

TALN-RÉCITAL 2013

TALN : Traitement Automatique des Langues Naturelles

RÉCITAL : Rencontres des Étudiants Chercheurs en
Informatique pour le Traitement Automatique des Langues

Actes de la conférence TALN-RÉCITAL 2013

Volume 2 : RÉCITAL 2013

Éditeurs

Florian Boudin

Loïc Barrault



17 au 21 juin 2013
Les Sables d'Olonne, France

Sous l'égide de l'ATALA (Association pour le Traitement Automatique des langues).

Avant-propos

Il est maintenant une tradition dans la communauté de l'ATALA de venir fouler tous les dix ans les côtes à l'ouest de la France. Ainsi après la Côte d'Amour en 2003, nous sommes heureux d'accueillir sur la Côte de Lumière la 20^e conférence TALN et la 15^e édition de RÉCITAL.

L'organisation de TALN et RÉCITAL 2013 a été assurée par les équipes TALN du LINA (Laboratoire d'Informatique de Nantes Atlantique) et LST du LIUM (Laboratoire d'Informatique de l'Université du Maine). Cette organisation conjointe est une bonne illustration de la synergie de ces deux équipes mais aussi de la dynamique du TALN dans la région des Pays de la Loire.

Cette année, avec 127 soumissions à TALN (dont 70 articles longs et 57 articles courts), la conférence a confirmé une fois encore son attractivité. Le processus d'évaluation, qui a demandé un travail important, a été réalisé consciencieusement pour arriver à une sélection de 36 articles longs et 35 articles courts. Nous remercions chaleureusement les membres des comités de lecture et de programme de TALN pour le travail réalisé. Outre ces communications, 13 démonstrations et 4 ateliers viennent accompagner la conférence. La conférence sera aussi ponctuée par l'intervention de deux conférenciers invités : Josiane Mothe et Alexander Fraser que nous tenons aussi à remercier de leur présence.

Poursuivant sur la démarche initiée lors de l'édition précédente, la conférence RÉCITAL a ciblé un large spectre de publications (états de l'art, travaux préliminaires, etc.). L'accent a une nouvelle fois été mis sur la pédagogie et l'échange direct en fournissant des relectures explicatives et non anonymes aux auteurs. Cette formule fonctionne bien puisque l'on recense un total de 25 soumissions parmi lesquelles 18 ont été sélectionnées (6 présentations orales et 12 posters). Nous remercions les membres du comité de programme de RÉCITAL pour les précieux retours qui, nous en sommes convaincus, sont très appréciés des jeunes chercheurs en TALN.

Cette conférence sur un site distant, mais presque à mi-chemin entre Le Mans et Nantes, a nécessité un travail d'organisation important. Que les membres du comité d'organisation trouvent ici la reconnaissance du travail réalisé.

Nous n'oublions pas non plus les partenaires institutionnels et privés qui se sont joints à nous pour faire de cette 20^e conférence TALN et de cette 15^e édition de RÉCITAL une véritable réussite.

Comme il est de tradition aux Sables d'Olonne, nous souhaitons à l'ensemble des conférenciers « bon vent ».

Emmanuel Morin
Yannick Estève
Organisateurs de TALN 2013

Florian Boudin
Loïc Barrault
Organisateurs de RÉCITAL 2013

Comité d'organisation de TALN-RÉCITAL

Président de TALN

Emmanuel Morin LINA Université de Nantes, France

Vice-Président de TALN

Yannick Estève LIUM Université du Maine, France

Présidents de RÉCITAL

Florian Boudin LINA Université de Nantes

Loïc Barrault LIUM Université du Maine

Membres

Denis Béchet	LINA	Université de Nantes
Fethi Bougares	LIUM	Université du Maine
Adrien Bougouin	LINA	Université de Nantes
Nathalie Camelin	LIUM	Université du Maine
Béatrice Daille	LINA	Université de Nantes
Colin De La Higuera	LINA	Université de Nantes
Paul Deléglise	LIUM	Université du Maine
Estelle Delpech	LINA	Université de Nantes
Alexandre Dikovskiy	LINA	Université de Nantes
Chantal Enguehard	LINA	Université de Nantes
Rima Harastani	LINA	Université de Nantes
Mohamed Hatmi	LINA	Université de Nantes
Amir Hazem	LINA	Université de Nantes
Nicolas Hernandez	LINA	Université de Nantes
Firas Hmida	LINA	Université de Nantes
Christine Jacquin	LINA	Université de Nantes
Ophélie Lacroix	LINA	Université de Nantes
Antoine Laurent	LIUM	Université du Maine
Elizaveta Loginova	LINA	Université de Nantes
Daniel Luzzati	LIUM	Université du Maine
Sylvain Meignier	LIUM	Université du Maine
Laura Monceaux	LINA	Université de Nantes
Simon Petitrenaud	LIUM	Université du Maine
Emmanuel Planas	LINA	Université Catholique de l'Ouest
Solen Quiniou	LINA	Université de Nantes
Anne-Françoise Quin	LINA	Université de Nantes
Holger Schwenk	LIUM	Université du Maine
James Scicluna	LINA	Université de Nantes
Christophe Servan	LIUM	Université du Maine
Prajol Shrestha	LINA	Université de Nantes
Déborah Sourdillat	LINA	Université de Nantes
Annie Tartier	LINA	Université de Nantes

Comité de programme RÉCITAL

Florian Boudin	LINA, Université de Nantes
Loïc Barrault	LIUM, Université du Maine
Adrien Lardilleux	Affinity Engine
Amir Hazem	LINA, Université de Nantes
Anne Vilnat	LIMSI, Université Paris-Sud
Antoine Jean-Yves	LI, Université de Tours
Aurélien Max	LIMSI, Université Paris-Sud
Benoît Favre	LIF, Université d'Aix-Marseille
Cécile Fabre	ERSS, Université Toulouse 2
Cédric Lopez	VISEO, Grenoble
Charlotte Roze	ALPAGE, INRIA Paris–Rocquencourt & Univ. Paris 7
Christian Raymond	IRISA, Université de Rennes
Christine Jacquin	LINA, Université de Nantes
Christophe Servan	LIUM, Université du Maine
Crabbé Benoît	ALPAGE, INRIA Paris–Rocquencourt & Univ. Paris 7
Denis Maurel	LI, Université de Tours
Denis Bechet	LINA, Université de Nantes
Didier Schwab	LIG, Université Grenoble 2
Farah Benamara	IRIT, Université de Toulouse
Guy Lapalme	RALI, Université de Montréal, Canada
Jorge Mauricio Molina Mejia	LIDILEM, Université Stendhal – