

Site Web : <http://www.cog.brown.edu/>

Courriel : coglingInfo@brown.edu.

Nombre de postes : 78

Projets de recherche

La faculté compte des universitaires de différentes disciplines en psychologie cognitive, linguistique théorique et computationnelle, psycholinguistique et neuroscience computationnelle. Elle travaille avec les départements des mathématiques, d'informatique, de neurosciences, de philosophie et de psychologie. Cette configuration de la faculté offre des possibilités incroyables pour l'étude au sein et entre les disciplines. Il est par exemple possible de combiner la formation formelle dans la linguistique théorique avec la recherche expérimentale en psycholinguistique, neurolinguistique ou acquisition du langage. Les étudiants appliquent les techniques de modélisation neuronale sur réseau à des résultats empiriques en perception, cognition, langage et parole. Le groupe de vision dispense une formation relative aux approches informatiques, psychophysique et écologiques à la perception et à l'action, une combinaison qui ne se trouve nulle part ailleurs. Le groupe de la parole couvre les perspectives expérimentales, développementales, neurolinguistiques et évolutives dans le domaine de la perception et la production de la parole.

Principaux domaines de recherche

- Acquisition du langage.
- Linguistique computationnelle (traitement du langage naturel).
- Neurolinguistique.
- Perception et production de la parole.
- Syntaxe et sémantique.

Haskins Laboratories

Haskins Laboratories, Yale University, 300 George Street, suite 900, New haven, CT 06511

Tél. : (203) 865 6163 Téléc. : (203) 865 8963

Site Web : <http://www.haskins.yale.edu/index.html>

Personne-ressource : Carol A. Fowler

Courriel : carol.fowler@haskins.yale.edu

Nombre de postes : 122

Projets de recherche

La recherche menée dans les laboratoires Haskins contribue à l'avancement de la technologie de la parole. Les projets de recherche actuels effectués dans les laboratoires sont interreliés et combinés pour former un programme de recherche exhaustif avec un seul but : la communication humaine par la parole et la lecture.

Nature and Acquisition of the Speech Code and Reading. Ce programme a pour principal objectif de comprendre comment le régime linguistique, spécialisé biologiquement pour la parole et l'écoute, s'adapte à la lecture et l'écriture.

Sources of Sound in Speech. Explore la coordination cinématique de l'articulation et la régulation de la pression d'air dans la production de la parole.

Perception and Production of Rhythmic Microstructure. Met l'accent sur la synchronisation sensométrique et la production de rythmes.

Links Between Production and Perception in Speech. Examine le lien entre la production et la perception de la parole en utilisant des données d'articulation, la modélisation et la synthèse.

Task Dynamics: Goals and Time in Speech Production. Développe le compte fourni par le modèle dynamique des aspects temporels de la production de la parole.

Imitation: A Tool for Studying Speech Perception. Utilise la tendance démontrée des adultes à imiter la parole qu'ils entendent en tant qu'outil pour explorer la nature de la perception de la parole.

Dynamic Patterns of Speech Aerodynamics in Children. Explore le développement du contrôle laryngé chez les enfants de 4 à 10 ans.

Human-Computer Interaction Institute

Human-Computer Interaction Institute (HCII) School of Computer Science Carnegie Mellon University 5000 Forbes Avenue Pittsburgh, PA 15213-3891

Tél. : (412) 268-3830 Téléc. : (412) 268-1266

Site Web : <http://www.hcii.cmu.edu/index.html>

Personne-ressource : Anne Watzman

Courriel : aw16+@andrew.cmu.edu

Nombre de postes : 148

Projets de recherche

Carnegie Mellon est spécialisé en interaction homme-machine, notamment les outils logiciels interface utilisateur, les modèles cognitifs, la reconnaissance de la parole, la compréhension du langage naturel, le graphique informatique, la reconnaissance des gestes, la visualisation des données, les agents intelligents, la conception d'interface visuelle, les multimédias, les travaux conjoints assistés par ordinateur, la musique et la drama informatiques, les tuteurs intelligents, la rédaction technique et les conséquences organisationnelles et sociales sur la technologie.

Le Human-Computer Interaction Institute continue de favoriser et d'effectuer la recherche et la formation multidisciplinaire en interaction homme-machine. La portée de ses travaux comprend la conception et le test de nouveaux outils et de nouvelles technologies pour appuyer l'organisation et l'activité humaine et développer la théorie dans le domaine. La recherche comprend des études empiriques et analytiques du comportement ainsi que des études du comportement chez les groupes et les organisations afin d'étayer la conception et l'évaluation de nouvelles technologies. Les étudiants des départements d'appui trouvent dans le HCII une source riche de possibilités. L'institut propose également un cadre d'alliances avec l'industrie dans les domaines de la recherche et du développement.

USC Natural Language Group

USC information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina Del Rey, CA 90292

Tél. : (310) 448-8731 Téléc. : (310) 823-6714

Site Web : <http://www.isi.edu/natural-language/nlp-at-isi.html>

Personne-ressource : Eduard Hovy

Courriel : hovy@isi.edu

Nombre de postes : 30

Projets de recherche

Adgen : Génération avancée pour répondre aux questions.

AGILE-MT : Traduction automatique basée sur un algorithme statistique.

AQ2-Time : Algorithmes d'orientation temporelle pour le texte du langage naturel.

AQ2-Wordnet : Développement de WordNet pour répondre aux questions.

EDC : Collecte de donnée sur l'énergie.

eRule : Technologie de traitement de l'information pour l'établissement électronique de règles.

eRulemaking : Technologie de traitement du langage pour l'établissement électronique de règles.

GRAMMAR : Modélisation avancée du langage pour la traduction automatique.

GrowOnto : Recherche d'information dans le gouvernement à l'ère du numérique : ontologies et schémas heuristiques autoextensibles.

ICT-Chat : À venir.

IL-Annot : Annotation interlingue.

ISI-NL : Production et analyse sémantique pour des agents autonomes.

ITR-LM : Technologie de traduction pour la modélisation multilingue.

Nitrogen : Production hybride de langage naturel statistique et symbolique.

OmegaOntology : À venir.....

PDP : Une approche maximale à base d'entropie au passage du discours.

International Computer Science institute, Speech Group

International Computer Science Institute 1947 Center Street. Suite 600 Berkeley, CA 94704

Tél. : (510) 666-2900 Téléc. : (510) 666-2956

Site Web : <http://www.icsi.berkeley.edu/about/index.html>

Personne-ressource : Nelson Morgan

Courriel : isci@berkeley.edu

Nombre de postes : 46

Projets de recherche

EARS Rich Transcription of Conversational Speech Le groupe de la parole travaille conjointement avec SRI et l'Université de Washington pour générer des transcriptions lisibles de la parole conversationnelle dans plusieurs langues en intégrant les majuscules, la ponctuation et les marqueurs de paroles.

EARS Novel Approaches Les chercheurs d'ICSI étudient les remplacements de l'enveloppe spectrale standard en tant que représentation de choix de la parole (d'ordinaire avec la transformation cepstrale).

Speech Processing for Meetings Les chercheurs d'ICSI tentent de concevoir des algorithmes et des systèmes de reconnaissance de la parole à partir des réunions ainsi que les méthodes de recherche d'information et d'autres applications qu'une telle reconnaissance permettrait.

Speaker Identification Ce projet porte sur la recherche de caractéristiques vocales pour la reconnaissance du locuteur les technologies connexes de la parole. Il vise ainsi à dépasser les caractéristiques spectrales à court terme qui sont d'habitude à niveau bas et qui dominent les systèmes actuels de reconnaissance du locuteur pour mettre l'accent sur des sources à plus haut niveau d'information du locuteur, y compris la prononciation et l'utilisation idiosyncrasique des mots, les modes prosodiques et les gestes vocaux.

Clarissa Clarissa est un navigateur vocal à part entière qui permet aux astronautes de mieux se servir de leurs mains et de leurs yeux dans les tâches qu'ils effectuent lorsqu'ils naviguent au moyen de commandes vocales.

Regulus Regulus est une trousse de source ouverte pour construire une grammaire à commande vocale. Le développement de la grammaire est effectué principalement en utilisant des méthodes à base d'exemples qui simplifient beaucoup le processus.

Speech Technology for Developing Countries Les chercheurs de l'ICSI mettent au point des technologies de reconnaissance de la parole pour les « nouvelles régions ». Dans le cadre de cet effort, ils ont mis au point de simples agents de reconnaissance pour le tamil, une langue parlée par plus de 50 millions de personnes au sud-est de l'Inde. Le système est un exemple de la capacité de l'ICSI à concevoir et déployer rapidement des prototypes de paroles à peu de frais au moyen d'une technologie de source ouverte.

SmartWeb Sous l'égide de l'institut allemand d'intelligence artificielle (DFKI), le projet SmartWeb porte sur l'accès à des services Web sémantiques sur les appareils portatifs comme les téléphones portables. L'entrée et la sortie de la parole sont bien adaptées aux appareils portatifs et constitueront le point de mire de SmartWeb. Le projet a pour principal objectif de permettre à un utilisateur de poser une question au moyen d'un appareil portatif et

de recevoir directement la réponse à partir d'une information qui vient du Web.

Linguistic Grammars Online

Stanford University, CSLI, Ventura Hall, 220 Panama Street, Stanford, CA, 94305-4101

Tél. : (650) 723-0758 Téléc. : (650) 723-0758

Site Web : <http://www-csli.stanford.edu/>

Personne-ressource : Byron Reeves

Courriel : reeves@stanford.edu

Nombre de postes : 32

Projets de recherche

Le laboratoire CSLI LinGO a pour mission d'élaborer des grammaires linguistiques précises selon le cadre HPSG et des outils polyvalents pour l'ingénierie de la grammaire, le profilage, le passage et la production. Les travaux préliminaires du laboratoire portaient sur la construction d'une grammaire générale de l'anglais (English Resource Grammar ou ERG) et le développement du système d'ingénierie de grammaire LKB. Le CSLI a également utilisé LKB en tant que banc d'essai pour plusieurs grammaires d'enseignements et plus petite envergure pour d'autres langues, y compris le japonais et l'espagnol.

L'ERG est le pilier de la recherche actuelle sur les expressions multimot (MWEs). Elle est financée conjointement par les laboratoires des sciences de la communication NSF et la NTT, Ippon Telegraph et Telephone Corporation. L'ERG a été également déployée dans la construction d'une nouvelle treebank *riche* et *dynamique* pour le HPSG, initiative de LinGO Redwoods. Nous rassemblons notre expérience de rédaction grammaticale à l'échelle de diverses langues afin de trouver une matrice de grammaire qui permettra d'obtenir des grammaires précises d'envergure pour les langages naturels.

Le laboratoire CSLI LinGO est l'un des membres fondateurs de l'international Deep Linguistic Processing with HPSG Initiative (DELPH-IN), un réseau de chercheurs universitaires et d'institutions de recherche commerciale qui ciblent des applications de traitement précis du langage naturel. HPSG et la sémantique de récursion minimale (MRS) fournissent le cadre scientifique commun aux membres de la DELPH-IN. Bon nombre des ressources LinGO font partie du dépôt à source ouverte DELPH-IN.

Au passé, l'ERG servait au système de traduction automatique Vermobil et à un projet financé par la NSF relativement à la production de la parole assistée par ordinateur pour les gens qui ne peuvent pas communiquer en raison d'une invalidité.



Maryland Information and Network Dynamics Lab (MIND)

Maryland information and network Dynamics, 8400 Baltimore Avenue, College Park, Maryland 20740

Tél. : (301) 405-2525 Téléc. : (301) 405-7662

Site Web : <http://www.mindlab.umd.edu/index.shtml>

Personne-ressource : Ashok K. Agrawala

Courriel : agrawala@cs.umd.edu

Nombre de postes : 31

Projets de recherche

L'Université de Maryland a créé le Maryland Information and Network Dynamics (MIND) Lab, un laboratoire qui vise à définir, à mettre au point, à évaluer et à déployer de nouvelles technologies d'information.

Le laboratoire MIND sert d'appui aux efforts de collaboration et à l'expérimentation opérationnelle sur les systèmes de taille raisonnable en déployant l'infrastructure de recherche et les bancs d'essai qui peuvent servir à résoudre les problèmes relatifs aux systèmes échelonnables distribués et hétérogènes.

Le laboratoire MIND est fier d'avoir le Semantic Web Agents Group (MINDSWAP). Ce groupe est présidé par James Hendler. Il a été formé en 2002 pour étudier les technologies de pointe pour la prochaine génération du WWW. Une partie de ses travaux comprend la création et la visualisation d'ontologie, le balisage des photos, le raisonnement et les régimes de confiance distribués.

MIT Media Laboratory

The Media Laboratory, Building E15, 77 Massachusetts Avenue, Cambridge, MA 02139-4307

Tél. : (617) 253-0365

Site Web : <http://www.media.mit.edu/>

Personne-ressource : Alexandra Kahn

Nombre de postes : 140

Projets de recherche

Le laboratoire Média vise à mettre en service une technologie d'apprentissage et d'expression par les gens et les machines. Il met l'accent sur les technologies qui améliorent la qualité de vie de l'ère numérique et aident les gens à construire leurs propres outils d'expression. Le laboratoire préconise un processus qui fait appel à l'imagination et la réalisation, la critique et le réflexe.

La recherche réalisée au laboratoire Média comprend des développements interconnectés dans plusieurs disciplines inhabituelles. Elle explore actuellement les nouvelles frontières, telles que le sans fil, les communications « virales », le prêt-à-porter connecté, des machines capables de raisonner, les nouvelles formes d'expression artistique et la façon d'apprentissage des enfants. Ces thèmes soulignent un avenir où des bits de partition numérique interagissent harmonieusement avec les atomes de notre monde physique et un avenir où nos machines non seulement répondent à nos commandes, mais elles comprennent également nos émotions, un avenir où l'innovation numérique devient le domaine de tout le monde.

Research Cluster on Language, Center for Cognitive Science, Rutgers State University

Rutgers Center for Cognitive Science (RuCCS), Center for Cognitive Science & Lab of Vision Research, Psych Bldg Addition, Busch Campus, Rutgers University - New Brunswick , 152 Frelinghuysen Road, Piscataway, NJ 08854-8020

Tél. : (732) 445-0635 Téléc. : (732) 445-6715

Site Web : http://ruccs.rutgers.edu/ruccs/research_language.html

Personne-ressource : Sue Cosentino

Courriel : suecos@ruccs.rutgers.edu

Nombre de postes : 61

Projets de recherche

Le groupe de langage cherche à pousser les limites grammaticales, computationnelles, psychologiques et neuronales à la compréhension du langage, qui interagissent pour astreindre l'architecture du système de traitement du langage. Bruce Tesar (Linguistics, RuCCS) a adapté les techniques dynamiques de passage à base de programmation pour travailler avec les théories grammaticales à partir de la théorie d'optimalité. Les algorithmes Tesar étayent simultanément notre compréhension du traitement et de l'apprentissage du langage. Les algorithmes de passage jouent un rôle important dans les modèles Tesar d'apprentissage du langage. Jerry Fodor (philosophie, RuCCS) a affirmé que certaines propriétés formelles sont nécessaires pour toute architecture computationnelle quant au traitement linguistique humain, ce qui fournira d'autres contraintes aux modèles proposés. Les études empiriques fournissent non seulement des données sur le comportement que le groupe tente d'expliquer, mais elles dégagent également des preuves cruciales sur l'architecture de traitement et ses limites, ainsi que les structures cognitives sous-tendant ces connaissances. Karin Stromswold (Psychologie, RuCCS) et Fodor apportent leurs contributions en réalisant des expériences psycholinguistiques afin d'étudier la structure des connaissances lexiques, syntaxiques et sémantiques. Stromswold va plus loin en explorant les propriétés de l'architecture de traitement du langage par l'utilisation de techniques de neuroimagerie fonctionnelle (TEP et IRM).

Robust Speech Processing Group Center for Spoken Language Research

Robust Speech Processing Group Center for Spoken Language Research, University of Colorado Boulder, Campus Box 594, Boulder, Colorado 80309-0594

Tél. : (303) 735-5148 Téléc. : (303) 735-5072

Site Web : <http://cslr.colorado.edu/rspl/>

Personne-ressource : John H.L. Hansen

Courriel : John.Hansen@colorado.edu

Nombre de postes : 24

Projets de recherche

- Questions de robustesse dans le repérage des phrases, recherche de thèmes.
- Classification des accents et analyse des dialectes dans la reconnaissance de la parole.
- Analyse de la corde vocale ou de la pathologie du tractus
- Amélioration de la parole/évaluation de la qualité des données pour les écouteurs normaux ou avec des troubles auditifs.
- Reconnaissance de la parole dans le bruit, le stress ou sous l'effet Lombard.
- Questions de robustesse dans les interfaces homme-machine, téléphone portable, codification vocale : GSM, CELP, etc.
- Synthèse de la parole et normalisation du parleur.
- Classification, évaluation et synthèse de la parole dans le stress.

The Signal, Speech and Language Interpretation Lab at the University of Washington

Department of Electrical Engineering University of Washington Box 352500 Seattle, WA 98195-2500

Tél. : (206) 543-2150 Téléc. : (206) 543-3842

Site Web : <http://ssli.ee.washington.edu/ssli/>

Personne-ressource : Mary Ostendorf

Courriel : mo@ee.washington.edu

Nombre de postes : 33

Projets de recherche

Domain-dependent speech synthesis – Combiner la prédiction prosodique (formulation et accent) avec la sélection d'unités, les modèles du gradient de la prominence, les modèles d'intonation adaptative.

Acoustic and pronunciation modeling for LVCSR – Utiliser la structure à syllabes et linguistique de haut niveau, autres solutions aux HMM.

Language modeling for LVCSR – modèles dynamiques mixtes à base de dialogues, formation à partir de données clairsemées.

Low-power and efficient software algorithms – Optimiser au niveau logiciel pour réduire la consommation d'électricité des systèmes de la parole qui fonctionnent sur des appareils portatifs.

Statistical modeling for information extraction from speech - identification d'entités nommées dans des données qui contiennent de nombreuses erreurs en utilisant des modèles linguistiques.

Speech Analysis and Interpretation Laboratory, University of Southern California

University of Southern California, 3740 McClintock Avenue, Room EEB430, Los Angeles, CA 90089-2564

Tél. : (213) 740-6432 Téléc. : (213) 740-4651

Site Web : <http://sail.usc.edu/index.php>

Personne-ressource : Shrikanth S. Narayanan

Courriel : shri@sipi.usc.edu

Nombre de postes : 26

Projets de recherche

Le groupe de recherche se consacre aux questions théoriques et aux applications pratiques dans les domaines suivants :

- Traitement de la parole et du langage et reconnaissance automatique de la parole.
- Modélisation de production de la parole, acoustique d'articulation, synthèse de la parole.
- Mod.lisation et traitement biomédicaux du signal, imagerie et instrumentation pour la recherche en matière de parole.
- Interaction homme-machine, dispositifs, système set interfaces multimodales-multimédia.

Human-like Speech Processing Nouvelles orientations dans la reconnaissance de la parole par une initiative de recherche universitaire multidisciplinaire (MURI).

Dynamics of Vocal Tract Shaping Nouvelles méthodes pour étudier la production de la parole humaine en utilisant l'imagerie par r.sonance magnétique en temps réel.

USC Transonic Solutions Permettre la communication humaine via la traduction vocale automatisée pour des domaines critiques tels que le triage médico-chirurgical.

Initial Target Languages: *Anglais, farsi.*

Content-based approach to Indexing, Query and Retrieval of Music par le traitement des signaux et les méthodes à base de connaissances.

Spoken Language Interactions in Preschoolers: Rapprocher les difficultés linguistiques, sociales et technologiques.

Robust Speech Processing for Immersive Interactions.

Automated Synthesis of Expressive Speech for Military Training.

Speech Technology and Research Laboratory

SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493

Tél. : (650) 859 2388 Téléc. : (650) 859 5984

Site Web : <http://www.speech.sri.com/>

Personne-ressource : Kristin Precoda

Courriel : precoda@speech.sri.com

Nombre de postes : 26

Projets de recherche

- La parole dans un environnement bruyant (SPINE).
- Prosodie pour les systèmes de dialogue.
- Reconnaissance de la parole conversationnelle riche en vocabulaire.
- Reconnaissance et compréhension des réunions.
- Événements cachés dans la parole spontanée.
- Extraction de l'information à partir de la parole.
- Agent de communication SRI.
- Reconnaissance du locuteur.
- Mappage de réunion, collecte et annotation de données.

Spoken Language Systems

Spoken Language Systems, MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street 32-G434, Cambridge, MA 02139 USA

Tél. : (617) 253 3049 Téléc. : (617) 258 8642

Site Web : <http://groups.csail.mit.edu/sls//sls-blue-noflash.shtml>

Personne-ressource : Marcia Davidson

Courriel : marcia@csail.mit.edu

Nombre de postes : 25

Projets de recherche

Depuis sa formation en 1989, le groupe des systèmes du langage verbal (SLS) se consacre à la recherche pour la création d'une technologie qui permet aux hommes d'interagir avec les ordinateurs en utilisant le langage naturel parlé. Pour atteindre cet objectif, le groupe s'active dans plusieurs projets de recherche dont :

- Les systèmes conversationnels pour l'interaction homme-machine.
- Recherche multilingue et apprentissage du langage.
- Traitement et interaction multimodaux.
- Technologie et formation dans le domaine du langage humain.
- Prototype de système conversationnel rapide.
- Traitement systématique axé sur l'homme.

Développement de la technologie de base

Afin de promouvoir sa recherche sur les systèmes de langage verbal aux fins d'interaction homme-machine, le groupe SLS a mis au point sa propre série de technologie de parole de base. Ces technologies comprennent :

- Reconnaissance de la parole (SUMMIT).
- Compréhension du langage naturel (TINA).
- Modélisation de dialogues.
- Génération de langage (GENESIS).
- Synthèse de la parole (ENVOICE).

Les technologies de base de la parole peuvent être intégrées pour créer des systèmes conversationnels en utilisant l'architecture GALAXY pour les systèmes de la parole conversationnelle. La création de GALAXY a permis aux développeurs de créer rapidement des systèmes conversationnels pour diverses applications.

Organismes de financement et associations

American association of artificial intelligence

American Association for Artificial Intelligence, 445 Burgess Drive, Suite 100, Menlo Park, California 94025

Tél. : (650) 328 3123 Téléc. : (650) 321 4457

Site Web : <http://www.aaai.org/>

Personne-ressource : Carol McKenna Hamilton

Mission

American Association for Artificial Intelligence 445 Burgess Drive Suite 100 Menlo Park, California 94025

Fondée en 1979, l'American Association for Artificial Intelligence (AAAI) est une société scientifique sans but lucratif qui a pour mission de promouvoir la compréhension scientifique des mécanismes relatifs au comportement intelligent et représentatif et leur incarnation dans les machines. Le traitement du langage naturel est l'un des principaux champs d'étude de l'AAAI. Cette dernière vise à faire comprendre davantage au public l'intelligence artificielle, à améliorer la formation et l'apprentissage des utilisateurs de l'intelligence artificielle, et à orienter les planificateurs et les bailleurs de fonds de la recherche à l'importance et au potentiel des développements actuels et des orientations en matière d'intelligence artificielle.

Parmi les principales activités de l'AAAI, mentionnons l'organisation et le financement des conférences, des symposiums et des ateliers, la publication trimestrielle d'une revue à l'intention de tous les membres, la publication de livres, de comptes rendus et de rapports, et enfin l'attribution de subventions, de bourses et d'autres prix d'honneur.

Applied Voice Input/Output Society (AVIOS)

Applied Voice Input/Output Society (AVIOS), P.O. Box 20817 Sa Jose, CA 95160

Tél. : (408) 323 1783 Téléc. : (408) 323 1782

Site Web : <http://www.avios.com/>

Personne-ressource : Peggie Johnson

Courriel : info@avios.org

Mission

AVIOS a été créée en 1981 en tant qu'association mutuelle professionnelle sans but lucratif. Depuis, et durant la plupart des années, la conférence annuelle d'AVIOS a été le seul forum consacré aux applications pratiques de la technologie avancée de la parole (au-delà de la recherche de base). La conférence est devenue récemment l'AVIOS-SpeechTEK Spring Conference grâce à un partenariat avec AmComm.

Après avoir établi cette forte base, AVIOS met désormais l'accent sur son rôle d'association mutuelle professionnelle mondiale.

Mission

AVIOS est une société professionnelle qui se consacre à promouvoir le développement et la diffusion d'applications de la technologie de la parole dans le monde réel. AVIOS crée des liens entre les utilisateurs, les développeurs et les chercheurs visant à faire avancer la technologie de la parole en tant qu'interface de choix. AVIOS permet à ses membres d'apprendre davantage sur les développements du marché et techniques et d'échanger des communications informelles. Grâce à Avios, les chercheurs découvrent ce dont ont besoin les développeurs, les spécialistes en commercialisation et les utilisateurs finaux dans les produits et les services automatiques liés à la parole. Les développeurs et les spécialistes en commercialisation apprennent à définir les applications appropriées relatives à la technologie de la parole et la façon de les présenter.

Association for machine translation of America

Priscilla Rasmussen, AMTA Focal Point, 3 Landmark Center, East Stroudsburg, PA 18301

Tél. : (570) 476 8006 Téléc. : (570) 476 0860

Site Web : <http://www.amtaweb.org/>

Personne-ressource : Priscilla Rasmussen

Courriel : focalpoint@amtaweb.org

Mission

AMTA est une association destinée à chaque personne qui s'intéresse à la traduction des langues en utilisant un ordinateur, notamment les gens qui ont besoin de traduire des textes, les développeurs de systèmes commerciaux, les chercheurs, les bailleurs de fonds et les personnes qui étudient, évaluent et comprennent la science de la traduction automatique (TA), et les personnes qui veulent sensibiliser le public aux importants principes et techniques scientifiques. L'association permet aux développeurs commerciaux, aux utilisateurs, que ce soit les traducteurs ou les personnes qui utilisent de temps à autre la TA, et aux chercheurs de se rencontrer dans une ambiance conviviale et de partager les idées, les rêves et les souhaits.

AMTA a des membres au Canada, en Amérique du Sud et aux États-Unis. Il constitue une partie régionale d'un réseau mondial présidé par l'association internationale pour la traduction automatique (IAMT). L'AMTA a deux sœurs : la Asian-Pacific Association for Machine Translation (AAMT) et la European Association for Machine Translation (EAMT). Le conseil de l'IAMT jouit d'une représentation équilibrée qui couvre toutes les régions. Les personnes, les institutions et les corporations peuvent s'adhérer à l'une des trois associations régionales et devenir automatiquement membres du réseau de l'IAMT.

AMTA facilite l'accès des chercheurs au corpora exploitable par une machine et la coopération pour l'échange de conventions de formats et d'encodage de textes, ainsi qu'aux discussions sur l'établissement de critères de référence pour l'évaluation de la technologie.

Les membres de l'AMTA s'engagent à mettre au point des programmes et du matériel de formation ainsi qu'à tenir des listes sur les dernières technologies et innovations.

IEEE Signal Processing Society

IEEE Signal Processing Society, 445 Hoes Lane, Piscataway, NJ, 08854

Tél. : (732) 562 3888 Téléc. : (732) 235 1627

Site Web : <http://www.ieee.org/organizations/society/sp/>

Personne-ressource : Alfred O. Hero

Courriel : sp.info@ieee.org

Mission

La mission de la Signal Processing Society consiste en la théorie et l'application en matière de filtrage, programmation, transmission, estimation, repérage, analyse, reconnaissance, synthèse, enregistrement et reproduction de signaux par des dispositifs ou des techniques numériques ou analogues. Le terme « signal » comprend l'audio, la vidéo, la parole, l'image, la communication, la géophysique, le radar, les signaux médicaux et musicaux et autres.

La Signal Processing Society est un organisme international dont l'objectif est : de faire avancer et publier l'information et les ressources scientifiques actuelles; sensibiliser les spécialistes en traitement des signaux; permettre aux gens d'interagir et d'échanger les idées.

La Signal Processing Society est un organisme scientifique qui constitue la principale source d'information et de ressources en matière de traitement des signaux à l'intention de la communauté mondiale. Pour ce faire, nous nous efforçons d'être une source unique de ressources de traitement des signaux, nous fournissons diverses ressources de grande qualité à divers utilisateurs dans des formats personnalisés selon leurs intérêts; nous nous adaptons à une communauté qui évolue rapidement, nous continuons à participer activement à la formation des professionnels en traitement des signaux à tous les niveaux.

Linguistic data Consortium

Linguistic Data Consortium, 3600 Market Street, Suite 810, Philadelphia, PA, 19104-2653

Tél. : (215) 898 0464 Téléc. : (215) 573 2175

Site Web : <http://www ldc.upenn.edu/>

Personne-ressource : Mark Liberman

Courriel : myl@ldc.upenn.edu

Mission

La Linguistic Data Consortium est un consortium d'universités, de compagnies et de laboratoires gouvernementaux de recherche. Il crée, recueille et distribue les bases de données de la parole et des textes, les lexiques et d'autres ressources aux fins de recherche et de développement. Le LDC est basé à l'Université de Pennsylvanie.

Le LDC compte plusieurs projets qui favorisent l'enseignement des langues, la recherche et le développement de la recherche.

ACE – En appui de l'Automatic Content Extraction Program, le LDC développe le corpora dans des entités annotées en anglais, en chinois et en arabe, la relation entre elles et les événements dans lesquels elles participent.

GALE - LDC conçoit des ressources linguistiques intégrées et développe l'infrastructure connexe pour favoriser les technologies d'exploitation linguistique dans le programme DARPA GALE (Global Autonomous Language Exploitation).

LCTL (Less Commonly Taught Languages) - LDC crée et distribue des ressources linguistiques incluant le texte monolingue et parallèle, les lexiques, les convertisseurs de codage, les segmenteurs de mots et de phrases, les agents d'analyse morphologiques, agents d'étiquetage d'entités nommées, annotations, infrastructure d'annotation et caractéristiques dans un nombre de langues qui sont moins enseignées. Dans sa première année, le LCTL cible sept langues qui sont parlées par plus d'un million de personnes, mais qui ne possèdent pas les ressources suffisantes pour mettre au point des technologies de langage humain : l'urdu, le thaï, le hongrois, le bengali, le punjabi, le tamil et le yoruba.

TIDES - LDC consacre jusqu'à une heure par jour à chacune des langues dans lesquelles Voice of America diffuse pour promouvoir le projet DARPA Translingual Information Detection Extraction and Summarization. Voici les sous-projets de TIDES :

Extraction – Création de corpus visant à promouvoir l'extraction d'entités, de relations et d'événements à partir de textes en tant que technologie TIDES et conjointement avec le programme Automatic Content Extraction.

Retrieval Group

100 Bureau Drive, Stop 8940, Gaithersburg, MD 20899-8940

Site Web : <http://www.itl.nist.gov/iaui/894.02/>

Personne-ressource : Ellen M. Voorhees

Courriel : ellen.voorhees@nist.gov

Mission

Le Retrieval Group de l'Information Access Division travaille avec le secteur industriel, universitaire d'autres agences gouvernementales afin de promouvoir l'utilisation de techniques efficaces et efficientes pour manipuler (largement) l'information textuelle non structurée, particulièrement par le parcours, la recherche et la présentation de l'information.

Activités (groupées en quatre domaines principaux) :

- Encourager la recherche sur la consultation d'information incluant de grands fichiers non structurés en fournissant des collections aux fins de test et en organisant la conférence TREC et son compte rendu.
- Continuer à créer des collections aux fins de test, en visant principalement les collections pour promouvoir les sous-tâches particulières de recherche d'information, telles que la recherche interlingue et la recherche multimédia.
- Concevoir une meilleure méthode d'évaluation pour l'accès à l'information, notamment des mesures d'évaluation améliorées pour comparer les systèmes en utilisant des collections aux fins de test et de nouvelles mesures d'évaluation aux fins de recherche interactive et d'opérations de parcours.
- Permettre un développement plus rapide de systèmes prototypes de recherche (secteur commercial, les universités et le gouvernement) par la distribution d'un moteur de recherche avancé (système PRISE) qui comprend la mise en place d'un serveur/client Z39.50.

Open Mind Initiative

Site Web : <http://www.openmind.org>

Personne-ressource : David G. Stork

Courriel : stork@openmind.org

Projets de recherche

La Open Mind Initiative est un cadre de travail conjoint pour la mise au point d'un logiciel intelligent au moyen d'Internet. En se fondant sur la méthode conventionnelle à source ouverte, il appuie les modules experts (qui fournissent les algorithmes), les développeurs d'outils (qui fournissent l'infrastructure logicielle et les outils) et les « surfeurs » non spécialisés (qui versent des données brutes).

Après des dizaines années de recherche, il existe toujours beaucoup de tâches dans lesquelles les ordinateurs sont pires que les hommes : reconnaître la parole, lire des textes imprimés et manuscrits, reconnaître les objets à partir d'images, comprendre les scènes, concevoir des plans complexes, résumer une histoire, etc. Il est clair qu'un logiciel de ce type serait extrêmement utile s'il venait à atteindre la performance humaine, ou même s'il améliorerait modestement ou aiderait la cognition humaine. La recherche initiale en intelligence artificielle s'est concentrée sur les petits problèmes de « jouets » comme les jeux. Il est devenu évident que nous avons besoin d'information contenue dans de très grandes séries de données.

Le principal objectif d'Open Mind consiste à concevoir un logiciel intelligent, en partie par la collecte de très grandes séries de données et la fourniture d'une infrastructure ouverte de différentes idées peuvent être essayées. Les données et le logiciel conséquent sont mis à la disposition de tout le monde.

Un autre objectif consiste à sensibiliser le public aux problèmes de la science cognitive, de l'informatique, de la reconnaissance de modes et des domaines connexes.

Finlande

Laboratoires

From data to Knowledge research unit
<p>Department of Computer Science, P.O. Box 68, (Gustaf Hällströmin katu 2b), FIN-00014, UNIVERSITY OF HELSINKI</p> <p>Tél. : (358) 9 191 51280 Téléc. : (358) 9 191 51120</p> <p>Site Web : http://www.cs.helsinki.fi/research/fdk/</p> <p>Personne-ressource : Esko Ukkonen</p> <p>Courriel : Esko.Ukkonen@cs.Helsinki.FI</p> <p>Nombre de postes : 58</p>
Projets de recherche
<p>L'unité de recherche From Data To Knowledge (FDK) élabore des méthodes pour former des connaissances utiles à partir de grandes quantités de données. L'unité est multidisciplinaire puisqu'elle intègre dans ses groupes de recherche l'excellence dans les algorithmes discrets, les techniques statistiques et les sciences d'application.</p> <p>Les principaux outils méthodologiques de l'unité de recherche constituent une combinaison d'harmonisation des modes et d'exploration de données. La combinaison des deux est unique dans le monde. Les travaux combinent les progrès conceptuels, les méthodes algorithmiques, statistiques et analytiques et les travaux empiriques : la théorie et la pratique main dans la main.</p>

Helsinki Institute for Information Technology, Basic Research Unit

Helsinki Institute for Information Technology (HIIT), Basic Research Unit (BRU), PO Box 68
(Gustaf Hällströmin katu 2b), FI-00014 University of Helsinki, FINLAND

Tél. : (358) 0 1911 Téléc. : (358) 9 191 51120

Site Web : http://www.cs.helsinki.fi/hiit_bru/

Personne-ressource : Ukkonen Esko

Courriel : Esko.Ukkonen@cs.helsinki.fi

Nombre de postes : 60

Projets de recherche

La Basic Research Unit (BRU) met l'accent sur la recherche à long terme de base en informatique, dans des domaines où les applications sont visibles dans d'autres sciences ou dans l'industrie.

Les groupes de recherche actuels travaillent sur l'analyse des données, les systèmes informatiques adaptatifs et la neuroinformatique. L'unité travaille en étroite collaboration avec les groupes d'application dans d'autres sciences (comme la génétique médicale, la biologie structurale et la neuroscience) et l'industrie. La BRU a été créée au début de 2002.

Laboratory of Acoustics and Audio Signal Processing

Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing,
Otakaari 5 A, 02150 Espoo, Finland

Tél. : (358) 9 451 2496 Téléc. : (358) 9 460 224

Site Web : <http://www.acoustics.hut.fi/>

Personne-ressource : Paavo Alku

Courriel : Paavo.Alku@hut.fi

Nombre de postes : 35

Projets de recherche

- Analyse et paramétrage de la production vocale.
- Traitement cérébral des sons de la parole.
- Algorithmes prédictifs pour la modélisation spectrale de la parole.
- Projet de l'Union européenne : Réseau Improving Human Potential (IHP) du Hearing. Organisation and Recognition of Speech in Europe (HOARSE).
- Kieliteknologian opetuksen verkosto (KIT)(réseau national de technologie de l'information).
- Graduate School of Language Technology.
- Analyse acoustique des émotions dans la parole.
- Expansion artificielle de la bande passante dans la parole téléphonique.
- Analyse acoustique de la dysphonie.
- Modèles paramétriques de la parole.
- Traduction de la voix en texte (reconnaissance de la parole phonémique).
- Acoustique neuronale.
- Avesound – reconnaissance automatique des espèces d'oiseaux par le son qu'ils produisent.

Fonctions cérébrales évoquées par la parole et les sons tridimensionnels

Mesure des réponses cérébrales à différents signaux de la parole et spatiaux.

Émotions dans la parole

Repérage et modélisation des émotions implantées dans la qualité de la voix du parleur.

Recherche relative à l'utilisation de la voix dans le travail

La recherche en matière d'utilisation de la voix dans le travail porte sur les caractéristiques vocales des professionnels de la parole avec des mesures objectives dans différentes circonstances.

Paramétrage de la source vocale

Le paramétrage des caractéristiques dans l'activité glottique, particulièrement dans le flux glottique.

Multimodal Interfaces Group, Adaptive Informatics Research center

Helsinki University of Technology Laboratory of Computer and Information Science P.O. Box
5400 FI-02015 HUT FINLAND

Tél. : (358) 9 451 3267 Téléc. : (358) 9 451 3277

Site Web : <http://www.cis.hut.fi/research/multimodal/>

Personne-ressource : Mikko Kurimo

Courriel : Mikko.Kurimo@hut.fi

Nombre de postes : 125

Projets de recherche

La recherche du Multimodal Interfaces Group vise à résoudre les problèmes de l'explosion d'information en fusionnant plusieurs modes accessibles perceptuels et relatifs à la rétroaction des utilisateurs. Les résultats seront appliqués dans des interfaces interactives (p. ex. dans la recherche multimodale axée sur le contenu). Le Multimodal Interfaces Group coordonne nos travaux relatifs aux sous-sujets pertinents, notamment la recherche d'images, la reconnaissance de la parole, la recherche d'information proactive et le traitement du langage naturel.

Projet Morpho : la découverte non supervisée de morphèmes et la morphologie d'un langage naturel ainsi que la segmentation de mots. Les morphes découvertes (segments de mots) ont été appliquées afin d'améliorer la reconnaissance de la parole en finlandais.

Semantic Computing Research Group

SCRG, Helsinki University of Technology (TKK) Media Technology Room 2540 P.O. Box 5500
FI-02015 TKKFinland

Tél. : (358) 9 451 3362 Téléc. : (358) 9 451 3356

Site Web : <http://www.seco.tkk.fi/>

Personne-ressource : Eero Hyvönen

Courriel : eero.hyvonen@tkk.fi

Nombre de postes : 21

Projets de recherche

Notre recherche vise la sémantique traitable par machine, c'est-à-dire que nous étudions les techniques.

Pour représenter les données et les connaissances de façon à ce que les machines peuvent « comprendre » leur sens et élaborer des méthodes algorithmiques visant à créer des applications intelligentes axées sur ces représentations.

Cette recherche est associée à plusieurs domaines de recherche et d'applications en informatique, entre autres l'intelligence artificielle et l'ingénierie de connaissances, le Web sémantique et les services Web, le traitement du langage naturel et la recherche d'information.

Notre groupe est actif à l'Université de technologie d'Helsinki, au département d'information de l'Université d'Helsinki et l'Institut de technologie d'information d'Helsinki (HIIT).

Le projet de recherche Intelligent Web Services (IWebS) étudie les possibilités des technologies relatives au Web sémantique et aux services Web pour annoter des services et fournir des services pertinents aux utilisateurs finaux.

Notre objectif vise à créer une base pour des ontologies nationales finlandaises relatives au Web sémantique. Les travaux couvrent ce qui suit :

- La recherche sur la technologie d'ontologie (incertitude dans les ontologies, le contrôle des versions d'ontologie, le développement d'ontologie distribuée);
- La création d'ontologies nationales centrales à partir de vocabulaires contrôlés, comme le Finnish Webthesaurus et MASA, vocabulaire contrôlé utilisé dans plusieurs musées finlandais, et l'application de plusieurs problèmes faisant l'objet d'études de cas.

De 2002 à 2004, nous avons étudié les points de vue relatifs au stockage et à la recherche d'information, comment les données riches sémantiquement devraient être annotées avec des métadonnées afin de faciliter la recherche d'information axée sur le contenu. Les bases de données artéfact culturelles et les dépôts de photographes sont utilisés en tant qu'études de cas.

France

Laboratoires

Analyse et traitement informatique de la langue française (ATILF)

ATILF - CNRS/Université Nancy 2, 44 avenue de la Libération, B.P. 30687, F 54063 NANCY CEDEX

Tél. : +33 3 83 96 21 76 Téléc. : +33 3 83 97 24 56

Site Internet : <http://www.atilf.fr/>

Personne-ressource : Jean-Marie Pierrel

Courriel : jean-marie.pierrel@atilf.fr

Nombre de postes: 94

Projets de recherche

L'unité mixte de recherche ATILF (Analyse et Traitement informatique de la Langue Française) est née au 1er janvier 2001 du rapprochement de l'Institut National de la Langue Française (INALF - CNRS) et de LANDISCO (Langue Discours Cognition - Université Nancy 2 Rattachée au Département Sciences de l'Homme et de la Société du CNRS, elle est aussi laboratoire d'interface avec le Département Sciences et Technologies de l'Information et de la Communication. Elle est membre de la fédération CNRS - Institut de Linguistique Française et du consortium international TEI (Text Encoding Initiative).

- Linguistique informatique : les activités de cet axe sont centrées sur la définition, la gestion et l'usage de ressources linguistiques informatisées en traitement automatique des langues.
- Développement informatique des outils, des bases et des ressources informatisées avec une gestion unifiée de nos ressources codées en XML autour du logiciel STELLA développé au laboratoire, en particulier le Trésor de la Langue Française informatisé (TLFi) et la base de données textuelles Frantext.
- Ressources linguistiques et méta-données, maintenance et évolution avec deux chantiers principaux : veille lexicographique en vue d'assurer l'évolution nécessaire du Trésor de la Langue Française informatisé (TLFi) et enrichissement de Frantext (Base de données textuelles) en fonction des demandes des projets scientifiques.
- Traitement automatique de la langue avec comme objectif d'offrir des ressources (base de connaissances et outils d'analyse et de traitement) intégrables à des chaînes d'ingénierie linguistique.

École nationale supérieure de télécommunications, Groupe Information, Interaction, Intelligence

46 rue Barrault, 75634 Paris

Tél. : +33 1 45 81 78 70 Téléc. : +33 1 45 81 31 19

Site Internet : <http://www.infres.enst.fr/>

Personne-ressource : Nicolas Puech

Courriel : npuech@enst.fr

Nombre de postes: 20

Projets de recherche

Les systèmes de traitement de l'information sont entrés dans une phase où ils se caractérisent par le besoin de traiter de très grandes quantités d'informations et de connaissances, ce qui implique de disposer de moyens d'interaction conviviaux et performants, et des outils de traitement automatique intelligents. Dans ce contexte le groupe 3I, dont le nom est inspiré du GDR-PRC I3 créé en 1997 auquel le groupe participe, a centré ses activités de recherche dans les domaines suivants :

Information : recherche d'informations dans de grandes bases documentaires, optimisation d'accès à des informations multidimensionnelles, visualisation de grandes quantités d'informations

Interaction : reconnaissance de l'écriture et de la parole, dialogue humain-machine, interaction avec un monde virtuel

Intelligence : cet aspect caractérise qualitativement les traitements mentionnés ci-dessus; l'intelligence peut-être obtenue soit par des techniques spécifiques aux données manipulées, soit par le biais d'une approche cognitive.

- Recherche et visualisation de l'information.
- Information Interaction.
- Bases de données multidimensionnelles, spatiales et spatio-temporelles.
- Linguistique à base de corpus.
- Information Intelligence.
- Modélisation du langage et des stratégies de communication.

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Département Communication Homme-Machine

LIMSI-CNRS BP 133, F-91403 ORSAY CEDEX (France)

Tél. : (33)1 69 85 80 80 Téléc. : (33) 1 69 85 80 88

Site Web : <http://www.limsi.fr/Recherche/CHMdpgb.html>

Personne-ressource : Philippe Tarroux

Courriel : philippe.tarroux@limsi.fr

Nombre de postes : 31

Projets de recherche

Le Département Communication Homme-Machine réunit plusieurs groupes qui nous permettent de comprendre les divers modes de communications sous différents angles. Ces activités portent sur le traitement du langage parlé (analyse, synthèse et perception), le traitement du langage naturel (analyse, compréhension et production de textes, dialogues écrits, représentations sémantiques), la vision informatique, l'analyse et la synthèse des images, l'analyse et la synthèse des gestes, la communication multimodale, les facteurs humains et l'étude des processus cognitifs. Collabore à ce département des chercheurs dans le domaine de l'ingénierie (informatique, traitement des signaux et intelligence artificielle) et de la linguistique, de la psychologie cognitive et des spécialistes des questions socio-économiques qui touchent l'innovation.

Notre recherche en **reconnaissance de la parole** met l'accent sur la segmentation et la transcription de la parole continue ainsi que sur l'identification du locuteur et du langage dans le but d'annoter automatiquement et de structurer les documents audio. Les progrès relatifs à la reconnaissance de la parole, qui comptent sur la promotion de la recherche dans la modélisation acoustique phonétique, la modélisation lexicale et enfin la modélisation linguistique, sont réalisés dans un contexte multilingue (anglais, français, allemand, mandarin, portugais, espagnol, arabe, etc.). L'approche sous-jacente vise à élaborer des modèles et des algorithmes qui tiennent compte des diverses sources d'information afin de décoder globalement le signal auditif.

Le **dialogue parlé homme-machine** est un domaine de recherche à plusieurs facettes qui exige une modélisation de la parole spontanée, un processus de communication et l'élaboration d'un module de dialogue. Ces travaux ont abouti à des systèmes de dialogue pour l'accès à l'information. L'interaction peut compter entièrement sur la parole (par téléphone) ou peut être associée à d'autres moyens d'interaction, comme l'écran tactile (kiosque multimodal).

Laboratoire des sciences de l'information et des systèmes, INCOD

Domaine Universitaire de Saint-Jérôme, Avenue Escadrille Normandie-Niemen, 13397
MARSEILLE CEDEX 20

Tél. : (33) 04 91 05 60 30 Téléc. : (33) 04 91 05 60 33

Site Internet : <http://www.lsis.org>

Personne-ressource : Norbert Giambiasi

Courriel : norbert.giambiasi@lsis.org

Nombre de postes: 34

Projets de recherche

Les travaux de l'équipe INCOD concernent l'ingénierie des systèmes d'information et de connaissance en s'intéressant plus particulièrement à leurs aspects distribués, coopératifs et de sûreté. Les recherches de l'équipe s'articulent autour de quatre thèmes majeurs, dans lesquels s'impliquent, de façon non exclusive, ses chercheurs :

- Réutilisation en conception de systèmes d'information.
- Modélisation et simulation multi-agents pour l'aide à la décision.
- Recherche et intégration d'informations hétérogènes.

Laboratoire d'Automatique Documentaire et Linguistique

77454 Marne-la-Vallée, Université Marne-la-Vallée

Site Internet : <http://ladl.univ-mlv.fr/>

Personne-ressource : M. Mathis

Courriel : mmathis@univ-mlv.fr

Nombre de postes : 20

Projets de recherche

L'équipe d'informatique linguistique se situe dans la perspective du traitement automatique des textes en langues naturelles. Elle privilégie l'utilisation de données précises et explicites: dictionnaires, grammaires, par rapport à l'approximation à partir de données incertaines [Lap97b]. Les applications visées sont nombreuses mais les plus significatives sont liées à la recherche documentaire. Les trois orientations qui nous intéressent le plus actuellement sont les suivantes :

La production de dictionnaires électroniques d'autres langues que celles pour lesquelles on dispose déjà d'outils fiables ;

Les traitements intermédiaires entre l'analyse lexicale et l'analyse syntaxique en vue de l'accès aux informations dans les grandes bases de textes : reconnaissance et indexation de terminologie, levée d'ambiguïtés...

Ces travaux prennent leur place dans le contexte plus général du réseau RELEX de laboratoires de linguistique et d'informatique linguistique. Le réseau RELEX regroupe une douzaine de laboratoires, situés pour la plupart en Europe, qui collaborent à la constitution d'un inventaire d'informations linguistiques précises et exploitables dans les traitements automatiques, sur la base d'exigences méthodologiques : reproductibilité, exhaustivité, cumulativité. Dans cet ensemble, notre équipe, depuis sa création en 1987 sous le nom de Centre d'études et de recherches en informatique linguistique (CERIL), joue avec le Laboratoire d'automatique documentaire et linguistique (LADL) de Paris 7, un rôle central, par l'organisation commune de colloques et à travers des contributions croisées dans l'élaboration des outils fondamentaux : dictionnaires et logiciels.

Le programme global de recherches du réseau RELEX est simple. Les unités élémentaires des langues sont décrites dans des dictionnaires électroniques. Ces unités pouvant se combiner pour former des phrases et des textes, la description formelle des contraintes syntaxiques est mise en oeuvre sous la forme de lexiques-grammaires. Les descriptions linguistiques (grammaticales, morphologiques, syntaxiques, phonétiques) étant destinées à être exploitées dans des applications informatiques, les objectifs sont extrêmement concrets: spécifier les formes effectivement observables. Le passage de la description linguistique à la réalisation informatique est facilité par l'utilisation d'automates finis comme support commun.

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

LIMSI-CNRS BP 133, F-91403 ORSAY CEDEX (France)

Tél. : (33) 1 69 85 80 85 Téléc. : (33) 1 69 85 80 88

Site Internet : <http://www.limsi.fr/>

Personne-ressource : Patrick Le Quéré

Courriel : plq@limsi.fr

Nombre de postes: 180

Projets de recherche

Le LIMSI est une Unité Propre de Recherche du CNRS, associée aux universités Pierre et Marie Curie (Paris-6) et Paris-Sud (Paris-11). Le laboratoire accueille environ 120 permanents (chercheurs, enseignants-chercheurs et ITA-IATOS) et une soixantaine de doctorants. Il mène des recherches pluridisciplinaires en Mécanique et Energétique et en Sciences et Technologies de l'Information et de la Communication, couvrant un large spectre disciplinaire allant du "thermodynamique au cognitif", en passant par la mécanique des fluides, l'énergétique, l'acoustique et la synthèse vocale, le traitement de la langue parlée et du texte, la vision, la réalité virtuelle. Le LIMSI a l'ambition et la volonté d'allier excellence disciplinaire dans l'ensemble de ses champs scientifiques et développement de projets pluridisciplinaires, en partenariat avec d'autres organismes de recherche et des industriels, français et étrangers. Il s'investit tout particulièrement dans la formation par la recherche et la diffusion de l'information scientifique et technique.

Laboratoire de Linguistique Formelle

UMR 7110, Université de Paris 7, Tour centrale, Place Jussieu 75005 Paris

Tél. : (+33)(0)1 57 27 57 94

Site Internet : <http://www.llf.cnrs.fr/>

Personne-ressource : Alain Kihm

Courriel : alain.kihm@linguist.jussieu.fr

Nombre de postes: 67

Projets de recherche

Le Laboratoire de Linguistique Formelle (LLF) est un laboratoire de linguistique théorique dont le projet scientifique est centré sur la description systématique de langues appartenant à des familles typologiquement variées, la confrontation de théories linguistiques distinctes, les interfaces entre les composantes de la grammaire (syntaxe et phonologie, syntaxe et morphologie, syntaxe et sémantique, syntaxe et pragmatique).

Chacune des opérations de recherche (Morphophonologie et interfaces, Syntaxe et sémantique formelle, Unités signifiantes et structures énonciatives s'attachent à la description détaillée des phénomènes linguistiques dans un cadre grammatical explicite. Le domaine empirique va du phonème à l'énoncé. Les sous-disciplines de la linguistique représentées au sein du laboratoire sont : la phonologie, la morphologie, la syntaxe, la prosodie, la sémantique (lexicale, phrastique et discursive) et la pragmatique (pragmatique formelle et théorie de l'énonciation).

LORIA Laboratoire lorrain de recherche en informatique et ses applications

LORIA - Campus Scientifique - BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex

Tél. : +33 3 83 59 30 00 Téléc. : +33 3 83 27 83 19

Site Internet : <http://www.loria.fr/>

Personne-ressource : Yves Laprie

Courriel : yves.laprie@loria.fr

Nombre de postes: 450

Projets de recherche

Les thématiques scientifiques de l'équipe Langue et Dialogue s'articulent autour des modèles des techniques et des ressources permettant de mettre en oeuvre à court, moyen ou long terme des systèmes de dialogue homme-machine finalisés reposant sur une forte composante langagière. Dans ce cadre, notre activité se développe dans trois directions complémentaires : L'étude des mécanismes fondamentaux de la communication en langue naturelle seule ou accompagnée d'une désignation gestuelle (dialogue multimodal). Cette recherche s'effectue dans un contexte pluridisciplinaire alliant linguistique et informatique principalement ; La réalisation de systèmes de dialogue effectifs dans le cadre notamment de collaborations industrielles. Cette activité nous permet par ailleurs de disposer d'une plate-forme d'expérimentation pour la validation des différents modèles que nous concevons ; La définition d'outils et de méthodes génériques permettant d'étudier de façon fine des situations de dialogues réels, issus de la transcription d'expériences de simulation ou d'observations directes. Ce travail s'appuie sur une expérience acquise depuis plusieurs années sur la normalisation et la manipulation de ressources linguistiques (en particulier des "corpus").

Le but des membres de l'action Orpailleur est de concevoir des systèmes à base de connaissances pour raisonner et résoudre des problèmes dans des domaines d'application donnés. Les connaissances mises en oeuvre ont plusieurs formes et proviennent de sources différentes : elles peuvent être fournies de façon explicite par des spécialistes ou être extraites dans des bases de données de toutes natures. Les connaissances sont ensuite exploitées par des procédures de raisonnement pour résoudre des problèmes dans différents domaines d'application. L'action Orpailleur s'intéresse principalement à l'agronomie, l'analyse de textes scientifiques et techniques, la chimie organique et la planification de synthèses, la classification de signaux temporels, la médecine et la sidérurgie.

Laboratoire informatique d'Avignon

339 Chemin des Meinajariès – BP 1228 – 84911 AVIGNON Cedex 9

Tél. : 04 90 84 35 09 Téléc. : 04 90 84 35 01

Site Internet : <http://www.lia.univ-avignon.fr>

Personne-ressource : Renato De Mori

Courriel : marc.elbeze@lia.univ-avignon.fr

Nombre de postes: 66

Projets de recherche

Les axes principaux de recherche de l'équipe sont les suivants :

- Conception de modèles de langage permettant de minimiser les erreurs sur la génération d'hypothèses générées par un système de reconnaissance automatique de la parole ;
- Reconnaissance automatique de la parole pour l'accès à des annuaires électroniques de grande taille ;
- Utilisation du dialogue et de l'expansion des requêtes en recherche documentaire.

L'équipe PAROLE du LIA mène des travaux dans le domaine de la reconnaissance et la parole et du locuteur.

Projets du lia en Recherche d'Information

Moteurs de question/réponse. Le système de recherche SIAC peut être exploité pour trouver une zone d'un document et non un document dans son intégralité en réponse à une requête. Cette possibilité, couplée au module d'extraction d'entités nommées développé au LIA (F. Béchet) permettra au laboratoire de participer à la prochaine campagne d'évaluation TREC -10 (2001) pour les moteurs de question-réponse (cette problématique sera traitée dans le cadre de la thèse de F. Genet co-dirigée par M. El-Bèze, F. Béchet et P. Bellot).

Moteurs de recherche documentaire. Le LIA entend continuer à développer le moteur de recherche du système SIAC pour porter ses performances au niveau de celles obtenues par les meilleurs systèmes du moment. Pour mettre en pratique le principe que ses systèmes doivent répondre aux critères fixés par le LIA, nous prévoyons de participer aux prochaines campagnes d'évaluation TREC et Amarylilis.

Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen

6, Boulevard du Maréchal Juin, 14050 CAEN Cedex

Tél. : +33 (0)2.31.56.74.35 Téléc. : +33 (0)2.31.56.73.30

Site Internet : <http://www.greyc.unicaen.fr>

Personne-ressource : Bruno Cremilleux

Courriel : Bruno.Cremilleux@info.unicaen.fr

Nombre de postes: 147

Projets de recherche

L'équipe "Données, Document, Langue" est une des sept équipes du laboratoire GREYC CNRS UMR 6072. Elle comporte 10 permanents, 17 doctorants et 1 chercheur associé. L'équipe a pour ambition de faciliter la production, l'accès et les usages du document électronique mais aussi de façon plus générale aux "données" et ses travaux concernent entre autres le document géographique, la langue et l'extraction de connaissances à partir de bases de données. Les activités scientifiques de l'équipe sont structurées en trois thèmes (brièvement présentés ci-dessous). L'équipe favorise les projets inter-thèmes ainsi que les collaborations avec d'autres équipes et laboratoires. Elle initie et développe des collaborations entre chercheurs issus de domaines et de communautés de pensée différentes autour des multiples facettes recoupant la production, la représentation, la diffusion et l'utilisation de l'information.

Le groupe "*fouille de données et apprentissage*" propose des modèles interactifs pour améliorer l'appréhension des informations contenues tant dans un document que dans de grandes collections de documents. Il s'intéresse aussi bien aux méthodes (algorithmique de l'extraction) qu'aux multiples usages (e.g., interprétation de données, caractérisation de classes, clustering) des résultats d'un processus d'extraction de connaissances à partir de données.

Partant du constat que le développement des technologies de l'information et de la communication mène à une situation paradoxale, offrant d'une part de plus en plus de facilités pour la production et la diffusion de documents électroniques, mais rendant d'autre part la recherche et l'accès à l'information de plus en plus complexe, le groupe "*document électronique composite*" étudie la nature et les représentations du document électronique, avec un intérêt particulier pour le document géographique.

Le groupe "*langue pour le document*" s'intéresse plus spécifiquement à l'information textuelle, élaborant des modèles linguistiques pour la recherche d'information, en articulation avec des travaux sémantiques plus fondamentaux. Les travaux du groupe comportent aussi bien des travaux théoriques visant à proposer des modèles informatiques, notamment en sémantique et en rhétorique et des aspects applicatifs plus directement liés au traitement automatique des langues et permettant de confronter les modèles au matériau linguistique.

Groupe de traitement du langage parlé, Laboratoire d'Informatique et de Mécanique pour les Sciences de l'Ingénieur

LIMSI – CNRS, B.P. 133, 91403 ORSAY CEDEX, France

Tél. : +33 1 69 85 80 80 Téléc. : +33 1 69 85 80 88

Site Internet : <http://www.limsi.fr/tlp/>

Courriel : Personne-ressource@limsi.fr

Nombre de postes: 30

Projets de recherche

Les recherches du groupe visent à augmenter notre compréhension des processus de la communication parlée et à développer des modèles appropriés au traitement automatique de la parole. Nos problèmes scientifiques concernent aussi bien les modélisations acoustique, lexicale et syntaxique, que le lien entre parole et sens, ainsi que la modélisation des processus de communication. Ces problèmes, par essence pluridisciplinaires, nécessitent des compétences en traitement du signal, en acoustique, en phonétique, en linguistique et en informatique. Nos recherches nous amènent à développer des systèmes multilingues de traitement du langage parlé assurant des fonctions variées telles que la reconnaissance de la parole, l'identification de la langue et du locuteur, le dialogue oral homme-machine et l'indexation de documents audio et audiovisuels.

- TC-STAR (Technology and Corpora for Speech to Speech Translation)
- CHIL (Computers in the Human Interaction Loop)
- HUMAINE: Network of Excellence (Emotions and human-machine interaction)
- EARS RT project (Rich transcription project)
- AUDIOSURF (RNTL) (Indexation Automatique et Recherche Intuitive dans des Documents Audio)
- STIC-SHS MIDL (Modélisation pour l'identification des langues)
- AMITIES (Automated Multilingual Interaction with Information and Services)
- MEDIA-EVALDA (Project Technolanguage sur l'évaluation du dialogue hors et en contexte)
- Processing multilingual broadcast audio for information access
- LVCSR (Large Vocabulary Speech Recognition)

Organismes et associations françaises

Association francophone de la communication parlée

AFCP, Université D'Avignon, 74 rue Louis Pasteur, 84029 Avignon cedex 1

Tél. : +33 (0)4 90 16 25 00 Téléc. : +33 (0)4 90 16 25 10

Site Internet : <http://www.afcp-parole.org/>

Personne-ressource : Jean-Francois Bonastre

Courriel : Personne-ressource-afcp@afcp-parole.org

Mission

L'*Association Francophone de la Communication Parlée* est une association à but non-lucratif (loi du 1er juillet 1901) consacrée au soutien, au développement, à la diffusion et à la promotion des différentes spécialités des sciences de la communication parlée, dans la communauté francophone.

Bien que les sujets semblent très divers, peu d'entre eux peuvent être étudiés sans faire appel simultanément à plusieurs disciplines : traitement du signal, probabilités, informatique, phonétique, linguistique.... L'association s'emploie à rassembler les chercheurs de ces disciplines, à favoriser leurs échanges et à aider à la formation de jeunes chercheurs, par l'organisation de séminaires et par la publication d'ouvrages spécialisés. Une rencontre biennale "Les Journées d'Etudes sur la Parole" (JEP) permet une synthèse de l'ensemble des travaux de la communauté.

Plus généralement, pour atteindre ses objectifs, l'AFCP se donne, dans son domaine de compétences, les moyens suivants : stimulation et facilitation des échanges scientifiques ; organisation de conférences, de journées scientifiques et d'expositions ; organisation de cours, de séminaires et d'écoles thématiques ; publication et aide à la publication de revues et d'ouvrages ; diffusion de l'information à l'intention des acteurs institutionnels et du grand public ; promotion des relations entre recherche et industrie ; coopération avec d'autres organisations scientifiques ; ainsi que toute action conforme à ses objectifs.

Association pour la Recherche Cognitive

ARCo, Université de Technologie de Compiègne, COSTECH, BP 319 - 60203 Compiègne Cedex

Tél. : 03 44 23 43 66 Téléc. : 03 44 23 52 12

Site Internet : <http://www.utc.fr/arco/index.php>

Personne-ressource : Olivier Gapenne

Courriel : olivier.gapenne@utc.fr

Mission

L'Association pour la Recherche Cognitive (ARCo), créée en 1981, a pour but de promouvoir la recherche interdisciplinaire sur la cognition. Elle comprend un groupe de travail sur le traitement automatique de la sémantique des langues. Les membres de l'ARCo exercent dans des domaines qui se rattachent aux sciences humaines, aux sciences de l'ingénieur et aux sciences de la vie : psychologie cognitive, linguistique, intelligence artificielle, logique, neurosciences, sociologie, philosophie.

Depuis sa création, l'association anime activement les débats essentiels à la vie et à la cohérence d'une communauté interdisciplinaire comme la nôtre. Elle veille à maintenir un pluralisme des disciplines; des approches théoriques, expérimentales et appliquées; des écoles, paradigmes et thèmes. L'ARCo cherche à promouvoir l'intégration dans notre communauté des jeunes chercheurs et des chercheurs isolés; elle oeuvre également à l'éclosion d'une dimension européenne des sciences cognitives. Depuis quelques années, les institutions scientifiques ont reconnu les Sciences Cognitives avec des programmes de recherche et des DEA spécifiques. Dans ce paysage renouvelé, le rôle spécifique de l'ARCo, conformément à son statut d'association régie par la loi 1901, est d'être plus que jamais un lieu d'ouverture et de transparence tant au niveau de sa propre administration qu'à celui des débats scientifiques. L'ARCo regroupe aujourd'hui environ 350 chercheurs de la recherche publique ou privée.

Association pour le traitement automatique des langues

ATALA? 45, rue d'Ulm 75230 PARIS CEDEX 5. - SIRET: 393 902 721 00017 - CCP 13 809
69 J PARIS

Site Internet : <http://www.atala.org>

Personne-ressource : Jean Véronis

Courriel : atala@biomath.jussieu.fr

Mission

L'ATALA fait office d'organisme centralisateur visant l'avancement des technologies relatives au traitement automatique des langues.

L'ATALA publie depuis 1960 la revue internationale Traitement Automatique des Langues (anciennement La Traduction Automatique puis TA Informations), avec le concours du CNRS. Cette revue paraît maintenant trois fois par an. Elle est envoyée gratuitement aux membres de l'ATALA à jour de leur cotisation. Sa diffusion (hors adhérents) est assurée par Lavoisier, qui a repris les activités de HERMÈS Science.

La conférence TALN est née en 1994 dans le cadre du pôle Langage Naturel du GdR-PRC Communication Homme-Machine. Elle a été créée par Philippe Blache, qui s'est chargé de ses trois premières éditions et en a lancé la dynamique. Depuis 1997, l'organisation de TALN tourne chaque année. Depuis 1999, elle se déroule sous l'égide de l'ATALA. Un Comité permanent, créé en 1998, recueille les candidatures pour l'organisation des éditions successives de la conférence, et assure sa pérennité. Depuis 1999, la conférence RÉCITAL (Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues) se déroule sous la forme d'une session étudiante de TALN.

L'ATALA organise chaque année depuis 1960 plusieurs Journées d'Étude qui comprennent une série d'exposés, pouvant être éventuellement accompagnés de "démonstrations". Elles sont l'occasion de rencontres et de discussions entre différents "acteurs" du domaine (universitaires, industriels, chercheurs, étudiants). Chaque journée, placée sous la responsabilité d'un organisateur, est consacrée à un thème particulier touchant aux recherches théoriques ou aux applications en Linguistique, en Informatique et dans les divers domaines du traitement automatique des langues : Communication homme-machine, Dictionnaires électroniques, Indexation automatique, traduction automatique, Reconnaissance de la parole, synthèse vocale, etc.

European Language Resources Association (ELRA)

55-57 rue Brillat Savarin, 75013 Paris, France

Tél. : +33 1 43 13 33 33 Téléc. : +33 1 43 13 33 30

Site Internet : <http://www.elra.info/fr/index.html>

Personne-ressource : Khalid Choukri

Courriel : choukri@elda.fr

Mission

L'association européenne pour les ressources linguistiques, ELRA (European Language Resources Association) est la force motrice pour la mise à disposition des ressources linguistiques pour la communauté de l'ingénierie linguistique, et pour l'évaluation des technologies en ingénierie linguistique.

ELRA travaille ainsi activement à l'identification, la distribution, la collecte, la validation, la standardisation, la production et à l'amélioration constante des ressources linguistiques. ELRA apporte également son soutien aux campagnes d'évaluation, et au développement d'un domaine scientifique propre aux ressources linguistiques et à l'évaluation.

Ces différentes activités sont menées à bien par le corps opérationnel d'ELRA, l'agence pour la distribution des ressources linguistiques et pour l'évaluation, ELDA (Evaluations & Language resources Distribution Agency).

Les projets européens, internationaux et français auxquels ELDA participe abordent entre autres les aspects relatifs aux ressources linguistiques (la production de nouvelles ressources, la validation des ressources, ou la normalisation dans ce domaine), ainsi que la problématique liée à l'évaluation des technologies de la langue, à travers les campagnes d'évaluation en cours ou la mise en place de nouvelles plateformes d'évaluation.

ELDA est aussi très active au sein des projets visant à promouvoir le domaine des technologies de la langue, et récolte tous types d'informations ayant trait au marché.

Association Internationale de la communication parlée

Site Internet: <http://www.isca-speech.org>

Personne-ressource : Julia Hirschberg

Courriel : isca-president@isca-speech.org

Mission

ISCA est une Association à but non lucratif. L'objet de l'Association est de promouvoir le domaine de la science et de la technologie en communication parlée, dans un contexte international, à l'échelle mondiale. L'Association essaiera d'atteindre ses objectifs en

- Stimulant la recherche scientifique et l'éducation;
- Organisant des conférences et des expositions;
- Organisant des cours;
- Publiant et en encourageant la diffusion de traités scientifiques;
- Promouvant l'étude des différentes langues;
- Promouvant les échanges scientifiques;
- Collaborant avec d'autres organisations;
- Promouvant les relations entre recherche et industrie;
- Être au service des besoins locaux, régionaux et globaux.

Réseau Francophone de l'Ingénierie de la Langue (FRANCIL)

Francil, LIMSI-CNRS, BP133 ? 91403 Orsay Cedex, France

Tél. : (33) (1) 69-85-80-64/80-66 Téléc. : (33) (1) 69-85-80-88

Site Web : <http://www.limsi.fr/Recherche/FRANCIL/frcl.html>

Personne-ressource : Joseph Mariani

Courriel : mariani@limsi.fr

Nombre de postes :

Projets de recherche

L'Agence francophone pour l'enseignement supérieur et la recherche a mis en place un programme destiné à renforcer ses activités dans le domaine du Génie Linguistique. Le génie linguistique est une partie du traitement électronique de l'information. Dans ce contexte, le traitement automatique des langues est un secteur en plein développement. Il comprend les recherches et développement en matière d'analyse et de génération de textes, de reconnaissance, de compréhension et de synthèse de la parole. Il inclut les applications relatives à la gestion de documents, à la communication entre l'humain et la machine, à l'aide à la rédaction, à la traduction assistée par ordinateur. Il comporte des enjeux de type industriel et économique, de type scientifique et technologique mais présente aussi une dimension culturelle très spécifique. Il est toujours préférable de maîtriser la langue qui sert à exprimer les résultats des recherches, surtout si cette recherche porte sur la langue elle-même. La langue française et la francophonie sont donc très directement concernées par cet enjeu.

Le réseau FRANCIL a pour objectifs fondamentaux de :

Assurer la solidarité entre les chercheurs du monde francophone dans une perspective de co-développement ;

- Aider au désenclavement des chercheurs, des laboratoires et des équipes de recherche des pays francophones et en particulier des pays du Sud ;
- Favoriser la coopération entre les équipes, les laboratoires des universités, des grandes écoles et des grands centres de recherche, qu'ils soient nationaux, internationaux, publics ou privés ;
- Contribuer à la qualité et au rayonnement de la recherche francophone ;
- Favoriser la production scientifique en langue française ;
- Permettre, dans toute la mesure du possible, aux chercheurs s'exprimant en français mais n'appartenant pas à des pays francophones de collaborer avec leurs homologues francophones.

Portail Technolangue

ELDA (Agence pour l'évaluation et la distribution de ressources linguistiques) 55-57 rue Brilat-Savarin 75013 Paris

Tél. : 01 43 13 33 33 Téléc. : 01 43 13 33 30

Site Internet: <http://www.technolangue.net>

Personne-ressource : Valérie Mapelli, ELDA

Courriel : info@technolangue.net

Mission

L'action Technolangue concerne le domaine du traitement informatisé des langues (TIL).

Ce domaine, également connu sous le nom de "traitement automatique des langues" (TAL) ou "ingénierie linguistique", vise à fournir des technologies utilisées dans de nombreuses applications qui traitent l'information sous forme écrite ou orale, monolingue ou multilingue.

Comme exemples d'applications, on peut citer : la traduction automatique, l'analyse de documents, la recherche et le filtrage d'information, la dictée vocale, la synthèse de la parole, le dialogue homme/machine en langage naturel,...

Technolangue est une action commune aux trois réseaux de recherche et d'innovation technologique (RNRT, RNTL, RIAM), financée dans un cadre interministériel.

Son objectif principal est de mettre en place de manière pérenne une infrastructure de production et diffusion de ressources linguistiques, d'évaluation des technologies de la langue écrite et orale, de participation aux instances nationales et internationales de normalisation et de standardisation et de veille informationnelle sur le domaine.

Les ministères concernés sont le Ministère chargé de la Recherche et des Nouvelles Technologies, le Ministère délégué à l'Industrie et le Ministère de la Culture et de la Communication.

- Volet "Ressources linguistiques" visant à renforcer l'infrastructure permettant la production, la validation et la diffusion de ressources langagières ;
- Volet "Évaluation" visant à créer une infrastructure d'évaluation qui prenne en compte les technologies de traitement automatique de la langue ;
- Volet "Normes et standards" facilitant l'accès à la connaissance des normes et standards développés pour ce domaine et soutenant une participation française à leur constitution ;
- Volet "Veille" pour la mise en place d'une veille technologique permettant de connaître l'état de l'art des technologies de la langue et de leurs usages dans différentes applications, dans le contexte international.

Hong Kong

Laboratoires

Holiday Centre for Intelligent Applications of Language Studies

The Halliday Centre for Intelligent Applications of Language Studies, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

Tél. : (852) 2788 8722 Téléc. : (852) 2788 8706

Site Web : hallidaycentre.cityu.edu.hk/

Personne-ressource : Dr. Jonathan Webster

Courriel : ctjjw@cityu.edu.hk

Nombre de postes : 27

Projets de recherche

Modélisation métalinguistique – conception de nos propres sources métalinguistiques pour modeler le langage en contexte, avec une attention particulière à la conception de modèles détaillés plus explicites en mesure de favoriser de nouvelles mises en application computationnelle et de nouveaux outils computationnels.

Inde

Laboratoires

AU-KBC Research Center, Madras Institute of Technology
<p>The AU-KBC Research Centre, Madras Institute of Technology, Anna University, Chromepet, Chennai (Madras)-600 044, India.</p> <p>Tél. : (91) 044 2223 2711/4885 Téléc. : (91) 044 2223 1034</p> <p>Site Web : http://www.au-kbc.org/index.html</p> <p>Personne-ressource : M. Murali</p> <p>Courriel : mmurali@au-kbc.org</p> <p>Nombre de postes : 44</p>
<p>Projets de recherche</p>
<p>Le groupe de traitement du langage naturel met l'accent sur la conception d'outils, de technologies, de produits et de systèmes pour faciliter l'utilisation des ordinateurs et de l'internet dans la vie de tous les jours. Dans le contexte indien, le principal aspect de nos travaux consiste à trouver des moyens de rendre l'information en ligne accessible aux personnes qui ne parlent pas l'anglais. Le groupe travaille actuellement à créer les ressources lexiques telles que les dictionnaires, Wordnet et le corpora étiqueté, qui sont essentielles pour les chercheurs qui travaillent dans divers domaines du traitement du langage naturel. Les chercheurs en linguistique et en informatique sont des membres du traitement du langage naturel.</p> <p>Le groupe a plusieurs objectifs à long terme, notamment :</p> <ul style="list-style-type: none">- La conception de systèmes de traduction automatique entre le tamil et d'autres langues, particulièrement l'anglais et le hindi.- Créer des ressources lexiques en tamil qui sont essentielles pour les chercheurs et les développeurs.- Concevoir des outils de base pour les travaux computationnels en tamil, comme l'agent d'analyse de morphe, agent d'étiquetage du mot d'ancrage, etc.- Application d'outils de traitement du langage naturel pour l'extraction d'information à partir de textes propres aux domaines pour concevoir des systèmes d'extraction d'information pour divers domaines comme la médecine et l'agriculture.- Les réalisations du groupe en presque deux ans et demi depuis le début en janvier 2000.- Le système de traduction automatique entre le tamil et le hindi a un rendement d'un intervalle de 75 %.- Le groupe dispose d'un moteur de recherche en tamil, le premier en son genre dans le langage indien.- L'agent d'analyse de pointe du tamil peut manipuler près de 3,5 millions de formes lexicales, y compris les mots composés d'une précision de plus de 95 %.- Le module d'extraction de l'entité nommée fait partie de l'extraction d'information pour la littérature médicale.

Center for Development of Advanced Computing

Centre for Development of Advanced Computing Pune University Campus Ganesh Khind
Pune - 411 007 India.

Tél. : (91) 20 2570 4100 Téléc. : (91) 20 2569 4004

Site Web : <http://www.cdac.in/>

Personne-ressource : Shri S. Ramakrishnan

Projets de recherche

C-DAC a été un pionnier dans le développement et la prolifération de l'utilisation des langues indiennes dans les ordinateurs. Cette technologie est maintenant développée pour inclure les solutions de traitement multimédia et multilingues qui couvrent tout un éventail d'applications qui ont été commercialisées avec succès, entre autres la publication, l'impression et le traitement de mots, les suites d'applications de bureau avec des interfaces de langage pour les logiciels populaires de tierce sur diverses plates-formes de commande, les courriels, la traduction automatique, l'apprentissage de langue, la vidéo et le contenu de télévision et des multimédias dans les langues indiennes. Les deux groupes de C-DAC qui ont travaillé au développement de technologies multilingues sont :

Graphics and Intelligence based Script Technology (GIST) Le groupe GIST de CDAC facilite l'utilisation des langues indiennes dans la traduction automatique. Afin de rester à l'avant-garde des technologies sans le monde, le groupe a adopté les derniers concepts qui lui permettront de rester au courant du monde de l'internet.

Applied Artificial Intelligence (AAI)

Le groupe AAI Group de C-DAC est engagé dans plusieurs activités telles que les systèmes de compréhension axés sur les connaissances, le système de traduction automatique, l'extraction et la recherche d'information, les tuteurs du langage, solutions accessibles sur le Web et plusieurs autres solutions de développements logiciels stratégiques.

Computer Vision and Pattern Recognition Unit, Indian Statistical Institute of Calcutta

COMPUTER VISION AND PATTERN RECOGNITION UNIT INDIAN STATISTICAL INSTITUTE? 203, B. T. ROAD, Calcutta – 700035 INDIA

Tél. : (91) 33 577 7694 Téléc. : (91) 33 577 6680, 3035

Site Web : <http://www.isical.ac.in/~cvpr/>

Personne-ressource : Bidyut Baran Chaudhuri

Courriel : bbc@isical.ac.in

Nombre de postes : 20

Projets de recherche

Les principaux domaines de recherche actuels de l'unité comprennent la reconnaissance du caractère optique, le traitement du langage naturel, la reconnaissance de formes, le traitement d'images et les outils de calcul souple.

L'unité **CVPR** est fière d'être le premier à avoir conçu un système de reconnaissance des **caractères optiques** (OCR) pour la reconnaissance de scripts combinés du bangla et le devnagari (hindi), les scripts les plus populaires de l'Inde. Le système effectue la correction verticale de documents, ligne de script, segmentation des mots et des caractères, la reconnaissance des caractères et enfin, la correction d'erreurs. Cette unité a pour principal domaine d'intérêt le problème général de l'analyse des documents numériques. Les nouveaux algorithmes d'estimation et de correction des effets d'obliquité ont été conçus, sans oublier la conception d'un système de repérage et de calcul automatiques des formules mathématiques dans les documents. Les questions relatives à l'analyse de documents imprimés et manuscrits (incluant la segmentation, l'extraction d'information et d'éléments, la compression de données, le traitement et la compréhension des dessins au trait) sont en cours d'étude alors qu'un système d'extraction automatique de méta-information à partir d'images a été déjà mis au point.

Le **traitement des langages naturels** des langues indiennes a connu des développements pilotes. On a mis au point et développé un puissant correcteur d'orthographe qui corrigeait non seulement les fautes d'orthographe dans les documents enregistrés dans un ordinateur de la langue bangla, mais aussi dans les documents en langue hindi. Parmi d'autres importants développements, mentionnons un dictionnaire vocal et un éditeur parlant pour la langue bangla, un processeur morphologique du bangla et un synthétiseur informatisé de la parole en bangla. L'un des nouveaux objectifs consiste à concevoir des idées informatiques pour les personnes invalides, telles que les malentendants, les personnes muettes et les non voyants. Parmi d'autres études, mentionnons l'analyse statistique d'un large corpus de texte en bangla, l'analyse des formes d'erreurs humaines de prononciation, ainsi que l'analyse d'omatonopées bangla. Les travaux d'élaboration d'un **dictionnaire informatisée trilingue** (hindi, bangla et anglais) sont en pleine ébullition.

Irlande

Laboratoires

Digital Signal Processing, University College Dublin
Digital Signal Processing Research, Department of Electronic and Electrical Engineering, Engineering Building, Belfield, Dublin 4 Ireland. Tél. : (353) 1 7161909 Téléc. : (353) 1 2830921 Site Web : http://dsp.ucd.ie/ Personne-ressource : Anthony Fagan Courriel : tony.fagan@ucd.ie Nombre de postes : 33
Projets de recherche
<ul style="list-style-type: none">• Annulation de la diaphonie pour les communications haute vitesse de données sur du cuivre par paires torsadées;• Vérification et identification audio-visuelle;• Diagnostic de pathologie de la corde vocale au moyen d'analyse multimodale des signaux;• Reconnaissance audio-visuelle automatique de la parole.

Information Management Group, Dublin College University

School of Computing, Dublin City University, Glasnevin, Dublin 9

Tél. : (353) 1 700 5984

Site Web : <http://www.computing.dcu.ie/research/information.html>

Personne-ressource : Stephen Blott

Courriel : Stephen.Blott@computing.dcu.ie

Nombre de postes : 29

Projets de recherche

L'Information Management Group compte trois principaux thèmes de recherche : Multimédia numérique, ingénierie de bases de données et systèmes interopérables, et technologies éducatives.

Le Centre for Digital Video Processing (CDVP) étudie et développe les techniques ainsi que les outils visant à analyser et indexer automatiquement l'information vidéo numérique, ainsi qu'à réaliser des opérations à base de contenu et les que le parcours, la recherche et le résumé. Les chercheurs en ingénierie de bases de données et en systèmes interopérables conçoivent des modèles formels et informels pour élaborer des bases de données et des couches interopérables entre les systèmes hétérogènes d'information. La recherche en ingénierie de systèmes d'information met l'accent sur des modèles et des méthodes pour augmenter la valeur des données dans les économies réseautées et axées sur l'information. La recherche en technologies éducative vise actuellement à améliorer l'utilisation d'ICT dans l'enseignement supérieur par la recherche-action.

Knowledge and Data Engineering Group, Trinity College Dublin

KDEG Department of Computer Science Trinity College Dublin

Tél. : (353) 1 6082091

Site Web : https://www.cs.tcd.ie/research_groups/kdeg/index.php

Personne-ressource : Vincent P. Wade

Courriel : Vincent.Wade@cs.tcd.ie

Nombre de postes : 31

Projets de recherche

Le groupe met l'accent sur la recherche et les applications relatives à la gestion des connaissances et des données. Le groupe étudie plus particulièrement les systèmes adaptatifs (hypermédia), la gestion intelligente des communications et du traitement ubiquiste, le Web sémantique et l'interopérabilité, l'orchestration Web et l'automatisation du processus opérationnel.

Particulièrement, le KDEG a été le pionnier de la recherche dans quatre domaines d'application :

- Gestion des services de télécommunications;
- Gestion de la formule Smart Space (calcul ubiquiste);
- Hypermédia adaptative et apprentissage en ligne;
- Informatique de la santé.

Language and Intelligence Group, Dublin City University

School of Computing, Dublin City University, Glasnevin, Dublin 9

Tél. : (353) 1 700 8929

Site Web : <http://www.computing.dcu.ie/research/language.html>

Personne-ressource : Dónal Fitzpatrick

Courriel : Dónal.Fitzpatrick@computing.dcu.ie

Nombre de postes :

Projets de recherche

Traitement de la parole et du langage, y compris la traduction automatique, le traitement de la parole, les modèles computationnels de la sémantique, treebanks, syntaxe formelle, traitement numérique des signaux, apprentissage de langage assisté par ordinateur, passage et traitement probabilistiques du langage naturel.

Intelligence artificielle, y compris Artificial Minds, les modèles computationnels de cognition, la représentation des connaissances, l'interaction homme-machine, science cognitive, les origines de l'intelligence, les réseaux neuronaux et les agents autonomes.

National Center for Language Technologies

National Centre for Language Technology (NCLT), School of Computing, Dublin City University, Dublin 9, Ireland.

Tél. : (353) - 1 700 5074 Téléc. : (353) 1 700 5442

Site Web : <http://www.computing.dcu.ie/research/nclt/index.html>

Personne-ressource : Prof. Josef van Genabith

Courriel : nclt@computing.dcu.ie

Nombre de postes : 20

Projets de recherche

Traduction automatique statistique et à base de règles (SMT, RBMT), Traduction automatique axée sur des exemples (EBMT), mémoires de traduction (TMs), renforcement des systèmes existants de traduction automatique, traduction assistée par machine (MAT), traduction assistée par ordinateur (CAT), langages contrôlés.

Intégrer la technologie CL/NLP/HLT dans CALL, CALL pour les langages menacés de disparaître, CALL for Remedial Learners.

Traduction automatique statistique et à base de règles (SMT, RBMT), Traduction automatique axée sur des exemples (EBMT), mémoires de traduction (TMs), renforcement des systèmes existants de traduction automatique, traduction assistée par machine (MAT), traduction assistée par ordinateur (CAT), langages contrôlés.

Algorithmes automatiques d'annotation forme-structure, extraction de trame de sous-catégorisation, acquisition d'envergure

Automatic Feature-Structure Annotation Algorithms, Subcategorisation Frame Extraction, Wide-Coverage Robust Probabilistic Unification Grammar Acquisition, PCFG-Based LFG Approximation, HPSG Acquisition, Multilingual Treebank-Based Grammar Acquisition.

Théorie de la représentation du discours, sémantique fondée sur la logique linéaire, traitement de formes logiques à partir de Treebanks, Systèmes de réponse aux questions à domaine ouvert.

Caractérisation du locuteur, classification audio, recherche et programmation, interfaces homme-machine.

Recherche d'information multilingue.

TCD Computational Linguistics Group

Computational Linguistics Lab, O'Reilly Institute G27, Department of Computer Science, Trinity College, University of Dublin, Dublin 2 Ireland.

Tél. : (353) 1 608 2866 Téléc. : (353) 1 677 2204

Site Web : <https://www.cs.tcd.ie/>

Personne-ressource : Carl Vogel

Courriel : carl.vogel@tcd.ie

Nombre de postes : 40

Projets de recherche

Le Computational Linguistics Group met l'accent sur le langage et le raisonnement. Il s'intéresse à la compréhension et la modélisation des capacités cognitives humaines, ainsi qu'à l'ingénierie d'applications qui peuvent être réalisées en tant que fruit de la recherche scientifique et de la philosophie du langage.

Israël

Laboratoires

Department of Computer Science, Bar-Ilan University
Department of Computer Science, Bar-Ilan University, Ramat Gan, 52900, Israel. Tél. : (972) 3 531 8866 Téléc. : (972) 3 736 0498 Site Web : http://www.cs.biu.ac.il/ Personne-ressource : Amihood Amir Courriel : amir@cs.biu.ac.il Nombre de postes : 41
Projets de recherche
Apprentissage par machine, théorie de l'autoorganisation, classification de textes. Exploration de données, exploration de textes, apprentissage par machine et analyse des liens. Harmonisation multidimensionnelle des formes telles qu'elles sont appliquées au traitement de textes, vision informatisée et multimédia, biologie computationnelle.

Department of Computer Science, University of Haifa

Department of Computer Science, University of Haifa, Mount Carmel, 31905 Haifa, Israël

Tél. : (972) 4 824 0259 Téléc. : (972) 4 824 9331

Site Web : <http://cs.haifa.ac.il/welcome.html>

Personne-ressource : Daphna Stern

Courriel : daphna@cs.haifa.ac.il

Nombre de postes : 44

Projets de recherche

Les domaines de recherche du département comprennent les algorithmes, la complexité computationnelle, les mathématiques discrètes, l'analyse numérique, les approximations, le calcul parallèle, les systèmes d'exploitation, les compilateurs, les graphiques informatisés, la vision artificielle, le traitement d'images, la vision en couleurs, CAD, les algorithmes de graphe, l'intelligence artificielle, les réseaux neuronaux, la biologie computationnelle, la linguistique computationnelle et le traitement du langage naturel.

Israël Institute of Technology, Computational Linguistics Group

Israel Institute of Technology, Haifa 32000, Israel

Tél. : (972) 4 829 3379

Site Web : <http://www.cs.technion.ac.il/~lcl/>

Personne-ressource : Ran El-Yaniv

Courriel : rani@cs.technion.ac.il

Nombre de postes : 35

Projets de recherche

Probabilistic Morphological Analyzer for Hebrew Undotted Texts L'analyse morphologique des mots dans un texte constitue la première étape des plus naturelles applications linguistiques. En raison de la riche morphologie de l'hébreu et de l'impertinence du script sans pointillé qui cause une grande ambiguïté morphologique, le problème n'a toujours pas trouvé de solution satisfaisante. Nous remarquons que le problème de l'analyse morphologique des textes hébreux est similaire au problème bien étudié qu'est l'étiquetage du mot d'ancrage en anglais. Nous avons ainsi appliqué certaines approches utilisées pour résoudre ce problème d'analyse morphologique. Les travaux ont abouti à une analyse morphologique exacte d'environ 96 % des mots. Ce résultat ressemble à ceux qui ont été observés dans l'étiquetage probabilistique de mot d'ancrage en anglais, en utilisant un très petit corpus de formation – 500 mots seulement par rapport à un corpora d'un million de mots utilisés pour l'étiquetage en anglais.

Corpus Based Analysis of Hebrew Le projet vise à étudier les moyens de concevoir de meilleurs outils pour l'analyse morphologique et syntaxique de l'hébreu en utilisant des techniques basées sur le corpus. Le projet fait partie d'une initiative conjointe avec le groupe NLP de l'Institut d'informatique de l'Université Hebrew, Jérusalem, présidé par le professeur Eli Shamir.

Extensions and Implementations of Natural Logic Ce projet consiste à élaborer un système interface pour le langage naturel, qui est fondé sur la logique naturelle : une logique qui opère directement sur des représentations syntaxiques sans traduction intermédiaire vers un formalisme logique plus connu – expressions monotones dans la grammaire catégorielle.

Semantics of Natural Language Temporal Questions and Interfaces to Temporal Database Systems Nous travaillons sur une interface de langage naturelle pour les bases de données temporelles, à partir d'un traitement sémantique des questions temporelles. En utilisant l'interface, l'utilisateur peut exprimer des questions dans le langage naturel qui seront traduites automatiquement et soumises à la base de données. Le projet est financé par FIRST, administré par l'académie des sciences d'Israël.

Knowledge Center for Processing Hebrew

KCPH, Taub Building, Technion Israel Institute of Technology, Haïfa 32000

Tel: (972) 4 829 4313 Téléc. : (972) 4 829 3900

Site Web : <http://www.mila.cs.technion.ac.il/>

Personne-ressource : Arnon Cohen

Courriel : mila@cs.technion.ac.il

Nombre de postes : 20

Projets de recherche

Les activités de recherche et de développement du centre sont entreprises par la plupart des chercheurs en linguistique computationnelle d'Israël, en plus de plusieurs autres chercheurs de domaines d'études connexes. Nous visons à établir une infrastructure computationnelle pour traiter l'hébreu dans diverses frontières :

- Un grand dépôt de textes hébreux modernes qui est étiqueté sur les plans morphologique et syntaxique.
- Des normes d'étiquetage et de présentation de textes annotés en hébreu.
- Des réseaux sémantiques multilingues de mots.
- Des systèmes d'analyse de la syntaxe et la morphologie à partir de règles linguistiques et de systèmes basés sur des modèles statistiques.
- Des méthodes d'évaluation du rendement des systèmes de traitement du langage naturel, à savoir l'hébreu.
- Un logiciel de classement des textes hébreux.
- Grammaires computationnelles qui fourniront la fondation de l'analyse syntaxique.

Les étudiants qui font de la recherche prendront également part à la recherche et au développement. En outre, les ressources qui seront développées dans le centre seront mises à la disposition des instituts universitaires d'Israël aux fins d'enseignement et de recherche. Ainsi, le centre servira également de catalyseur dans l'élaboration de programmes pédagogiques qui permettront de former plus de personnel dans le domaine de la linguistique computationnelle en Israël.

Italie

Laboratoires

Center for Scientific and Technological Research, Cognitive and Communication Technologies Division

ITC-irst Via Sommarive, 18 38050 Povo (Trento) Italy

Tél. : (39) 0461 314 444 Téléc. : (39) 0461 302 040

Site Web : <http://www.itc.it/irst/Renderer.aspx?targetID=164>

Personne-ressource : Fabio Pianesi

Courriel : pianesi@itc.it

Nombre de postes : 20

Projets de recherche

La Cognitive and Communication Technologies Division mène la recherche sur le traitement linguistique afin d'élaborer une interface et des technologies de connaissances avancées. Sont en étude des interfaces axées sur l'utilisateur qui sont multilingues et multimodales et des technologies sémantiques de connaissances basées sur le contexte.

Objectifs de la recherche

Le TCC a pour objectif de concevoir un modèle idéal simplifié dans lequel l'information enregistrée en format linguistique est obtenue, explorée et mise à la disposition de la population. Pour atteindre cet objectif, il y a besoin d'étude et de développement de technologies pour :

- une interaction homme-homme et homme-machine. La recherche met l'accent sur la multimodalité et le multilinguisme, ainsi que l'interaction avec des environnements virtuels et physiques.
- une représentation d'information flexible pour les utilisateurs finaux. Cette représentation combine la contribution de différents médias, l'adaptation aux caractéristiques des utilisateurs et le caractère changeant de l'interaction.

L'analyse automatique de textes tels que les documents, les courriels et les pages Web, ainsi que l'extraction automatique de l'information à partir de ces textes.

Cognitive and Communication Technologies Division, ITC-Irst

ITC-irst Via Sommarive, 18 ? 38050 Povo (Trento) Italy

Tél. : (39) 0461 314 444 Téléc. : (39) 0461 302 040

Site Web : <http://tcc.itc.it/>

Personne-ressource : Fabio Pianesi

Courriel : pianesi@itc.it

Nombre de postes : 47

Projets de recherche

La Cognitive and Communication Technologies (TCC) division de l'ITC-Irst est un groupe de recherche européen réputé dans des domaines tels que le traitement du langage naturel, l'interaction homme-machine et les systèmes de dialogue, la multimodalité et la génération du langage naturel, la production et la maintenance des ressources linguistiques, et enfin la théorie linguistique. Le TCC fait partie de l'ICT-Irst et est membre du European Network of Excellence in Natural Language and Speech (ELSNET).

À l'heure actuelle, nos activités sont organisées en deux principaux domaines :

- TEXTEC (Technologies for Text Processing) porte sur les questions de base et appliquées dans le traitement de textes : réponse aux questions, désambiguïsation du sens des mots et extraction d'information. Cette recherche inclut également la construction et la maintenance de ressources linguistiques : lexique et wordnets, grammaires, corpora, etc.
- i3p (Intelligent and Interactive Information Presentation) concerne l'étude et la conception de systèmes pour présenter l'information aux utilisateurs, en mettant l'accent sur des aspects comme l'adaptation et l'appui de l'interaction entre le langage et d'autres formes (images, vidéo, animations, etc.). Les scénarios varient, d'une pr.sentation d'information sur le Web à une présentation mobile. La TCC est également active dans le domaine de la traduction automatique, particulièrement la traduction vocale.

TCC se consacre à la recherche de base et appliquée et possède des liens étroits avec plusieurs grands centres de recherche de partout dans le monde. La division a entretient des relations régulières et fructueuses avec les entreprises privées, particulièrement celles qui œuvrent dans le monde Web, les télécommunications et les publications multimédias. De plus, la TCC compte une participation de longue date aux projets financés par l'Union européenne.

Department of computer and systems science, University of Rome

Via Eudossiana 18 00184 Roma

Tél. : (39) 06 44 585360 Téléc. : (39) 06 44 585367

Site Web : <http://www.dis.uniroma1.it/research.html>

Personne-ressource : Tiziana Toni

Courriel : abhaya@ics.mq.edu.au

Nombre de postes : 127

Projets de recherche

L'Integrative Modelling, Simulation and Data Analysis group a été établi à DIS en 2004. La recherche puise dans les méthodologies provenant de diverses disciplines telles que la modélisation dynamique à grande échelle, la simulation et l'optimisation. Les principaux domaines de recherche sont : systèmes intégrés et réactifs en temps réel dans la gestion automatisée, l'optimisation dans la biologie des systèmes, le traitement d'images et la vision artificielle, les algorithmes complexes et l'exploration de données.

DAMSO : Le laboratoire d'analyse de données, de modélisation intégrative, de simulation et d'optimisation vise à élaborer des modèles et à tester des algorithmes efficaces de traitement de données en temps réel à partir de l'ingénierie industrielle et biosystémique. Responsable du laboratoire : le professeur Alberto De Santis. Lieu : DIS-Buonarroti.

Systèmes intégrés et réactifs en temps réel dans la gestion : Les systèmes intégrés sont des éléments électroniques intégrés dans une installation physique. Les systèmes intégrés les plus problématiques, soit les systèmes réactifs en temps réel, contrôlent le comportement de l'installation pour que les modèles et les techniques de contrôle ainsi que les outils de simulation soient nécessaires lorsque les algorithmes sont montés sur une architecture HW-SW.

Optimisation computationnelle dans la biologie des systèmes : L'optimisation computationnelle dans la biologie des systèmes est un programme de recherche qui relie les biologistes et les ingénieurs dans une approche multidisciplinaire de l'analyse systématique à grande échelle et du phénomène biologique dynamique et complexe. La recherche porte particulièrement sur l'analyse des propriétés dynamiques des gènes, des métabolites et des protéines qui interagissent dans la cellule.

Traitement de signaux et vision artificielle : le traitement numérique de signaux porte sur les aspects théoriques et pratiques de la représentation et l'extraction d'information à partir de données réelles. Il a d'importantes répercussions sur la plupart des domaines scientifiques et techniques des sciences de la gestion. Les applications actuelles d'intérêt du groupe de recherche concernent l'ingénierie industrielle, l'interaction homme-machine, la technologie médicale, la science de matériels et le diagnostic de pannes des structures mécaniques.

- *Robotique cognitive et raisonnement relatif aux actions* : c'est-à-dire la définition de méthodes et de techniques pour raisonner sur les actions et pour concevoir des agents incorporés (particulièrement les robots mobiles) capables de réaliser des tâches complexes

dans des environnements réels.

Department of Information Engineering, University of Padova

Via Gradenigo 6/B, 35131, Padova, Italy

Tél. : (39) 49 827-7601 Téléc. : (39) 49-827 7699

Site Web : <http://www.dei.unipd.it/wdyn/?IDsezione=2>

Personne-ressource : Paolo Tenti

Courriel : paolo.tenti@dei.unipd.it

Nombre de postes : 68

Projets de recherche

Recherche d'information : La recherche d'information (RI) est la science qui porte sur la réunion, l'indexation, l'organisation et la recherche efficaces et efficientes à partir de milliards de collections de documents de toutes les données multimédias ou uniquement des multimédias qui répondent aux besoins en information de millions d'utilisateurs. Les activités de recherche dans ce contexte sont comme suit :

- La recherche d'information contenue dans les documents éparpillés dans les réseaux : ces derniers fournissent l'accès aux collections de données structurées et non structurées. Dans le dernier cas, si un service de RI est accessible sur un réseau P2P, un utilisateur peut chercher les documents pertinents stockés dans des homologues connectés à son propre ordinateur, téléphone mobile ou PDA. Un modèle de pondération est étudié puis proposé pour réduire l'exploration du réseau et optimiser la recherche.
- L'évaluation de la recherche d'information interlingue : Il est question dans ce programme d'étudier la radicalisation dans un contexte multilingue. En raison des différentes langues des collections, la radicalisation nécessite des travaux plus complexes qu'un contexte classique monolingue. Deux méthodes de radicalisation du langage, l'une basée sur les modèles de Markov caché et l'autre, sur l'analyse des liens, ont été élaborées pour concevoir des radicaliseurs sans connaissances linguistiques a priori sur la morphologie du langage, en les inférant directement à partir des collections de données.
- La production, l'analyse et l'exploration automatiques des liens : Comme les connaissances ne se trouvent pas que dans les entités, mais également entre elles, les activités portent sur la transformation d'une série de documents en un hypertexte, l'addition de liens Web au contenu sémantique, la fourniture d'un mode de recherche d'information efficace et l'exploration de la topologie graphique à partir de sites Web. Les algorithmes de production et d'analyse des liens ont été conçus et testés.

Intelligence artificielle : La recherche met l'accent sur deux domaines précis de l'intelligence artificielle, notamment le raisonnement temporel et le traitement du langage naturel. En ce qui concerne le raisonnement temporel, la recherche vise à l'élaboration de systèmes automatiques capables de représenter et de raisonner en matière de connaissances temporelles en cas d'incertitude et d'imprécision. De nouveaux modèles d'intégration de l'information temporelle quantitative et qualitative, touchés par l'incertitude et l'imprécision, font l'objet d'étude au moyen du paradigme FCSP (Fuzzy Constraint Satisfaction Problem). Les systèmes réalisés sont appliqués aux applications de planification et de programmation ainsi qu'au domaine médical pour la reconnaissance des maladies à partir des formes temporelles des symptômes d'un patient. Dans le domaine du traitement du langage naturel, la recherche porte sur la conception de modèle statistique et la conception d'algorithmes spécialisés pour l'analyse syntaxique automatique du langage naturel.

European Academy Bolzano, Institute for Specialised Communication and Multilingualism

EURAC research Viale Druso 1, 39100 Bozen/Bolzano - Italy

Tél. : (39) 0471 055 055 Téléc. : (39) 0471 055 059

Site Web : <http://www.eurac.edu/Org/LanguageLaw/Multilingualism/index.htm>

Personne-ressource : Marcello Soffritti

Courriel : msoffritti@eurac.edu

Nombre de postes : 22

Projets de recherche

Depuis sa création en 1993, le département de recherche a élargi son étendue originale de « Langage et loi » pour couvrir graduellement le multilinguisme et la communication spécialisée. Ainsi, le département a été en mesure de se placer fermement au sein de la communauté internationale dans le domaine de la linguistique appliquée. Les activités précédentes étaient consacrées à la recherche sur la terminologie, la traduction et les langages spéciaux en vue de créer une base scientifique pour le multilinguisme dans les domaines juridique et administratif du Sud-Tyrol. Les divers projets du département peuvent être divisés en trois principaux domaines de recherche :

Bilinguisme et multilinguisme : Une réalité telle que la population du Sud-Tyrol, qui se distingue par ses trois groupes linguistiques et culturels, offre un panorama d'un grand intérêt pour l'étude des phénomènes associés à l'apprentissage du langage et personnes-ressources. Quelles sont les difficultés associées à l'apprentissage de plus d'une langue étrangère? Pourquoi certaines personnes apprennent-elles une langue étrangère plus facilement que d'autres? Comment les nouvelles technologies peuvent-elles favoriser l'apprentissage linguistique? Il s'agit là de quelques-unes des questions fondamentales relatives au « bilinguisme et au multilinguisme » dans le cadre de projets incluant l'analyse de la compétence linguistique et la conception de nouveau matériel d'apprentissage de l'allemand et de l'italien. En outre, l'interaction avec la linguistique computationnelle apporte une contribution importante par l'application de nouvelles technologies à l'apprentissage de langues et la création de matériel personnalisé et convivial.

Communication spécialisée : Outre les projets conventionnels consacrés à la terminologie juridique et administrative en italien et en allemand, le domaine de recherche sur la « communication spécialisée » met l'accent sur les nouveaux secteurs et les nouvelles méthodologies, y compris la formation de deuxième cycle en traduction, l'élaboration d'une terminologie juridique et administrative en ladin et les travaux terminologiques sur l'environnement et la planification spatiale en français, en italien, en allemand et en slovène. En plus de bénéficier des études actuelles dans la linguistique de corpus, la recherche est complétée par des ressources offertes par la linguistique computationnelle, intégrant ainsi de nouvelles méthodes de collecte et d'analyse des textes, d'extraction de la terminologie et de représentation de concepts.

Technologies de langage : L'apprentissage de langue assisté par ordinateur est un domaine de recherche multidisciplinaire qui a recours à la technologie informatique pour apprendre une langue. La linguistique computationnelle exploite les algorithmes et les ordinateurs afin de traiter les phénomènes linguistiques dans les grands corpora de textes. Les techniques modernes telles que la modélisation et l'adaptation utilisateur sont appliquées à nos projets afin de personnaliser l'accès et de faciliter les processus d'orientation et d'apprentissage.

Institute of Cognitive Sciences and Technologies

ISTC, V.le.Marx 00137 Rome Italy

Tél. : (39) 06 860 90 260 Téléc. : (39) 06 82 47 37 15

Site Web : <http://www.istc.cnr.it/>

Personne-ressource : Albano Leoni

Courriel : leoni@ip.rm.cnr.it

Nombre de postes : 92

Projets de recherche

L'Institute of Cognitive Sciences and Technologies (ISTC) est le résultat d'une fusion entre diverses institutions telles que l'ancien institut de psychologie, l'ancien institut de phonétique et de dialectologie de Palova et certains groupes des technologies biomédicales de Rome, LADSEB à Padova et du groupe Solid State Electronics de Rome. Il compte ainsi plusieurs groupes de recherche dont plusieurs étudient les technologies langagières, que ce soit le traitement des signaux ou l'exploration de données.

Les groupes de recherche qui composent l'ICST sont :

Cognitive Processes and New Technology Group, Division of Artificial Intelligence Cognitive Modelling and Interaction, Environmental Psychology Unit, Gesture and Language Laboratory (GaLL), Group for the Psychological Study of Interpersonal Relations, Human Development and Society Research Group, Laboratory on Agent-Based Social Simulation (LABSS), Laboratory for Applied Ontology (LOA), Laboratory of Artificial Life and Robotics - Stefano Nolfi, Language Development and Disorders Laboratory (LaDD), Neuropsychology of Language and Deafness Unit (NLS) - Virginia Volterra, Planning and Scheduling Team (PST), Sign Language Laboratory (SLL), Speech and Multimodal Communication Laboratory, T-cube: Trust, Theory and Technology Group, Unit of Cognitive Primatology and Primate Center, Laboratorio di Psicologia della Partecipazione Infantile.

Instituto di Elaborazione della Informazione

Pisa Research Area, Via Moruzzi, 1, 56124 , PISA

Tél. : (39) 050 3152876 Téléc. : (39) 050 3152810

Site Web : <http://www.iei.pi.cnr.it/>

Personne-ressource : Piero Maestrini

Courriel : maestrini@iei.pi.cnr.it

Nombre de postes : 120

Projets de recherche

L'Institut mène des recherches en informatique, en technologie de l'information et dans des domaines d'application connexes. Les activités de recherche portent particulièrement sur les secteurs ci-dessous. La plupart de activités sont réalisées dans le cadre de projets de recherche nationaux et internationaux. Les projets relatifs aux technologies du langage portent sur les domaines suivants :

- Modélisation de la recherche d'information multimédia (y compris les textes, les images et les données vidéo) à partir de modèles terminologiques logiques et probabilistiques;
- La conception de dépôts multimédias de haute performance, notamment les méthodes d'accès à des hiérarchies de stockage à plusieurs niveaux et des méthodes d'accès et d'interrogation multilingues dans des bases d'information multimédias;
- Méthodes et outils formels de spécifications pour élaborer des applications de bases de données;
- Interopérabilité et métadonnées pour des réseaux de bibliothèques.

Interactive Sensory System Division, SSI

Trento, Italy

Tél. : (39)-0461-314566 Téléc. : (39)-0461-314591

Site Web : <http://ssi.itc.it/home.htm>

Personne-ressource : Gianni Lazzari

Courriel : lazzari@itc.it

Nombre de postes : 56

Projets de recherche

SSI vise à étudier de nouvelles méthodes et à appliquer des systèmes d'acquisition, de traitement et d'organisation d'information sensorielle et environnementale. Les défis à relever sont les nouvelles applications fondées sur la technologie du langage parlé, la vision, la prédiction et la modélisation de logiciels.

CASPER – Comprehensive Automatic SPEech Recognition

Cette recherche a pour objectifs d'améliorer la technologie en matière d'extraction de l'information linguistique à partir de données audio et à faire face aux questions qui font de la reconnaissance automatique de la parole dans des domaines sans contrainte un problème entier. Une attention particulière est accordée aux questions relatives à la robustesse, la généralité et l'efficacité des systèmes de transcription automatique.

Projects FAME – Facilitating Agents in Multicultural Exchange
TC-STAR – Technologie et corpora pour la traduction vocale
PF-STAR : Preparing Future Speech Translation Research (projet de l'Union européenne)
PEACH - Personal Experience with Active Cultural Heritage
WebFAQ -Web Flexible Access and Quality

DITELO – Dialogue over the TELEphOne

Dialogo su TELEfonO sont liés à des agents de reconnaissance de la parole dans les applications téléphoniques dans le but principal de les inclure entre autres dans les systèmes de centres d'appels. Cela signifie la conception de modèles acoustiques et linguistiques et la proposition de capacités de dialogue. **Projets CoralRom** – corpora de référence intégré pour les langues romanes
Homey – surveillance à domicile par des systèmes de dialogue intelligents
WILMA – projet Wireless Internet and Location Management Architecture
PF-STAR - Preparing Future Speech Translation Research (projet de l'Union européenne).

HERMES – Traitement interlingue de langage

La recherche HERMES vise à faire avancer et à développer les technologies de traduction automatique dans les domaines d'application suivants : système de langage parlé, traduction de documents, accès à l'information. Le **projet TC-STAR** – Technologie et corpora pour la traduction vocale
PF-STAR - Preparing Future Speech Translation Research
FAME - Facilitating Agent for Multicultural Exchange
WebFAQ – Web : Flexible Access and Quality

SHINE – Interaction vocale dans un environnement bruyant

Cette recherche vise à étudier la robustesse dans la reconnaissance de la parole et l'analyse acoustique.

À l'heure actuelle, les principales activités portent sur la conception de systèmes d'interaction conversationnelle de vive voix dans une voiture, ainsi que l'interaction à distance dans un environnement bruyant et réverbérant.

ISTI, Knowledge Discovery and Delivery Laboratory

ISTI • Area della ricerca CNR, via G. Moruzzi 1, 56124 PISA, Italy

Tél. : (39) 050 3152878 Téléc. : (39) 050 3152811

Site Web : <http://www.isti.cnr.it/ResearchUnits/Labs/kdd-lab/>

Personne-ressource : Fosca Gianotti

Courriel : Fosca.Giannotti@isti.cnr.it

Nombre de postes : 27

Projets de recherche

Le laboratoire a pour objectif principal de concevoir de nouvelles méthodes computationnelles pour l'analyse de séries de données volumineuses et compliquées, c.-à-d. concevoir des méthodes qui aident à extraire les connaissances à partir de données et à les présenter aux utilisateurs finaux. Les méthodes de construction des connaissances sont fondées sur des approches algorithmiques et statistiques, des techniques de bases de données et des méthodes d'apprentissage automatique. La recherche de l'unité peut être vue comme une combinaison de cinq domaines de recherche :

Knowledge Discovery Support Environment - Data Mining Query Languages.

La plupart des applications d'analyse de données à forte intensité de connaissances nécessitent la combinaison de deux types d'activités : l'acquisition de connaissances et le raisonnement relatif aux connaissances acquises selon les règles du domaine. Les techniques d'exploration de données sont une réponse à la première question, puisqu'elles permettent d'extraire l'information à partir de connaissances implicites de données brutes à un niveau d'abstraction élevé. Nous avons besoin d'environnements d'appui de la découverte de connaissances et de langues d'interrogation et d'exploration de données capables d'intégrer l'extraction de connaissances et la manipulation de connaissances, afin de développer des applications d'analyse verticale des données qui intègrent la technologie d'exploration des données. Le laboratoire explore dans ce contexte plusieurs orientations :

LDL-Mine est un langage de base de données logique (à base de règles, compatible avec SQL) qui facilite l'intégration de l'exploration et de l'interrogation en permettant un traitement des connaissances extraites qui concorde avec les connaissances propres au domaine.

Frequent Pattern Discovery in Databases. Cette recherche vise la découverte de formes fréquentes. L'objectif comporte deux volets : étudier la façon de concevoir une primitive très expressive pour la découverte de formes fréquentes dans les bases de données qui peut être ajoutée à un langage d'interrogation relationnel; étudier les techniques d'optimisation qui rendent la primitive suffisamment efficace pour l'intégrer dans un système de gestion de bases de données.

KDDML-MQL. MQL est un langage algébrique comme SQL pour la précision des problèmes d'extraction des connaissances. Il favorise la combinaison de plusieurs formes de connaissances grâce à des opérateurs intégrés et la composition d'étapes du processus de découverte des connaissances. MQL est basé sur un KDDML (KDD Markup Language) de niveau inférieur qui exploite XML pour la représentation des processus d'extraction et des formes extraites.

Istituto di Linguistica Computazionale del CNR

ILC-CNR? Istituto di Linguistica Computazionale del CNR Area della Ricerca CNR Via Giuseppe Moruzzi N° 156124, Pisa, ? Italy

Tél. : (39) 050 315 2872 Téléc. : (39) 050 315 2834

Site Web : <http://www.ilc.cnr.it/indexflash.html>

Personne-ressource : Nicoletta Zamorani Calzolari

Courriel : direttore@ilc.cnr.it

Nombre de postes : 53

Projets de recherche

L'Istituto di Linguistica Computazionale del CNR (ILC-CNR) œuvre dans le domaine de la linguistique computationnelle depuis 1967, lorsqu'une division de linguistique computationnelle a été fondée au Centro Nazionale Universitario di Calcolo Elettronico (CNUCE). En 1978, l'ILC est devenu autonome et ne faisait donc plus partie du CNR. L'ILC – CNR a été l'un des principaux promoteurs de la notion des ressources de langage en tant que partie centrale de l'infrastructure linguistique (et ses incidences culturelles, économiques et politiques). Il a coordonné les principales initiatives liées aux ressources et à la normalisation du langage et a souvent été le promoteur de nouveaux paradigmes dans le domaine. L'ILC – CNR a conçu plusieurs types de corpora et de lexiques et les ontologies respectives, a élaboré une chaîne complète d'outils visant le traitement robuste de la langue italienne, l'acquisition d'information à partir de corpora et la désambiguïsation du sens des mots. Il a également mis au point des technologies pour divers domaines d'application (réponse aux questions, recherche d'information, exploration de textes, extraction monolingue et multilingue de la terminologie, structuration et acquisition d'ontologie, résumé, filtration de documents Web, préservation du patrimoine culture; par le traitement numérique d'images et des techniques numériques de bibliothèque, etc.)

Les activités de l'ILC-CNR s'articulent autour de sept domaines de recherche :

- Élaboration de normes et création de ressources linguistiques computationnelles;
- Modèles et méthodes de traitement des langages naturels et prototypes monolingues et multilingues orientés vers l'application;
- Outils et méthodes computationnels pour la recherche en sciences humaines, avec un accent sur les disciplines linguistiques et littéraires et la lexicographie;
- Matériel de bibliothèques et philologie computationnelle;
- Extracteur linguistique : un observatoire virtuel de l'italien contemporain;
- Architecture des technologies de langage pour la promotion de l'italien dans la société des connaissances;
- Traitement de langage naturel et accès naturel aux données.

Knowledge Representation Research Center, Free University of Bozen-Bolzano

KRDB Research Centre Faculty of Computer Science Free University of Bozen-Bolzano
Piazza Domenicani, 3 I-39100 Bozen-Bolzano BZ, Italy

Tél. : (39) 0471 016 120 Téléc. : (39) 0471 016 129

Site Web : <http://www.inf.unibz.it/krdb/>

Personne-ressource : Enrico Franconi

Courriel : franconi@inf.unibz.it

Nombre de postes : 22

Projets de recherche

Au cours des dernières années, les applications de connaissances et de bases de données ont graduellement convergé vers des technologies intégrées qui tentent de repousser les limites de chaque discipline. La recherche dans la représentation des connaissances (KR) s'est initialement concentrée sur les formalismes logiques qui traitent habituellement des bases de connaissances relativement petites. Elle fournit également d'importants services de déduction. Le langage de structuration de l'information est très expressif. Par exemple, la recherche sur les langages formels pour les ontologies est inspirée de la KR et la recherche en sémantique computationnelle du langage naturel. À l'opposé, la recherche sur les systèmes d'information et les bases de données (DB) porte principalement sur le stockage efficace et la recherche de puissants langages d'interrogation, ainsi que sur l'échange et la visualisation de beaucoup de documents (multimédias). Toutefois, les représentations de données étaient relativement simples et le raisonnement relatif à la structure et au contenu des documents a joué un rôle mineur seulement.

Cette distinction entre les exigences relatives à la représentation des connaissances et aux bases de données s'estompe rapidement. D'un côté, pour qu'un système de KR moderne soit utile dans les applications réalistes, il doit être en mesure de manipuler les grandes séries de données et de fournir des langages d'interrogation expressifs. Ces options donnent à penser que ces techniques élaborées dans le domaine des bases de données pourraient être utiles pour les systèmes de KR. D'un autre côté, l'information stockée dans le Web, dans des bibliothèques numériques et dans des dépôts de données est maintenant très complexe et inclut des structures sémantiques profondes, nécessitant des langues et des méthodes de modélisation plus intelligentes ainsi que des services de raisonnement sur les représentations complexes qui favorisent la conception, la gestion, l'accès flexible et l'intégration. En conséquence, il y a eu un vaste mouvement réclamant une vue intégrée logique des technologies relatives à la représentation des connaissances et aux bases de données. Les technologies KRDB offrent des formalismes prometteurs pour résoudre plusieurs problèmes relatifs à la modélisation conceptuelle de données et la conception d'ontologie, l'accès intelligent à l'information et le traitement des interrogations, l'intégration d'information, systèmes homologues, données semiestructurées, systèmes d'information distribués et accessibles sur le Web, services en ligne, logique computationnelle, linguistique computationnelle basée sur la logique et bio informatique.

Le centre de recherche KRDB de la faculté d'informatique de l'Université de Bozen-Bolzano a été fondé en 2002. Il vise à devenir un centre d'excellence international dans le domaine de la

recherche de base et appliquée sur les technologies de KRDB, et à proposer à des entreprises sélectionnées des idées et des technologies novatrices sur la recherche entreprise dans le centre.

Laboratory fo Applied Ontology

Polo Tecnologico Via Solteri, 38 38100 Trento, ITALY

Tél. : (39) 0461 436641 Téléc. : (39) 0461 435344

Site Web : <http://www.loa-cnr.it/index.html>

Personne-ressource : Nicola Guarino

Courriel : guarino@loa-cnr.it

Nombre de postes : 32

Projets de recherche

Le Laboratory for Applied Ontology (LOA) mène des recherches appliquées sur les fondations ontologiques de modélisation conceptuelle, en explorant le rôle des ontologies dans différents domaines, tels que : la représentation des connaissances, l'ingénierie des connaissances, la conception de bases de données, la recherche d'information, le traitement du langage naturel et le Web sémantique. Le groupe se distingue par une approche interdisciplinaire solide qui combine informatique, philosophie et linguistique et s'appuie sur la logique d'un paradigme unificateur. En ce qui concerne l'application, le groupe met l'accent sur l'utilisation des ontologies pour le commerce électronique, les systèmes d'information médicale, la modélisation des entreprises, l'intégration de ressources lexiques et l'accès au Web.

- ALgebraic Tools IN Formal Ontology (projet conjoint et d'échange)
- ILIKS: Interdisciplinary Laboratory on Interacting Knowledge Systems;
- MOSTRO, Modeling Security and Trust Relationships within Organizations;
- Metokis, Methodology and Tools Infrastructure for the Creation of Knowledge Units;
- SemanticMining Network of Excellence - Semantic Interoperability and Data Mining in Biomedicine.

Organismes et associations

Center for the Evaluation of Language and Communication Technologies

c/o BIC - Via dei Solteri, 38, 38100 Trento - Italy

Tél. : (39) 0461 405338 Téléc. : (39) 0461 405372

Site Web : <http://www.celct.it/>

Personne-ressource : Amedeo Cappelli

Courriel : cappelli@celct.it

Nombre de postes :

Mission

- Étudier et appliquer les processus d'évaluation dans le cadre de plusieurs technologies pertinentes de communication multimodale du langage naturel. Par exemple, la réponse à des questions multilingues, la traduction vocale, la recherche d'information multilingue à partir d'archives de multimédias, désambiguïsation du sens des mots, les technologies multimodales (émotions dans la parole et le langage et interaction d'agents synthétiques avec de nouveaux appareils).
- Lancer, organiser et promouvoir les campagnes d'évaluation.
- Suivre l'évolution des technologies de communication multimodale du langage naturel pour répondre en temps utile aux nouveaux besoins dans le domaine de l'évaluation.
- Éditer et publier les lignes directrices d'évaluation des technologies de communication multimodale du langage naturel.
- Collaborer activement avec d'autres institutions scientifiques et partenaires industriels.
- Fournir des services aux entreprises dans le domaine des technologies de communication multimodale du langage naturel.

FoLLI, Association of Logic, Language and Information

39100 Bolzano Italy

Tél. : (33) 3 83 58 17 90 Téléc. : (33) 3 83 41 30 79

Site Web : <http://www.folli.org/>

Personne-ressource : Carlos Araces

Courriel : carlos.araces@loria.fr

Mission

FoLLI constitue un point de rencontre de plusieurs domaines existants. De manière générale, FoLLI vise principalement le phénomène de l'information. Les structures d'information relatives aux études linguistiques, les mécanismes d'étude de l'informatique dans le traitement de l'information et la logique constituent l'étude mathématique des aspects représentationnels et informatiques. Cette relation générale est reflétée dans plusieurs domaines de recherche précis. Dans les domaines substantiels et florissants comme la théorie généralisée de quantificateurs la théorie type ou la sémantique dynamique, il est difficile de prédire le début et la fin d'une discipline. De plus, de nouveaux liens de ce type ne cessent d'émerger dans notre communauté.

FoLLI a pour principale activité d'agir en tant que société mère pour la communauté grandissante dans le domaine de la logique, du langage et de l'information. FoLLI sert de point de rencontre pour les membres de la communauté, finance les activités scientifiques dans le domaine et distribue l'information pertinente, etc.

FoLLI est active dans plusieurs initiatives internationales concrètes dans les domaines de la logique, du langage et de l'information. Plus particulièrement, elle organise chaque année l'école d'été européenne de la logique, du langage et de l'information (ESSLLI) et publie le journal de la logique, du langage et de l'information (JoLLI).

Tuscan Word Centre

The Tuscan Word Centre, Azienda Casanova Vellano 409, 51010 Pescia (PT), Italia

Tél. : (39) 55 29 5470

Site Web : <http://www.twc.it/>

Personne-ressource : John Sinclair

Courriel : jms@twc.it

Mission

Le Tuscan Word Centre (TWC) est une association sans but lucratif qui a pour objectif de promouvoir l'étude scientifique du langage. Le TWC a acquis sa réputation en offrant des cours abrégés de grande qualité et des manifestations similaires pour les chercheurs en langage et les industries langagières. un cours intensif de quatre jours intitulé **The Case for a Corpus - Spoken and Written Corpora** est préparé par 10 experts dans la conception et la construction de corpus. Comme toujours, l'accent est mis sur les questions pratiques de la manipulation de corpus et le logiciel de recherche.

Un séminaire guidé intitulé **Dial-a-Corpus - Compiling Corpora from the Web** est une manifestation dirigée par une équipe de six experts de réputation internationale. Le séminaire vise à trouver des conclusions raisonnables par des discussions et des essais pratiques. Les participants auront toutes les occasions de prendre part aux discussions et d'attirer l'attention des autres membres vers leurs propres besoins en matière de corpus.

Cours du TWC : Ces cours se concentrent sur l'utilisation de corpora électronique à des fins diverses, notamment : la traduction, assistance automatique ou par machine, traitement de langage, taggage, parsing, le soutien pédagogique en matière de langue, assistance à l'apprentissage de langue, la lexicographie et la référence langagière.

Autres activités du TWC : Services de conseillers et de consultants, évaluation des logiciels de traitement du langage, la conception et la gestion des projets, élaboration des produits langagiers, l'organisation de manifestations universitaires et professionnelles (p. ex. conférences, séminaires et ateliers de travail).

Ressources : Le TWC offre un accès à des ressources uniques et grandissantes dans le domaine du langage et de l'information, y compris de grands corpora en anglais et dans plusieurs autres langues européennes, un logiciel de pointe en matière de recherche selon les normes européennes, des outils spéciaux de traitement de corpus pour les analyses sophistiquées, un éventail d'ordinateurs et de logiciels privés, un groupe de conseillers spécialisés dans plusieurs aspects du langage et de l'information.

Japon

Laboratoires

Advanced Telecommunications Research Institute International ATR

ATR Spoken Language Communication Research Laboratories, 2-2-2 Hikaridai, "Keihanna Science City", Kyoto 619-0288 Japan

Site Web : www.slt.atr.co.jp

Personne-ressource : Selichi Yamamoto

Courriel : webmaster@slc.atr.jp

Nombre de postes : 30

Projets de recherche

ATR est un organisme autonome spécialisé dans la recherche et le développement de base et avancés dans le domaine des télécommunications. La recherche est exécutée sur demande des secteurs public et privé ou grâce à des fonds obtenus par des appels d'offres ouverts. ATR est également active dans l'application des résultats de recherche aux secteurs industriel et commercial.

- CHATR – Technologie de recréation de la parole humaine;
- ATR Matrix – Traduction automatique multilingue pour l'échange d'information;
- ATRCALL – Système d'apprentissage de langue assisté par ordinateur;
- Semantic Map – Visualisateur de réseaux communautaires;
- Recherche dans le domaine de technologies en traduction de langues basées sur le corpus;
- Élaboration de bases de données bilingues à grande échelle.

Computational Linguistics Laboratory

Computational Linguistics Laboratory, Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

Tél. : (81) 743 72 5240 Téléc. : (81) 743 72 5249

Site Web : cl.naist.jp

Personne-ressource : Yuji Matsumoto

Courriel : matsu@is.naist.jp

Nombre de postes : 36

Projets de recherche

- Acquisition de formes de traduction pratiques à partir de ressources langagières combinées;
- Acquisition d'ontologie robuste à partir de dictionnaires que des machines peuvent lire;
- Extraction d'information et classification de phrases appliquées aux abrégés MEDLINE;
- La Hinoki Treebank : Un treebank pour comprendre les textes;
- Élaborer un corpus de paraphrases pour la traduction de la parole.

Itaku Laboratory

Itakura Laboratory, Department of Information Electronics, Graduate School of Engineering,
Nagoya University, Furo-cho 1, Chikusa-ku, Nagoya, 464-8603 JAPAN

Tél. : (81) 52 789 3171 Téléc. : (81) 52 789 3172

Site Web : <http://www.itakura.nuee.nagoya-u.ac.jp/>

Personne-ressource : Fumitada Itakura

Courriel : itakura@nuee.nagoya-u.ac.jp

Nombre de postes : 24

Projets de recherche

Les technologies de traitement de la parole et des signaux acoustiques dans le contexte des communications humaines sont les principaux domaines d'intérêt du groupe de recherche.

Les systèmes suivants sont élaborés dans le cadre de la recherche :

- Système acoustique de réalité virtuelle;
- Système de reconnaissance de la parole ;
- Système de reconnaissance de la parole continue très riche en vocabulaire;
- Système robuste de reconnaissance de la parole;
- Traitement des signaux de la parole pour une parole robuste;
- Système de reconnaissance;
- Système de reconnaissance du locuteur;
- Analyse de scènes acoustiques;
- Masquage de données audio.

Kasuya Lab (Klab)

Utsunomiya University Faculty of Engineering, Graduate School of Engineering, 7-1-2, Yoto, Utsunomiya-city, Tochigi, 321-8585, JAPAN

Site Web : <http://www.utsunomiya-u.ac.jp/en/index.html>

Personne-ressource : Hideki Kasuya

Courriel : kasuya@klab.jp

Nombre de postes : 21

Projets de recherche

- Extraction automatique des paramètres de formants et de sources vocales à partir d'un modèle de production de la parole autorégressif avec des données exogènes;
- Modélisation de la nature apériodique des signaux de la parole;
- Dynamiques acoustique et articulatoire pour distinguer les voix des femmes de celles des hommes;
- Modélisation acoustique de la parole spontanée;
- Analyse sémantique fondée sur l'unification pour la compréhension robuste;
- Contrôle de temporisation relatif à la rétroaction par la voie de la parole pour une aisance dans la communication entre homme et machine;
- Conception et élaboration de bases de données;
- Analyse linguistique/paralinguistique/extralinguistique de la conversation informelle;
- Production de réponses très prosodiques.

Language Media Laboratory

Language Media Laboratory, Division of Media and Network Technologies
Graduate School of Information Science and Technology, Hokkaido University

Tél. : (81) 11 706 6534 Téléc. : (81) 11 709 6277

Site Web : http://sig.media.eng.hokudai.ac.jp/index_e.html

Personne-ressource : Kenji Araki

Courriel : araki@media.eng.hokudai.ac.jp

Nombre de postes : 26

Projets de recherche

Traitement du langage naturel

Méthode d'apprentissage inductive (IL-NLP)

Analyse syntaxique

Analyse morphologique

Analyse sémantique

Traduction automatique

Traduction de la parole

Traitement du langage parlé

Méthode d'entrée pour les téléphones cellulaires

Traitement d'information de la parole

Traitement robuste du langage parlé

Système de dialogue multimodal

Analyse de dialogues

Satisfaction des utilisateurs

Système d'information embarqué

Traitement et recherche d'information

Création automatique d'une base de données de sens commun

Recherche d'information relative aux émotions

Découverte du caractère illogique des émissions de parole

**National Institute of Advanced Industrial Science and Technology
(AIST)**

AIST Tokyo Headquarters 1-3-1, Kasumigaseki Chiyoda-ku, Tokyo 100-8921 Japan

AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

Tél. : (81) 29 862 6553 Téléc. : (81) 29 862 6559

Site Web : www.aist.go.jp

Personne-ressource : Hiroaki Kojima

Courriel : speechweb@m.aist.go.jp

Nombre de postes : 3200

Projets de recherche

- Analyse robuste des signaux de la parole, extraction de formes et séparation de la parole;
- Représentation de la parole au moyen d'un système universel intermédiaire de codes phonétiques;
- Codage verbal de la parole à base de formes et synthèse de la parole;
- Reconnaissance et recherche de parole au moyen de codes phonétiques universels.

Shikano Lab

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, Japan

Tél. : (81) 743 72 5280 Téléc. : (81) 743 72 5289

Site Web : <http://spalab.naist.jp/en/e-home.html>

Personne-ressource : Kiyohiro Shikano

Courriel : shikano@is.aist-nara.ac.jp

Nombre de postes : 42

Projets de recherche

- Conversion vocale entre les locuteurs, les langages et les émotions
- Système de dialogue homme-robot
- Système de guidage par un agent de parole
- Reconnaissance de commande de la parole pour l'ordinateur vestimentaire
- Adaptation non supervisé du locuteur dans des environnements bruyants

Tokyo Institute of Technology

Department of Computer Science, Graduate School of Information Science and Engineering
2-12-1 Oookayama, Meguro-ku, Tokyo 152-8552, Japan

Tél. : (81) 3 5734 3480 Téléc. : (81) 3 5734 3480

Site Web : <http://www.furui.cs.titech.ac.jp/english/index.html>

Personne-ressource : Sadaoki Furui

Courriel : furui@furui.cs.titech.ac.jp

Nombre de postes : 51

Projets de recherche

Reconnaissance de la parole

- Reconnaissance de la parole continue riche en vocabulaire;
- Reconnaissance de la parole spontanée;
- Reconnaissance robuste de la parole au moyen d'information prosodique.

Synthèse de la parole

- Synthèse de la parole naturelle et joliment tournée;
- Synthèse de la parole avec un style prosodique varié.

Système de dialogue parlé

- Système de dialogue multimodal pour la recherche d'information;
- Système informatique parallèle de dialogue parlé.

Organisations et Associations

Asian-Pacific Association for Machine Translation AAMT
c/o JEITA, Mitsui Sumitomo Kaijo Bldg., Annex, 3F, 3-11, Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062, JAPAN Tél. : (81) 3 3518 6418 Téléc. : (81) 3 3518 6472 Site Web : www.aamt.info Personne-ressource : Junichi Tsujii Courriel : aamt@aamt.info
Mission
L'association compte trois entités : les chercheurs, les manufacturiers et les utilisateurs des systèmes de traduction automatique. L'association s'emploie à concevoir des technologies de traduction automatique afin d'élargir l'étendue des communications globales efficaces. À cette fin, l'association œuvre dans l'élaboration, l'amélioration, la formation et la publicité relatives au système de traduction automatique.

Information Processing Society of Japan

Kagaku-kaikan (Chemistry Hall) 4F, 1-5 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062
JAPAN

Tél. : (81) 3 3518 8374 Téléc. : (81) 3 3518 8375

Site Web : <http://www.ipsj.or.jp/english/index.html>

Personne-ressource : Takashi Masuda

Courriel : through website mail only

Mission

Fondée en avril 1960, l'Information Processing Society of Japan (IPSJ) a été un chef de file dans les domaines techniques du traitement de l'information et de l'informatique pour les professionnels et les étudiants. L'IPSJ sera le fer de lance de l'évolution éclairée de l'informatique et la technologie dans une société de plus en plus informatisée, et contribuera à la création de nouvelles idées pour gérer l'imputabilité relative à la technologie d'information en évolution. Grâce à ses publications qui font autorité, à ses conférences et à d'autres activités, l'IPSJ jouera un rôle essentiel dans la prospérité du monde.

L'IPSJ favorise le développement des arts, des sciences, de l'industrie et des sciences humaines grâce à plusieurs activités relatives au traitement de l'information avec des ordinateurs et des communications et en fournissant les ressources et les possibilités de collaboration avec les sociétés sœurs.

Pays bas

Laboratoires

Computational Linguistics and AI Group, Tilburg University

Tilburg University, Faculty of Arts, Computational Linguistics & AI Group, P.O. Box 90153, 5000 LE Tilburg, Netherlands

Tél. : (31) 13 4663060 Téléc. : (31) 13 4663110

Site Web : <http://let.uvt.nl/research/ti/indexE.htm>

Personne-ressource : Harry C. Bunt

Courriel : Harry.Bunt@uvt.nl

Nombre de postes : 29

Projets de recherche

Le programme de recherche de la section de linguistique computationnelle et AI vise à étudier les aspects computationnels de l'acquisition, la représentation, le traitement et la communication de connaissances exprimées dans le langage naturel.

Le programme s'appuie sur l'idée que le langage, les connaissances et le traitement d'information sont mieux étudiés s'ils sont abordés dans une approche intégrée. Plusieurs phénomènes linguistiques ne peuvent être compris que si l'on considère les connaissances contextuelles et l'état dynamique de l'information des agents qui produisent et traitent le langage. D'un autre côté, la communication fondée sur le langage détermine plusieurs aspects de l'acquisition, de la repr.sentation et du traitement des connaissances par des agents intelligents. L'approche intégrée de l'étude du langage, des connaissances et de l'information et la façon dont ces facteurs sont associés par ordinateur prennent la forme dans plusieurs projets qui font partie de l'un des autres sous-programmes suivants :

- Théorie de dialogue et communication homme-machine
- Induction des connaissances linguistiques
- Pragmatique et sémantique computationnelles
- Reasonnement avec les connaissances et la croyance

Mathematics and Computer Science? Faculty, University of Twente

University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science
Faculty. Office P.O. Box 217 7500 AE Enschede

Tél. : (31) 53 489 3701 Téléc. : (31) 53 489 4571

Site Web : <http://www.ewi.utwente.nl/en/>

Courriel : BFD_EWI@ewi.utwente.nl

Projets de recherche

Multimédia, réalité virtuelle et interaction homme-ordinateur

La recherche dans ce domaine est effectuée par de nombreux groupes du département d'informatique. Le groupe de gestion des données s'intéresse de plus en plus au multimédia. Cette attention au multimédia et à la recherche d'information à Twente est unique dans les Pays-Bas. La recherche porte également sur le développement de solutions en gestion de données pour des environnements d'intelligence ambiante par l'infrastructure égale à égale de bases de données de multimédias pour favoriser un accès sensible au contexte et orienté vers les personnes dans des environnements distribués ou spéciaux. À l'université de Twente, l'interaction homme-ordinateur a toujours été ciblée vers l'interaction multimodale orientée par les possibilités et les impossibilités de l'environnement. Le point de départ est la présentation multimodale. Cette recherche est associée à la recherche dans le domaine des interfaces utilisateurs sociales et à un intérêt à l'égard des possibilités de la parole et du langage qui sont également présentes dans le domaine de la recherche multimédia.

Institute for Logic, Language and Computation

ILLC Office University of Amsterdam, Plantage Muidersgracht 24 1018 TV Amsterdam, Netherlands

Tél. : (31) 20 525 6051 Téléc. : (31) 20 525 5206

Site Web : <http://www.illc.uva.nl/>

Personne-ressource : Ingrid Van Loon

Courriel : ingridillc.uva.nl

Nombre de postes : 42

Projets de recherche

Plusieurs flots d'information propulsent le monde technologique d'aujourd'hui. Il est difficile pour la science contemporaine de fournir une plus grande vision, et le cas échéant, d'améliorer la pratique existante. Au cours de ce siècle, l'information est devenue un thème fondamental pour les études scientifiques dans plusieurs disciplines.

L'encodage, la transmission et la compréhension de l'information sont les principaux thèmes de la recherche à l'Institute for Logic, Language and Computation (ILLC) de l'Université d'Amsterdam. Le grand contexte dans lequel se voit l'ILLC est celui d'une science d'information à venir ou de l'informatique qui porte sur le flux d'information dans les langages formels et naturels ainsi que sur plusieurs moyens de communication, y compris la musique et les images de divers types.

La recherche à l'ILLC vise à élaborer des systèmes logiques capables de manipuler cette diverse information, en utilisant les visions dans des disciplines telles que la linguistique, l'informatique, la science cognitive et l'intelligence artificielle. Des méthodes supplémentaires sont également en cours, lorsqu'il y a lieu, de la statistique à la théorie d'argumentation.

Outre ses objectifs de recherche précis, l'ILLC vise à repousser les obstacles habituels entre les facultés et les disciplines, et à rallier les chercheurs en information dans le domaine de l'informatique, de la linguistique, de la philosophie et des sciences sociales. De plus, l'institut diffuse des normes logiques exactes en matière de clarté de la sémantique, de perspicuité algorithmique et la calculabilité efficace.

L'optique de la science d'information qui en résulte dépasse la carte de l'université. L'ILLC s'engage également à diffuser ses résultats au monde de l'enseignement général, de la formation professionnelle et de la recherche industrielle.

Dutch Research School for Information and Knowledge Systems

SIKS, Utrecht University, Institute for information and computing sciences, P. O. Box 80.089,
3508 TB Utrecht

Tél. : (30) 253 4083/1454 Téléc. : (30) 251 3791

Site Web : <http://www.cs.uu.nl/siks/>

Personne-ressource : Dr. R.J.C.M. Starmans

Courriel : office@siks.nl

Nombre de postes : 270

Projets de recherche

SIKS est l'école néerlandaise de recherche sur les systèmes d'information et de connaissances. Elle a été fondée en 1996 par des chercheurs dans le domaine de l'intelligence artificielle, des bases de données et des systèmes d'information ainsi que de l'ingénierie de logiciels. SIKS est une école de recherche interuniversitaire qui compte 12 groupes de 270 chercheurs, incluant 120 étudiants au doctorat.

SIKS a pour mission :

- Mener une recherche fondamentale et appliquée de haut niveau dans le domaine des sciences de l'information et de l'informatique, particulièrement dans le domaine des systèmes d'information et de connaissances;
- Organiser un programme pédagogique de haute qualité en quatre années à l'intention de ses étudiants au doctorat qui travaillent dans 10 universités des Pays-Bas ou dans des entreprises chefs de file dans le domaine de la technologie de l'information et des connaissances;
- Faciliter et stimuler la collaboration et la communication entre nos membres (étudiants au doctorat, chercheurs universitaires, chargés de recherche et membres associés) et entre l'école et ses intervenants, notamment les entreprises chefs de file dans le domaine de la technologie de l'information et des connaissances;

Le programme de recherche porte sur l'étude des systèmes d'information et des connaissances à partir de perspectives des domaines scientifiques suivants :

- Systèmes de bases de données, tels qu'ils sont étudiés en informatique;
- Systèmes d'information selon leur domaine d'application;
- Ingénierie des connaissances telle qu'elle est étudiée dans l'intelligence artificielle;
- Ingénierie de logiciels telle qu'elle est appliquée dans les systèmes d'information et des connaissances;
- Informatique théorique et logique appliquée, en tant que base des systèmes d'information et

des connaissances.

Human Computer Studies Laboratory, University of Amsterdam

Human-Computer Studies Laboratory, Matrix I, Kruislaan 419, 1098 VA, Amsterdam

Tél. : (31) 20 888 4689 Téléc. : (31) 20 525 6896

Site Web : <http://hcs.science.uva.nl/index.html>

Personne-ressource : Pieter Willems Adriaan

Courriel : pietera@science.uva.nl

Nombre de postes : 31

Projets de recherche

GARP Outils de communication et d'articulation des connaissances (pour le raisonnement relatif au comportement des systèmes).

HOPS est un projet triennal de déploiement en Europe de plates-formes publiques frontales commandées par la voix afin de permettre aux citoyens européens d'accéder à la plus proche administration publique.

Gestion des connaissances **METIS** : D'employés intelligents à des organismes intelligents.

IRIS Information Retrieval and Information, Universtiy of Nijmegen

Room A4021 (Floor #4) Toernooiveld 1 6525 ED Nijmegen the Netherlands

Tél. : (31) 24 365 34 56 Téléc. : (31) 24 365 33 56

Site Web : <http://osiris.cs.kun.nl/iris/web-docs/>

Personne-ressource : Nicole el Moustakim

Courriel : nicolem@cs.ru.nl

Nombre de postes : 27

Projets de recherche

Le projet PRONIR, financé par NWO, est réalisé en collaboration avec l'Université de Tilburg. Il porte sur deux domaines, dont la recherche par profil qui porte sur la capacité d'un système de recherche d'information à mettre au point un ensemble de ressources pour répondre aux besoins des chercheurs. L'IRIS élabore un modèle générique pour les ressources Web et conçoit une théorie générale pour une fonction efficace de courtage. Personne-ressource : Erik Proper.

Modèles de recherche : Le moteur de recherche est le courtier en information le plus connu sur Internet. Plusieurs approches ont été proposées selon certaines méthodes pour représenter le contenu des documents, un langage de recherche et une stratégie d'harmonisation. IRIS tente d'élaborer de nouveaux modèles pour les applications au-delà du moteur de recherche. Personne-ressource : Theo van der Weide.

The Leiden University Centre for Linguistics

Tél. : (31) 71 527 2105

Site Web : <http://www.ulcl.leidenuniv.nl/>

Personne-ressource : Vincent Van Heuven

Courriel : v.j.j.p.van.heuven@let.leidenuniv.nl

Nombre de postes : 135

Projets de recherche

Acquisition : The (morpho)syntactic development of children with a cochlear implant (NWO-Vidi)

Amérindien: Giving them back their languages: The endangered Amerindian languages of the Guianas (NWO-Endangered Languages)

Bantu : The phrasal phonology of Bantu languages (NWO-Veni)

Bantu : Word order and morphological marking in Bantu (NWO-Programme)

Degrees across Categories (NWO-Vidi)

Anglais : The codifiers and the English language: tracing the norms of Standard English (NWO-Vici)

Néerlandais : Dutch as a construction language (NWO-Vidi)

Ethiopie : The morpho-syntax of two modal categories in Omotic languages of south-west Ethiopia (NWO-Endangered Languages)

Europe Asie de l'Est : Spinoza-programme of Frederik Kortlandt (Spinoza)

Fataluku : The Fataluku Language Project (NWO-Endangered Languages)

Hittite : Connectives in Hittite texts (NWO-Veni)

Indo-européen : The Indo-European Etymological Dictionary (IED) (NWO, Brill)

Indonésie : Linguistic Variation in Eastern Indonesia: The Alor and Pantar Project (NWO-VIMP)

Syntactic Change : How and Why? (NWO-Vidi)

The Typological Database Project (NWO, participating universities)

Utrecht Institute of Linguistics OTS

Trans 10 3512 JK UTRECHT The Netherlands

Tél. : (31) 30 253 6006 Téléc. : (31) 30 253 6000

Site Web : <http://www-uilots.let.uu.nl/>

Personne-ressource : Martin Everaert

Courriel : Martin.Everaert@let.uu.nl

Nombre de postes : 52

Projets de recherche

L'OTS d'Utrecht Institute (UiL) a pour objectif de développer l'expertise scientifique dans les systèmes sous-jacents au langage, à la parole et à l'utilisation. À cette fin, sont indispensables l'étude de la structure de la faculté langagière humaine et la façon dont elle s'intègre dans la capacité générale cognitive de l'homme. La recherche est menée selon l'idée que le système informatique du langage ne peut pas provenir de considérations fonctionnelles des systèmes d'utilisation ou d'interprétation : les systèmes d'utilisation ou d'interprétation et de communication concordent avec plusieurs langages naturels et non naturels, et ne peuvent donc pas expliquer pourquoi le langage humain a les particularités qu'il a. Par ailleurs, un fait est sûr dans le langage humain : il peut servir à argumenter, à communiquer, à penser, etc. Une analyse ne peut être correcte si une analyse formelle du système informatique s'avère incompatible avec les faits de base dans l'utilisation et l'interprétation du langage, car les phrases réelles du langage humain peuvent être utilisées et interprétées. Par conséquent, une saisie correcte de l'interface entre le système formel et les systèmes d'interprétation et d'utilisation est un critère indispensable de la pertinence de toute théorie linguistique. Des facteurs similaires portent sur la relation entre la forme linguistique et son expression. La forme linguistique est compatible avec plusieurs modes d'expression. En raison d'un certain mode d'expression, une étude des propriétés est primordiale pour préciser les contraintes que le mode d'expression pose sur le système linguistique avec qui il interface.

La recherche de base de l'OTS est divisée en sept groupes en plus du programme de recherche spécial Language in Use.

- Syntaxe et sémantique
- Morphologie et phonologie
- Linguistique computationnelle et logique
- Phonétique
- Élaboration de systèmes de langage
- Utilisation du langage
- Psycholinguistique expérimentale

Organismes et associations

Acoustical Society of the Netherlands

NAG-secretariat P.O.Box 480 6200 AL Maastricht, The Netherlands

Tél. : (31) 43 3470956 Téléc. : (31) 43 3438781

Site Web : <http://www.nag-acoustics.nl/>

Personne-ressource : Ingrid Bonfrere

Courriel : secr@nag-acoustics.nl

Mission

Le Nederlands akoestisch Genootschap (NAG) est la société acoustique des Pays-Bas qui compte 500 membres de l'industrie acoustique néerlandaise, d'universités et d'instituts de recherche. Notre objectif principal est de promouvoir le développement et l'application de la science et des techniques relatives à l'acoustique.

Pour atteindre ses objectifs, le NAG tente, entre autres :

- d'organiser des réunions scientifiques environ tous les trois mois;
- de publier le procès-verbal de chaque réunion scientifique;
- de publier le NAG-almanak, y compris une liste de tous les membres et les bailleurs de fonds.

European Network of Excellence in Human Language Technologies

ELSNET, Trans 10, 3512 JK UTRECHT, The Netherlands

Tél. : (31) 30 253 6050 Téléc. : (31) 30 253 6000

Site Web : <http://www.elsnet.org>

Personne-ressource : Steven Krauwer

Courriel : steven.krauwer@let.uu.nl

Mission

L'European Network of Excellence in Human Language Technologies (ELSNET) a pour principal objectif de développer les technologies du langage humain dans un sens général en réunissant les principaux acteurs européens dans la recherche, le développement, l'intégration ou le déploiement dans le domaine de la technologie du langage et de la parole et dans des disciplines apparentées. Le rôle du réseau consiste à offrir un environnement qui permet l'exploitation optimale des ressources humaines et intellectuelles disponibles afin de faire progresser les technologies. Cet environnement comprend un nombre de structures (comités, groupes d'intérêts), d'actions (écoles d'été et ateliers de travail) et de services (site Web, listes de diffusion, bulletins, diffusion de l'information courtage de connaissances).

Royaume uni

Laboratoires

Center for communication Interface Research

Centre for Communication Interface Research The School of Engineering and Electronics The University of Edinburgh King's Buildings, Mayfield Road Edinburgh, EH9 3JL

Tél. : (44) 131 650 2783 Téléc. : (44) 131 650 2784

Site Web : <http://www.ccir.ed.ac.uk/index.html>

Personne-ressource : Mervyn Jack

Courriel : maj@ccir.ed.ac.uk

Nombre de postes : 20

Projets de recherche

Le Centre for Communication Interface Research (CCIR) fait partie de l'école de génie et d'électronique de l'Université d'Edinburgh.

La recherche du CCIR porte sur les personnes qui interagissent avec des systèmes informatisés.

CCIR mène des recherches sur le génie de l'utilisation et de dialogue pour les services de reconnaissance de la parole (vCommerce) pour les services d'Internet (eCommerce) et les services mobiles (mCommerce).

Le génie de l'utilisation est la discipline de créer des interfaces homme-ordinateur efficaces, confortables et selon les besoins des utilisateurs.

L'ingénierie du dialogue est l'application des principes scientifiques à la conception, à la construction et à l'entretien d'interfaces homme-ordinateur.

Les bailleurs de fonds commerciaux du CCIR favorisent la recherche sur la conception et l'évaluation expérimentale des interfaces homme-ordinateur. Le CCIR entreprend également des études plus générales des aspects humains de la conception d'interface homme-ordinateur.

The Center for Speech Technology Research

The Centre for Speech Technology Research University of Edinburgh 2 Buccleuch Place
EDINBURGH, EH8 9LW, UK

Tél. : (44) 131 650 4434 Téléc. : (44) 131 650 6626

Site Web : <http://www.cstr.ed.ac.uk/people/>

Personne-ressource : Steve Renals

Courriel : S.Renals@ed.ac.uk

Nombre de postes : 37

Projets de recherche

AMI : Augmented Multiparty Interaction

AMI est un projet européen intégré qui porte sur l'interaction multimodale améliorée avec l'ordinateur dans le contexte des réunions.

BARKS

Le projet BARKS explore l'utilisation de modèles de commutation dynamiques linéaires pour la reconnaissance de la parole automatique.

Cougar

Le projet Cougar étudie l'utilisation d'un domaine articulatoire pour le calcul des coûts de jointures en utilisant parallèlement des transitions unitaires dans la synthèse de la parole.

ESLASR

ESLASR vise à améliorer la qualité de la reconnaissance automatique de la parole en utilisant des HMM à configuration dispersée avec des caractéristiques articulatoires acoustiques.

Espresso

Nouveaux modèles acoustiques pour l'ASR

Expressions

Expressions vise à améliorer la qualité de la prosodie et l'intonation pour la synthèse de la parole par sélection d'unités.

The Festival speech synthesis system

The Edinburgh Speech Tools

Speech tools est une série de bibliothèques de base utilisée par Festival et diverses autres applications.

SVitchboard

Petites tâches en vocabulaire à partir du standard téléphonique.

TESS: Testing Evaluation of Speech Synthesis

TESS est un projet conçu pour étudier les processus psychoacoustiques sous-jacents à

l'évaluation humaine de la parole synthétique.

Center for vision, Speech and Signal Processing

Center for vision, Speech and Signal Processing, School of Electronics & Physical Sciences,
University of Surrey, Guildford, GU2 7XH, Surrey, UK

Tél. : (44) 1483 686030 Téléc. : (44) 1483 686031

Site Web : <http://www.ee.surrey.ac.uk/CVSSP/index.php>

Personne-ressource : Joseph Kittler

Courriel : J.Kittler@surrey.ac.uk

Nombre de postes : 78

Projets de recherche

Le Centre for Vision, Speech and Signal Processing est l'un des principaux centres de recherche du School of Electronics and Physical Sciences de l'Université de Surrey. Il a pour objectif de faire avancer l'état de la connaissance dans les domaines du traitement des signaux multimédia et de la vision artificielle, en mettant l'accent sur les applications vidéo et audio.

Le groupe de recherche en traitement et interprétation des signaux multimédia réunit les intérêts de recherche dans le traitement des signaux et l'analyse des données sensorielles. La portée de la recherche dépasse les tâches conventionnelles du traitement des signaux que sont la réduction du bruit et la compression de l'information aux fins de communication et de stockage. Notre premier but est de développer des techniques qui permettent aux machines de comprendre le contenu des données et d'y donner suite (p. ex. la reconnaissance de la parole). La représentation du contenu que nous tentons d'extraire pour servir de base au processus décisionnel et à l'interprétation doit donc mettre l'accent sur l'information sémantique acheminée par les données. Nos travaux se distinguent également par l'exploitation de la complémentarité des multimédias et des données multimodales et la synergie résultant de son analyse et traitement.

Pattern Recognition & AI La technologie clé de l'interprétation des signaux multimédia. La méthode de validation prédictive du centre en termes de modélisation des classes de formes aide à la détection des anomalies et défaillances.

Automated Analysis of Audio-Visual Material Nous avons étudié plusieurs méthodes d'interprétation automatique du matériel de sport, en tant qu'exemple de l'analyse d'un contenu multimédia très structuré.

Speech and Audio Processing Les activités dans ce domaine comprennent l'analyse du temps fréquence et le filtrage adaptatif des signaux audio, entre autres la décomposition périodique apériodique de la parole et les mécanismes de reconnaissance de la parole à partir de gestes verbaux.

Biometric Personal Identity Authentication Authentification de l'identité en combinant l'image du visage, les caractéristiques vocales et les mouvements des lèvres. Une technologie pareille est nécessaire pour le contrôle d'accès, les applications intelligentes et les systèmes

de sécurité.

Natural Language and Information Processing Group

University of Cambridge, Computer Laboratory, William Gatesw Building, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, UK

Tél. : (44) 1 223 763500 Téléc. : (44) 1 223 334678

Site Web : <http://www.cl.cam.ac.uk/Research/NL/>

Personne-ressource : Caroline Matthew

Courriel : cb210@cl.cam.ac.uk

Nombre de postes : 20

Projets de recherche

Technologie robuste de parsing

- Parsing statistique robuste (RASP)
- Induction automatique grammaire/lexique
- Disambiguïsation
- Analyse sémantique

Traitement orienté-contraintes

- Système LKB et grammaire des ressources anglaises
- Sémantique compositionnelle non précisée
- Questions relatives à la représentation (défauts, règles lexiques, multilinguisme)
- Acquisition lexicale
- Outils de langage naturel Alvey

Résumé de textes

- Analyse robuste du discours (zonage argumentatif)
- Types de structures du discours et leurs rôles
- Effets des facteurs de contexte
- Programme DARPA/NIST TREC

Recherche et navigation motivées par la rhétorique

- Survol des textes électroniques; accès à des textes similaires
- Utilisation de l'information de citation aux fins de recherche

Génération et régénération de textes robustes

- simplification des textes
- Régénération de texte résumé cohérent
- Déclaration de résultats par l'extraction d'information

Acquisition et évolution du langage

- Acquisition de grammaire catégorielle
- Simulation en évolution du développement et du changement du langage

Évaluation des systèmes

- Méthode d'évaluation des systèmes NLIP (rôle du contexte de tâches)
- Évaluation sommaire par des expériences extrinsèques
- Évaluation sommaire par des unités de contenu (faits)

Exploration de textes et bioinformatique

- Conservation automatique de la littérature scientifique
- Parsing et extraction d'information

Natural Language Processing and AI

The Faculty of Mathematics and Computing, The Open University, Walton Hall, Milton Keynes, Buckinghamshire, MK7 6AA, UK.

Tél. : (44) 1908 652892 Téléc. : (44) 1908 652140

Site Web : <http://mcs.open.ac.uk/>

Personne-ressource : Anne De Roeck

Courriel : a.deroeck@open.ac.uk

Nombre de postes : 21

Projets de recherche

La recherche en intelligence artificielle (IA) du département d'informatique se divise en deux domaines généraux. Nous étudions la façon d'exploiter les ordinateurs pour modéliser les processus cognitifs et la façon dont les modèles peuvent s'appliquer à des domaines tels que la planification. Notre recherche dans le traitement du langage naturel vise à trouver les formes de textes en langage naturel et l'interprétation des textes et des figures.

Nous tentons particulièrement à trouver des réponses aux questions suivantes :

Quelles sont les conditions nécessaires pour une reconnaissance de sentience dans les entités artificielles telles que les robots?

Comment pouvons-nous élaborer des représentations analogiques, diagrammatiques ou connexionnistes pour le raisonnement sur l'action et la planification?

Comment les ordinateurs peuvent comprendre les diagrammes?

Quelles sont les formes qui peuvent être décelées dans les corps d'un texte réel?

Comment distiller les concepts des données textuelles?

Comment pouvons-nous modéliser l'interaction entre différents niveaux de représentation dans un système intelligent?

Un bon nombre des membres qui font partie de la recherche font partie de l'Open University Natural Language Processing Group.

Projets

CLinical E-science Framework

Designing adaptive infOrmation exTraction from text for KnOwledge Management.

European Learning Grid Infrastructure based on GRID technologies

HALO

Network for Euro-Mediterranean LAnguage Resources

Natural Language Processing Research Group

NLP Research Group, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK

Tél. : (44) 114 222 1814 Téléc. : (44) 114 222 1810

Site Web : <http://nlp.shef.ac.uk/>

Personne-ressource : Yorick Wilks

Courriel : y.wilks@dcs.shef.ac.uk

Nombre de postes : 45

Projets de recherche

Le groupe de traitement du langage naturel de l'Université de Sheffield NLP existe depuis dix ans. Il est l'un des plus grands groupes de recherche au Royaume-Uni; nous participons à des compétitions internationales sur le meilleur compteur informatique, le meilleur répondeur aux questions, etc. (Habituellement des compétitions aux États-Unis) et nous avons souvent gagné. Nous mettons l'accent sur l'utilisation de représentations déterminées par un code dans le contenu de sens, la croyance et les connaissances, sur les techniques d'apprentissage automatique pour dériver les données de sources comme le Web, et sur la fourniture d'architectures logicielles pour étayer la recherche en traitement du langage naturel. Notre architecture GATE a été installée dans 400 sites de par le monde. Nous voyons également dans le traitement du langage sur les ordinateurs une route essentielle pour comprendre comment fonctionne l'esprit, l'objectif conventionnel de l'intelligence artificielle. Le traitement du langage naturel est le principal moyen selon nous d'utiliser et de s'adapter au WWW, à savoir réunir les machines intelligentes et les connaissances, y compris toutes les connaissances scientifiques, avec les personnes par la technologie de la conversation et en étant le principal contributeur à l'avenir des jeux électroniques et aux loisirs.

Institute for Communication and Collaborative Systems

School of Informatics, Appleton Tower, Crichton Street, Edinburgh, EH8 9LE, Scotland, UK

Tél. : (44) 131 650 2690 Téléc. : (44) 131 651 1426

Site Web : <http://www.iccs.inf.ed.ac.uk/>

Personne-ressource : Johanna Moore

Courriel : j.Moore@ed.ac.uk

Nombre de postes : 213

Projets de recherche

L'Institute for Communicating and Collaborative Systems (ICCS) mène des recherches de base dans la nature de la communication entre les hommes d'un côté et entre les hommes et les machines de l'autre côté en utilisant le texte, la parole et les graphiques et dans la conception de systèmes de dialogue interactifs, au moyen de mécanismes informatiques et algorithmiques, avec des applications incluant le traitement du langage naturel, la présentation et la recherche d'information, la formation, les analyses musicales et l'instruction. L'ICCS fait partie du School of Informatics de l'Université d'Edinburgh.

Les domaines de recherche que couvre l'ICCS et le nom des chercheurs sont les suivants :

Annotation et technologie de balisage : Jean Carletta, Ewan Klein, Oliver Lemon, Henry Thompson, Richard Tobin

Traitement du langage naturel biomédical : Claire Grover, Ewan Klein, Bonnie Webber

Psycholinguistique computationnelle : Frank Keller, Jon Oberlander

Sémantique computationnelle : Ewan Klein, Mirella Lapata, Alex Lascarides, Oliver Lemon, Colin Matheson, Bonnie Webber

Agents de conversation : Oliver Lemon, Hiroshi Shimodaira, Mark Steedman

Discours et langage : Jean Carletta, Alex Lascarides, Oliver Lemon, Colin Matheson, Johanna Moore, Bonnie Webber

Technologie éducative : Johanna Moore, Helen Pain, Keith Stenning

Communication graphique : John Lee, Jon Oberlander, Keith Stenning

Raisonnement humain : Keith Stenning, Jon Oberlander

Extraction d'information : Claire Grover, Ewan Klein, Bonnie Webber

School of Computer Science, University of Manchester, IMG, NLE

School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester, M13 9PL, United Kingdom

Tél. : (44) 161 275 6154 Téléc. : (44) 161 275 6204

Site Web : <http://www.cs.manchester.ac.uk/>

Personne-ressource : Carole Goble

Courriel : carole@cs.man.ac.uk

Nombre de postes : 39

Projets de recherche

Information Management Group (IMG) Ce groupe est dirigé par la professeure Carole Goble et le professeur Norman Paton. La recherche qu'il effectue porte sur la conception, l'élaboration et l'utilisation de systèmes de gestion des données et des connaissances. Les applications stimulantes motivent et valident notre recherche, particulièrement le Web sémantique et les sciences en ligne. Le groupe se compose de 5 universitaires, 17 employés de recherche et 31 étudiants qui font de la recherche. Les trois millions de livres sterling injectés dans le programme de recherche actuel de l'IMG provient de diverses sources, y compris les conseils de recherche du Royaume-Uni, soit EPSRC et BBSRC, l'Union européenne, le Wellcome Trust, US DARPA, la DTI et l'industrie. Le groupe a de très bonnes relations avec de grandes entreprises industrielles, notamment AstraZeneca, GlaxoSmithKline, Millennium Pharmaceuticals, IBM, Oracle and Sun Microsystems et les petites et moyennes entreprises comme geneticXchange. Outre la collaboration avec des entreprises réputées, le groupe pris part à la création de deux entreprises en démarrage : Sagitus Solutions et Network Inference. Ces deux entreprises ont récemment bénéficié de 4 millions de dollars en capital risque de Nokia Ventures. L'IMG joue un rôle prépondérant dans la normalisation et les forums industriels, de chef de file du W3C Semantic Web et de co-président des groupes de travail relativement au forum industriel sur la pharmacologie et la biotechnologie I3C et du Global Grid Forum. Les membres du groupe siègent à plusieurs comités britanniques et internationaux d'élaboration des politiques et de financement gouvernementaux et communautaires.

Ingénierie du langage naturel Il s'agit de conception d'outils pour appuyer les personnes qui font œuvre utile du texte du langage naturel. Le projet MultiFlora vise à prouver que l'extraction d'information peut être améliorée par l'analyse de plusieurs textes parallèles. Cette technique, appliquée aux descriptions botaniques taxonomiques, peut être un outil utile dans l'informatique de la biodiversité. Multiflora vise à i) former l'assise dans un modèle de domaine structuré, en tirant profit de l'expertise en informatique médicale à Manchester et ii) à élargir la portée taxonomique, en contribuant considérablement à la recherche sur la biodiversité.

Scool of computing, Robert Gordon University

Scool of computing, Robert Gordon University, Schoolhill, Aberdeen, AB10 1FR, Scotland, UK

Site Web : <http://jabber.comp.rgu.ac.uk/research/aboutus.php>

Personne-ressource : Susan Craw

Courriel : S.Craw@comp.rgu.ac.uk

Nombre de postes : 27

Projets de recherche

- Agents Web adaptatifs.
- Algorithmes pour la résolution de contraintes distribuée
- Développer le raisonnement sur les taxonomies restreintes
- Faciliter l'acquisition des connaissances pour la conception à base de cas
- Découverte de connaissances à partir de texte
- RECOP : Representing and Reformulating Constraint and Optimisation Problems
- Centre de technologies Web intelligentes
- Appuyer des groupes aux intérêts communs pour échanger l'information entre eux
- HORUS : Hieroglyphic Optical Recognition and Update System
- ProfileSkim – Un navigateur intelligent de documents
- Sentinel – Outils d'échange des connaissances par courriel
- Centre de technologies Web intelligentes

Computing Science at Aberdeen

Department of Computing Science, University of Aberdeen, Aberdeen AB24 3UE, Scotland

Tél. : (44) 1224 272295 Téléc. : (44) 1224 273422

Site Web : <http://www.csd.abdn.ac.uk/>

Personne-ressource : Sandra Edwards

Courriel : sedwards@csd.abdn.ac.uk

Nombre de postes : 84

Projets de recherche

Le département d'informatique jouit d'une réputation internationale établie en intelligence artificielle et en bases de données, notamment en gestion des connaissances et de l'information. Une bonne partie des travaux de recherche actuels du département vise les problèmes de gestion des connaissances et de l'information dans le contexte d'internet et de World Wide Web, notamment la recherche d'une solution qui aidera les hommes à prendre des décisions et à agir face à un grand nombre de connaissances et d'information.

Le département est inclus dans plusieurs projets conjoints internationaux. Il est également membre de l'Interdisciplinary Research Collaboration (IRC) in Advanced Knowledge Technologies qui est financé par l'EPSRC. Parmi les projets récents, mentionnons des travaux fondamentaux en sciences en ligne qui ont trait à l'élaboration d'un Web sémantique. Nos principaux thèmes de recherche actuels sont cités ci-dessous. Ces thèmes sont interreliés, de même que plusieurs membres travaillent sur plus d'un thème à la fois.

Agents logiciels intelligents Nous étudions la façon de créer des réseaux d'entités logicielles autonomes qui peuvent remplacer l'homme dans un environnement complexe, par l'acquisition de l'information, l'apprentissage et l'interaction avec d'autres agents. Exemple d'application : commerce en ligne, conception de marchés de travail électroniques robustes d'agents qui achètent et vendent des produits et services.

Technologies des connaissances Nous développons des procédures d'acquisition, de modélisation, d'utilisation et de réutilisation des connaissances. Exemple : ingénierie en ligne dans le domaine de l'industrie aérospatiale, présenter l'expertise accumulée de centaines d'ingénieurs à un concepteur qui travail sur un moteur jet, en modélisant et en réutilisant les connaissances significatives de l'intranet d'une entreprise.

Interprétation et communication des connaissances Nous recherchons les techniques d'extraction des connaissances à partir des flux de données, en ciblant les manifestations importantes et en communiquant avec les utilisateurs par la génération du langage naturelle et les graphiques. Exemples d'application : soins de santé en ligne, déceler les urgences des patients à partir de données en temps réel dans une unité de soins intensifs, afin d'aider le personnel médical à prendre les mesures nécessaires; bioinformatique, modélisation et visualisation des interactions de protéines en 3D.

Language and computation group

Department of Computer Science University of Essex Wivenhoe Park Colchester CO4 3SQ
United Kingdom

Tél. : (44) 1206 872770 Téléc. : (44) 1206 872788

Site Web : <http://cswww.essex.ac.uk/LAC/>

Courriel : genoffice@essex.ac.uk

Nombre de postes : 20

Projets de recherche

Les membres du Language and Computation group étudient la plupart des domaines de la linguistique humaine et de l'ingénierie du langage naturel.

Grammaires et parsing : La recherche met l'accent sur les formalismes grammaticaux à base de contraintes (Arnold, Borsley, Sadler, Spencer) et le parsing statistique et symbolique (Arnold, Borsley)

Sémantique et interprétation de la sémantique : Notre recherche porte sur les fondations logiques de la sémantique (Fox), sur les modèles computationnels motivés par la psychologie pour le traitement de la sémantique (Poesio), sur les modèles linéaires de la représentation lexicale et textuelle utilisés par exemple dans la résolution anaphorique (Poesio).

Applications : Certains secteurs d'intérêt comprennent la génération (p. ex. pour la génération dynamique d'une page Web, Poesio), la recherche d'information et la recherche sur le Web (Kruschwitz, Robinson), la traduction automatique (Arnold, Sadler), les systèmes de dialogue parlé (Kruschwitz, Poesio).

Projets

- L'acquisition de connaissances lexicales et ontologiques
- La recherche intelligente sur le Web
- Statistiques de fiabilité pour l'annotation de corpus; sémantique formelle et interaction sémantique-prosodie; sémantique compositionnelle au-dessous du niveau des mots; coordination; qualification temporelle
- Sémantique formelle, théorie du droit de propriété, pluriels, sous-spécification

National Center for Text Mining

University of Manchester, Oxford Road, Manchester, UK, M13 9P

Tél. : (44) 161 275 6100

Site Web : <http://www.nactem.ac.uk/>

Personne-ressource : Julia Chruszcz

Courriel : julia.chruszcz@manchester.ac.uk

Nombre de postes : 53

Projets de recherche

Le National Centre for Text Mining est le premier centre d'exploration de textes au monde qui est financé par le secteur public en réponse aux demandes des universitaires du Royaume-Uni. Le centre apporte également d'importantes contributions aux chercheurs dans le domaine d'exploration de textes, aux échelles nationales et internationales.

WordNet en arabe – vise à élaborer une version de WordNet en arabe, après le processus d'élaboration de **Princeton WordNet** et **Euro WordNet**.

Bio-MITA – vise à promouvoir la découverte des connaissances biologiques par l'exploration de l'association de termes fondée sur les textes.

BOOTStrep – permet d'améliorer les ressources comme les ontologies et les dictionnaires, nécessaires pour l'exploration de textes dans le domaine biomédical.

JST Project – étudie l'utilisation des techniques d'apprentissage par machine pour l'élaboration de systèmes de groupement et de classification.

Manchester Centre for Integrative Systems Biology – dirigera le développement de nouvelles technologies dans la biologie des systèmes, en partie par l'utilisation d'outils d'exploration de textes.

ProFClass-TM – étudie l'utilisation de la classification automatique de textes pour aider à l'attribution des protéines dans les catégories fonctionnelles.

Natural Language and Computational Linguistics

Department of Informatics, University of Sussex, Falmer, Brighton BN1 9QH UK

Tél. : (44) 1273 678029 Téléc. : (44) 1273 877873

Site Web : <http://www.informatics.susx.ac.uk/research/nlp/>

Personne-ressource : John A. Carroll

Courriel : J.A.Carroll@sussex.ac.uk

Nombre de postes : 24

Projets de recherche

COGENT : Production contrôlée de textes : Nous développons dans ce projet des techniques réflexives pour contrôler efficacement la génération à grande échelle. Le projet vise à offrir de nouvelles ressources substantielles qui favoriseront la recherche à venir dans ce domaine, et l'application pratique de généralités contrôlées à grande échelle.

Natural Habitats : Ce projet permet d'étudier la façon dont les techniques de traitement du langage naturel peut rendre possible la composition de services aux utilisateurs non techniques, en mettant l'accent sur la mise au point d'un outil interactif de composition de services qui utilise une interface de langage naturel.

Ranking Word Senses for Disambiguation – Models and Applications : Ce projet vise à trouver de nouvelles façons d'estimer les distributions de fréquence du sens des mots à partir d'un texte brut (qui n'est pas annoté).

Projets antérieurs

MEANING : Developing Multilingual Web-scale Language Technologies. **DEEP THOUGHT**: Hybrid Deep and Shallow Methods for Knowledge-Intensive Information Extraction. Ce projet vise à étudier les méthodes de combiner les méthodes superficielles robustes pour l'analyse du langage avec un traitement sémantique profond.

LUCY : Le projet vise à élaborer une base de données de l'anglais moderne écrit qui est analysé structurellement.

PSET: Practical Simplification of English Text.

CHRISTINE : Le **CHRISTINE** Corpus comprend un échantillon annoté, représentatif socialement, de la parole spontanée actuelle pour créer des ressources visant à étudier la structure dans la langue britannique d'aujourd'hui.

LEXSYS : Analysis of Naturally-occurring English Text with Stochastic Lexicalized Grammars. Le projet a permis la mise au point d'un système robuste de passage à grande échelle pour le

texte anglais, en exploitant plusieurs techniques statistiques qui font appel au corpora en ligne, les hiérarchies d'héritage pour imposer la structure aux données du traitement du langage naturel et les grammaires lexicalisées.

POLYLEX : Le projet a permis de mettre au point un lexique trilingue à base d'héritage pour le vocabulaire de base du néerlandais, de l'anglais et de l'allemand au moyen de réseaux pour partager l'information entre les langues à tous les niveaux de la description linguistique.

SPARKLE : Shallow Parsing for Acquisition of Lexical Knowledge.

SUSANNE : Surface and Underlying Structural Analysis of Natural English.

Oxford Computational Linguistics Group

Oxford University Computing Laboratory Wolfson Building Parks Road Oxford,

OX1 3QD UK

Tél. : (44) 1865 283520

Site Web : <http://www.clg.ox.ac.uk/>

Personne-ressource : Stephen Clark

Courriel : Stephen.Clark@comlab.ox.ac.uk

Nombre de postes : 20

Projets de recherche

Les intérêts du groupe couvrent tout un éventail de recherches actives sur la linguistique computationnelle et l'ingénierie linguistique, de la théorie à l'application, y inclus :

Parsage statistique

Modèles de probabilité

Parsage statistique avec la Combinatory Categorical Grammar (site de CCG)

Analyse sémantique de grande envergure

Étiquetage statistique

Agents d'étiquetage d'entropie maximale

Étiquetage du mot d'ancrage, reconnaissance d'une entité nommée,

supertaggage CCG

Traduction automatique

TA par parsage statistique (2005 JHU Workshop)

TA illustrée

Acquisition de connaissances

Apprendre les théories de domaines à partir du texte

Estimer les probabilités sélectives au moyen de WordNet

Développement à grande échelle de la grammaire et du lexique

Concevoir des grammaires LFG à grande échelle pour plusieurs langages

Le projet ParGram

Représentation sémantique pour les langues naturelles

Logiques de premier ordre et à ordre élevé pour la représentation sémantique des langues naturelles

Méthodes d'inférence pour ces représentations

Composition sémantique pour le LFG en utilisant la logique linéaire

Langage naturel pour la représentation des connaissances

UCL Department of Phonetics and Linguistics

Department of Phonetics and Linguistics, University College London, Gower Street, London, WC1E 6BT, United Kingdom

Tél. : (44) 20 7679 7172 Téléc. : (44) 20 7679 3262

Site Web : <http://www.phon.ucl.ac.uk/index.html>

Personne-ressource : Stefanie D. Anyadi

Courriel : stefanie@ling.ucl.ac.uk

Nombre de postes : 36

Projets de recherche

Le thème de notre recherche qui porte sur la reconnaissance de la parole portait sur la façon dont les connaissances du traitement de la linguistique humaine peut aider la mise au point d'une reconnaissance par machine de la parole à deux unités d'exécution : l'application d'une analyse linguistique traditionnelle dans le décodeur et l'analyse du rendement de reconnaissance de la machine et de l'homme.

Les travaux récents sur l'**analyse linguistique** ont porté sur l'intégration de l'analyse morphologique dans le décodeur de reconnaissance de la parole. Grâce à l'analyse morphologique, les petits lexiques de prononciations sont possibles tout en augmentant la dissimilarité des prononciations. En association avec un modèle de langue au niveau lexical, nous avons montré que cette approche peut améliorer la précision du mot. Voir l'article par Huckvale & Fang (2001).

Les travaux de **comparaisons homme-machine** visaient l'étude des aspects psycholinguistiques du traitement morphologique humain. Des parallèles intéressants peuvent être tracés entre le dispositif de reconnaissance de morphe et les résultats des expériences d'amorçage sur les auditeurs humains. Les travaux les plus récents ont été une recherche de la façon dont les ordinateurs (qui font du reitement d'information) comprennent la perception (voir le **séminaire**)

Les travaux actuels de Gordon Hunter et Mark Huckvale étudient les **propriétés statistiques des mots dans les tours de dialogue** dans l'espoir de concevoir des modèles linguistiques supérieurs particulièrement pour le dialogue et pour trouver les liens entre les modèles statistiques et l'analyse conversationnelle.

En réalisant nos **travaux prosodiques**, nous cherchions à trouver la façon de représenter la structure prosodique du texte dans une représentation phonologique hiérarchique, en étudiant la formulation prosodique (couper le texte en phrases intonationnelles) et en catégorisant et allouant les accents dans les hauteurs tonales dans la phrase. Nos travaux d'interprétation de la phonétique de ces structures incluent la modélisation de contours de fréquence fondamentaux et la prédiction de la durée des composants syllabiques en tant que fonction du

contenu segmental et du contexte de la phrase.

Les **travaux à venir** se pencheront sur la façon dont l'intonation est utilisée pour exprimer la structure des phrases. Ces travaux serviront aux *systèmes de la parole* dans lesquels le texte à prononcer est également généré par le système informatique.

University Center for Computer Corpus Research on Language

InfoLab 21, South Drive, Lancaster University, Lancaster, LA1 4WA UK.

Tél. : (44) 1524 510357 Téléc. : (44) 1524 510492

Site Web : <http://www.comp.lancs.ac.uk/computing/research/ucrel/>

Personne-ressource : Paul Rayson

Courriel : paul@comp.lancs.ac.uk

Nombre de postes : 27

Projets de recherche

UCREL (the University Centre for Computer Corpus Research on Language) est un centre de recherche de l'Université Lancaster. Il s'appuie sur l'expertise du département de linguistique et de la langue anglaise moderne et du département d'informatique. Pendant plus de vingt ans, nous avons été le fer de lance d'une approche de traitement des langues naturelles qui est basée sur l'information dérivée d'importants corps de textes naturels. Ces corps de texte, connus sous corpora, sont stockés dans l'ordinateur. La grande majorité des travaux de l'UCREL sont effectués dans ce paradigme basé sur le corpus. Les corpora servent à dériver les connaissances empiriques sur la langue qui peuvent compléter, et souvent supplanter, l'information tirée de source de référence et de l'introspection (Leech 1991; 1992).

Eu égard à leur grande pertinence à l'analyse quantitative, corpora peut fournir des renseignements sur les fréquences relatives de plusieurs aspects de la langue. Ces fréquences peuvent alors servir aux techniques d'analyse probabilistique, une autre importante caractéristique des travaux de l'UCREL qui mettent beaucoup l'accent sur les résultats pratiques. Nous avons opté pour la recherche basée sur le corps afin de contribuer à des applications pratiques telles que : la synthèse de la parole, la reconnaissance de la parole, la publication de dictionnaire de traduction assistée par ordinateur, l'analyse des entrevues d'enquêtes sociales et l'enseignement de langues assisté par ordinateur.

Nos travaux mettent l'accent sur les domaines suivants :

Anglais – Nous avons été un premier partenaire au consortium British National Corpus que nous exploitons aujourd'hui pour trouver de nouvelles analyses puisées à même des données sur la parole et la rédaction britannique d'aujourd'hui. Nous effectuons également des travaux basés sur le corps dans le domaine du développement historique de l'anglais et de l'apprentissage de l'anglais.

Langues étrangères modernes – Nous avons conçu, annoté et exploité les corpora des langues modernes telles que le français et l'espagnol, et nous travaillons actuellement (conjointement avec l'Université de Lodz) à la production d'un corps majeur du polonais contemporain.

Langues de la minorité, menacées de disparaître et anciennes – Nous sommes des pionniers dans les travaux de corps relatifs aux langues de la minorité non indigène du Royaume-Uni (p. ex. chinois, hindi et punjabi). De plus, nous étendons nos travaux aux langues de la minorité

indigène de l'Europe. Nous avons également effectué des recherches linguistiques assistées par ordinateur sur les langues anciennes comme le latin.

Suède

Laboratoires

Department of Linguistics, Lund University

Center for Languages and Literature, Linguistics/Language Technology, Lund University, Helgonabacken 12, Sweden

Tél. : (46) 222 04 38 Téléc. : (46) 222 42 10

Site Web : <http://www.ling.lu.se/index.html?language=English>

Personne-ressource : Joost van de Weijer

Courriel : studierektor@ling.lu.se

Nombre de postes : 42

Projets de recherche

Le département de linguistique de l'Université Lund comporte deux sections : phonétique et linguistique générale. Les deux sections disposent de programmes de premier et de deuxième cycles. La section de phonétique a vu le jour en premier. Sa première chaire d'enseignement date de 1950. La section de la linguistique générale a suivi en 1969. En 1983, une chaire professorale et une section de recherche sur le langage des enfants ont été établies au département avant d'être remplacées en 1998 par une chaire professorale en acquisition du langage.

Actuellement, le département compte environ 50 employés, notamment les enseignants, les chercheurs, les étudiants au doctorat et le personnel technique et administratif.

Le département a un grand profil de recherche et d'enseignement. L'étude de différents aspects de la parole et du traitement des langues constitue un domaine d'intérêt majeur pour le département. Elle porte particulièrement sur les thèmes suivants : des études qui s'appuient sur la production et la perception pour examiner la prosodie incluant les applications technologiques de la parole, l'acquisition de la langue (notamment les aspects cognitifs du processus d'acquisition), le processus de lecture, les études comparatives de l'oral et l'écrit et les processus cognitifs qui les régissent. Le département compte également la typologie linguistique et la description des langues moins connues et de la minorité, notamment l'étude des langues en Asie, particulièrement du Sud-Est de l'Asie et des Caucases.

Department of Linguistics, Stockholm University

CL Group, Department of Linguistics, Stockholm University, S-106 09 Stockholm, Sweden

Tél. : (46) 8 16 20 00 Téléc. : (46) 8 15 53 89

Site Web : <http://www.ling.su.se/eng/>

Personne-ressource : Kristina Nilsson

Courriel : kristina.nilsson@ling.su.se

Nombre de postes : 84

Projets de recherche

Treebanks parallèles C'est dans le cadre d'efforts de recueillir et d'engager les ressources pour la création de systèmes de traitement de langues naturelles pour le suédois que nous contribuons à l'initiative de création d'une treebank suédoise. En outre, nous menons des expériences sur les treebanks parallèles anglais-allemand-suédois. Nous nous intéressons à la réduction de l'effort manuel de compiler une treebank par un approfondissement du treebank automatique et la projection de l'annotation. Nous travaillons sur le transfert des lignes directrices, des méthodes et des outils de l'anglais et de l'allemand au suédois, et nous créons des outils de recherche, de vérification et de visualisation pour des treebanks parallèles.

Extraire des ontologies à partir de corpora parallèle Il existe plusieurs situations dans lesquelles il est très important que deux agents (humain ou artificiel) sachent qu'ils parlent de la même chose. Les mésententes peuvent être engendrées lorsque les agents parlent différentes langues ou utilisent le même mot pour un sens différent (homonymie et polysémie). Pour résoudre ce problème, nous pouvons utiliser l'ontologie qui associe, lorsqu'elle est informelle, les termes d'un domaine avec des nœuds à des sens uniques, ordonnés dans une structure hiérarchique. Voici une autre définition de l'ontologie (from Chandrasekaran et al., 1999).

Résolution de coréférence pour l'accès à l'information La résolution de coréférence peut être définie comme étant l'identification de l'*identité* entre les entités dans le texte. Les deux entités sont identifiées par une reconnaissance d'entité nommée et des références anaphoriques. Ainsi, grâce à la résolution de coréférence, l'information éparpillée dans un texte (ou si elle est éparpillée dans une collection de textes, par la résolution de coréférence entre les documents) peut être associée aux entités auxquelles elle réfère.

Classification des noms de produits selon l'ontologie Ce projet vise à développer un système (dans n texte d'information sur une compagnie ou un produit) qui suggère un ou plusieurs nœuds dans une classification des produits (Vocabulaire commun pour les marchés publics, CPV).

Stockholm University Information System (SuiS) SuiS est un système d'information de langage croisée basé sur l'ontologie. Il permet des recherches en anglais et en suédois dans tous les documents indexés par Google dans le domaine de l'Université de Stockholm.

Graduate School of Language Technology

Graduate School of Language Technology, Faculty of Arts, Göteborg University, Box 200 SE-405 30, GÖTEBORG

Tél. : (46) 31 773 1000

Site Web : <http://www.gslt.hum.gu.se/>

Personne-ressource : Robin Cooper

Courriel : director@gslt.hum.gu.se

Nombre de postes : 83

Projets de recherche

Le Graduate School of Language Technology (GSLT) est basé à la Faculty of Arts, Université Göteborg. Il est le fruit d'une collaboration entre des centres de renommée dans la technologie linguistique en Suède. L'école vise à intégrer la recherche sur la parole et le langage et à fournir une base éclairée aux fondations théoriques et à la recherche orientée vers les applications.

- Recherche d'information relative au langage croisé de l'amarigna, l'anglais et le suédois;
- Analyse du phénomène de rétroaction dans la communication entre hommes et entre hommes et machines;
- Étiquetage des relations et des entités sémantiques.

Étudier si et comment certaines méthodes automatisées texte-linguistiques peuvent apporter plus d'outils pertinents à la recherche d'information et apporter des résumés uniformes plus adaptés aux interrogations et aux utilisateurs que les méthodes qui sont souvent comprises dans les systèmes d'information.

Interprétation sémantique et reconnaissance de la parole, interfaces de la parole pour des environnements dynamiques.

Interface-utilisateur multisensorielle du langage naturel entre un homme et un robot intelligent.

Amélioration du modèle vectoriel obtenu par LSI en utilisant l'information linguistique qui pourrait être extraite automatiquement à partir des données textuelles brutes.

Études portant sur les relations entre la structure de documents, le balisage et le genre dans lesquelles la relation entre la structure des documents et le balisage est effectuée selon des techniques d'apprentissage fondées sur la mémorisation, et la relation entre la structure des documents et le genre est effectuée selon le groupement et les techniques d'apprentissage de la machine.

Un modèle fondé sur les connaissances de la variation de prononciation du suédois standard central créé à partir de règles phonologiques.

Des systèmes multilingues et interculturels du dialogue parlé.

Indexation automatique.

**Human Computer Interaction and Language Engineering
Laboratory (HUMBLE), Swedish Institute of Computer Science**

HUMBLE, SICS, Uppsala Science Park, SE-751 83 Uppsala, Sweden

Tél. : (46) 8 633 15 00 Téléc. : (46) 8 751 72 30

Site Web : <http://www.sics.se/humble/>

Personne-ressource : Jussi Karlgren

Courriel : jussi@sics.se

Nombre de postes : 20

Projets de recherche

Les projets utilisent la technologie linguistique pour faire comprendre davantage certains aspects du texte au système. Nous étudions la structure thématique du texte lui-même. Pour ce faire, nous utilisons principalement la technologie d'extraction de l'information.. Nous étudions également l'accumulation des connaissances sémantiques en utilisant des algorithmes statistiques en collaboration avec le laboratoire ARC.

Dans la même veine, nous mettons l'accent sur l'importance d'une personnalisation systématique des outils à la tâche et au lieu de travail et à la proposition d'outils qui enrichissent la compétence professionnelle humaine. À cette fin, nous utilisons et développons des méthodes de rassemblement de connaissances importunes dans un milieu de travail.

Nous mettons actuellement l'accent sur certains thèmes de recherche concrets : amélioration des données, recherche d'information conjointe, recherche d'information multilingue et la gestion de thesaurus ou l'attribution des mots clé.

Interaction and Presentation Laboratory

Interaction and Presentation Laboratory, Department of Numerical Analysis and Computing
Science Royal Institute of Technology S-100 44 Stockholm, Sweden

Tél. : (46) 8 790 62 80 Téléc. : (46) 8 10 24 77

Site Web : <http://www.nada.kth.se/iplab/>

Personne-ressource : Ulla-Britt Lindqvist

Courriel : ubl@nada.kth.se

Nombre de postes : 43

Projets de recherche

IPLab a été créé en 1985 dans le département d'analyse numérique et d'informatique (NADA) du Royal Institute of Technology (KTH). Il s'appuie sur un groupe de chercheurs de diverses disciplines et d'étudiants en informatique, en linguistique, en psychologie, en sociologie et en conception. Ces deux groupes ont un intérêt commun pour l'interaction homme-informatique. Ci-dessous les projets de recherche actuels de IPLab :

Rédaction : Aide informatique pour l'arédaction et la lecture des processus

Dans le cadre de ce projet, une recherche réunissant diverses disciplines est menée sur la rédaction et la lecture assistées par ordinateur, ainsi que sur le résumage et la production de textes automatiques.

CSCW : travaux de collaboration assistés par ordinateur

Cette recherche porte sur des aspects de la conception et l'utilisation d'applications collectives ainsi que les systèmes de communication médiatisés par ordinateur.

POO : Programmation orientée objet pour les systèmes distribués

Cette recherche met l'accent sur les modes de conception pour les applications interactives et distribuées, particulièrement dans le domaine CSCW.

HRI : Interaction homme-robot

Ce domaine concerne l'interaction entre les hommes et les robots autonomes.

CL : Communication et apprentissage

Cette recherche porte sur les aspects fondamentaux de la communication et l'apprentissage humains, et la façon dont la technologie est utilisée dans différents contextes d'apprentissage.

PUI : Interfaces-utilisateurs perceptuelles

Cette recherche porte sur la conception d'interfaces-utilisateurs avec des capacités perceptuelles, c'est-à-dire des interfaces qui permettent aux utilisateurs de communiquer avec les ordinateurs au moyen de la parole ou de gestes faciaux ou du corps.

UCSD : Développement de système orienté utilisateur

Cette recherche met l'accent sur le développement d'un système orienté utilisateur.

Natural Language Processing Laboratory, Linköping University

Department of Computer and Information Science, Linköping University, S-583 81 Linköping, SWEDEN

Tél. : (46) 13 28 24 22 Téléc. : (46) 13 14 22 31

Site Web : <http://www.ida.liu.se/labs/nlplab/>

Personne-ressource : Lars Ahrenberg

Courriel : lah@ida.liu.se

Nombre de postes : 20

Projets de recherche

- Développement de ressources génériques pour la technologie langagière (LATOS);
- Interaction multimodale pour les services d'information (MIFIS);
- Systèmes multimodaux de dialogue (MDS) pour les applications industrielles;
- Système de traduction basé sur corpus (KOMA).

Speech and Communications Technology Group

Royal Institute of Technology (KTH) Dep. of Speech, Music and Hearing, Institutionen för Tal, Musik och Hörsel, Speech and Communications Tcehnology Group, Lindstedtsvägen 24 SE-100 44 STOCKHOLM, SWEDEN

Tél. : (46) 8 790 7879 Téléc. : (46) 8 790 7854

Site Web : <http://www.speech.kth.se/speech/>

Personne-ressource : Björn Granström

Courriel : bjorn@speech.kth.se

Nombre de postes : 33

Projets de recherche

- Reconnaissance du locuteur dans la page principale du projet de téléphonie.
- Reconnaissance basée sur la biométrie des personnes sur Internet.
- Interaction de la parole dans la télécommunication.
- Ordinateurs dans la boucle de l'interaction humaine.
- Mobilité universelle dynamique pour les interfaces de la parole adaptée.
- Un système d'enseignement de la technologie des systèmes du dialogue parlé.
- Initiative Web de voix sur HP.
- Visage parlant synthétisé dérivé de la parole à l'intention des durs d'oreille qui utilisent les voix téléphoniques.
- Page d'accueil du projet de synthèse multimodale de la parole.
- Interfaces multimodales et multisensorielles pour l'interaction avec les modèles musculaires squelettiques.

Suisse

Laboratoires

Artificial intelligence Laboratory, Federal Institute of Technology in Lausanne

Artificial Intelligence Laboratory, Institute of Core Computing Science, School of Computer and Communication Science, Swiss Federal Institute of Technology, IN (Ecublens), CH-1015 Lausanne (Switzerland)

Tél. : (41) 21 693 2735 Téléc. : (41) 21 693 5225

Site Web : <http://liawww.epfl.ch/>

Personne-ressource : Boi Faltings

Courriel : Boi.Faltings@epfl.ch

Nombre de postes : 20

Projets de recherche

La recherche dans le domaine du traitement des langues naturelles du Laboratoire d'intelligence artificielle (LIA) met l'accent sur l'exploration des textes (extraction des connaissances à partir de données textuelles), la production automatique d'outils de syntaxe et l'évaluation des outils NLP. Nos méthodes d'exploration des textes s'appuient sur des techniques conçues pour retirer l'information au moyen d'une approche sémantique distributionnelle. Dans de telles méthodes, les proximités sémantiques sont dérivées de matrices de co-fréquence calculées dans un large corpora textuel. Diverses mesures de similarité servent à caractériser la proximité entre les interrogations et les documents qui sont représentés d'une façon unifiée, entre autres les projections des termes pertinents dans un espace vectoriel à grandes dimensions. Les méthodes de production automatique des outils de syntaxe sont destinées à mettre appliquer les techniques et les modèles probabilistiques qui sont exécutés dans un corpora textuel (textes bruts ou annotés) afin d'adapter divers algorithmes génériques à des applications précises : étiquetage du mot d'ancrage, reconnaissance de la parole, recherche d'information, etc.

- Structuratin automatique de données textuelles : Applications à l'exploration de textes
- INSPECT – Intégration de modèles linguistiques acoustiques et avancés dans les systèmes de compréhension de la parole;
- STING : Évaluation de l'innovation scientifique et technologique et des progrès en Europe par des brevets;
- NLP-FPGA - coprocessor de Hardware NLP;
- INFOVOX – serveurs vocaux interactifs pour les applications avancées de la téléphonie informatique;
- EXTRACT : Extraction automatisée de l'information à partir d'annonces classées dans les journaux;
- GRACE – Évauation de l'étiquetage du mot d'ancrage;
- ELSE - Évaluation de l'ingénierie du langage et de la parole;

- Analyse syntaxique probabilistique axée sur les données;
- Semantique distributionnelle : Application à la recherche à partir de grandes bases textuelles;
- ISIS – Conception de serveurs vocaux avancés;
- Outils industriels pour le traitement des langues naturelles.

Département de linguistique, Université de Genève, Laboratoire d'analyse et de technologie du langage

2, rue de Candolle, CH - 1211 Genève 4 (Landolt, 7e étage, bureau L 705)

Tél. : 022 379 73 63 Téléc. : 022 379 79 31

Site Web : <http://www.latl.unige.ch/>

Personne-ressource : Eric Wehrli

Courriel : Eric.Wehrli@lettres.unige.ch

Nombre de postes : 20

Projets de recherche

Les travaux de recherche entrepris au LATL ces dernières années ont principalement porté sur des problèmes liés au traitement de la syntaxe des langues naturelles, à la fois dans une perspective de recherche pure, et dans une perspective d'application à des problèmes typiques des industries de la langue. Ces travaux ont abouti d'une part à l'élaboration d'un modèle cognitif d'analyse (cf. Walther 1997), et d'autre part à la réalisation d'analyseurs syntaxiques (anglais, allemand, français), s'appuyant sur des modèles chomskyens de linguistique formelle. Ces outils d'analyse, et les bases lexicales qui les accompagnent, constituent les outils fondamentaux sur lesquels s'articulent plusieurs projets de développement de prototypes et de systèmes de traitement de la langue, aussi bien écrite que parlée.

IPS (Interactive Parsing System)

Le projet IPS consiste à développer des analyseurs syntaxiques interactifs pour différentes langues, basés sur le modèle linguistique de la théorie GB. Les objectifs du projet IPS sont à la fois théorique et pratique. Le projet FipsVox (Gaudinat & Wehrli 1997) vise à développer un système de synthèse de la parole à partir d'un texte ('text-to-speech system'). Ce qui le distingue de la plupart des autres systèmes de synthèse du français, c'est avant tout sa base linguistique, à savoir l'analyseur syntaxique (F)IPS. À côté des objectifs d'ordre pratique (lecture de textes dans les formats MS-Word ou HTML), ce projet comporte également des préoccupations d'ordre théorique, notamment celle de montrer l'importance des données linguistiques pour le traitement de la parole.

En effet, malgré les progrès considérables accomplis récemment en synthèse de la parole, les systèmes actuels présentent encore des lacunes importantes, principalement en ce qui concerne les aspects prosodiques (intonation, durée, etc.), mais également au niveau du traitement des homographes et de celui de la liaison. Ces lacunes, largement responsables du caractère peu naturel de la voix de synthèse, peuvent porter préjudice à la compréhension d'un énoncé et limitent donc considérablement les possibilités d'application pratique des systèmes de synthèse vocale.

Institute of Computational Linguistics, University of Zurich

interthurerstr. 190, CH-8057 Zurich, Switzerland
University of Zurich, Institute of Computational Linguistics

Tél. : (41) 44 635 4331 Téléc. : (41)-44 635 6809

Site Web :

Personne-ressource : Manfred Klenner

Courriel : klenner@cl.unizh.ch

Nombre de postes : 20

Projets de recherche

Sémantique computationnelle : un thème unificateur

Le point de mire de recherche de notre groupe porte sur la *sémantique computationnelle*, c'est-à-dire sur les méthodes de traitement du sens des émissions de parole naturelles, soit une tâche aussi théorique que pratique. Sur le plan théorique, les modèles informatiques de la compréhension de la langue peuvent mettre en évidence sur la façon dont les hommes comprennent l'introduction orale des données, un élément crucial de l'intelligence humaine. Ce faisant, nous serions beaucoup plus sur l'intelligence en général.

Sur le plan qui nous intéresse davantage, c'est-à-dire pratique (application des méthodes), l'application éventuelle consisterait à disposer de systèmes qui répondent aux questions arbitraires (formulées dans un langage naturel) sur les textes, soit les « systèmes de réponse aux questions ». Ces systèmes seront extrêmement utiles pour les encyclopédies, les manuels techniques, les rapports administratifs, les documents juridiques, les rapports de service et d'essai, les journaux, les messages de distribution de nouvelles et même les dossiers documentaires.

Un système idéal de réponse aux questions serait en mesure de comprendre les textes dans leur intégralité avec toute l'information implicite qu'ils contiennent, de comprendre parfaitement la question (y compris la situation dans laquelle elle a été énoncée), et de dériver toutes les réponses correctes à partir de l'information textuelle. Il ne fait pas de doute, ce genre de système dépasse de loin ce qui est possible dans un avenir proche. Nous avons besoin de tâches plus simples et plus faciles à gérer; des tâches qui sont aussi intéressantes sur le plan théorique que pratique.

Organismes et associations

European Association for Machine Translation

EAMT 40, boulevard de Pont-d'Arve? CH-1221 Genève, Suisse

Télec. : (41) 22 379 8689

Site Web : <http://www.eamt.org/>

Personne-ressource : Bente Maegaard

Courriel : bente@cst.dk

Mission

La European Association for Machine Translation (EAMT) est une organisation au service de la communauté croissante des personnes intéressées par la traduction automatique (TA) et les outils de traduction, y compris les utilisateurs, promoteurs et chercheurs de cette technologie de plus en plus viable.

L'EAMT, de concert avec l'AMTA et l'AAMT, publie un bulletin, MTNI, et organise des ateliers de travail et des conférences, entre autres le sommet biennuel sur la traduction automatique. L'organisation compile également, par l'entremise de l'IAMT, les listes des entreprises et des produits qui sont distribuées gratuitement ou à un coût nominal à ses membres. L'EAMT détient également une liste de distribution, mt-list@eamt.org, en tant que forum public de discussion sur la technologie dans le domaine de la traduction.

European Association for Speech, Signal and Image Processing

EURASIP European Association for Signal, Speech, and Image Processing, EPFL-STI-LTS
Station 11 CH-1015 Lausanne Switzerland
Site Web : <http://www.eurasip.org>
Personne-ressource : Peter Grant
Courriel : info@eurasip.org

Mission

La European Association for Signal, Speech and Image Processing (EURASIP) a été fondée le 1^{er} septembre 1978 afin d'améliorer la communication entre les groupes et les individus qui travaillent dans le domaine pluridisciplinaire et dynamique du traitement des signaux en Europe et ailleurs d'une part, et d'échanger et diffuser l'information dans le domaine partout dans le monde. L'association européenne existe pour intensifier les efforts des chercheurs et fournir une plate-forme savante et professionnelle pour la diffusion et la discussion de tous les aspects du traitement des signaux.

Domaines d'intérêt

- Une théorie de signaux en temps discret et continu;
- Applications du traitement des données;
- Systèmes et technologie;
- Communication orale;
- Communication et traitement des images.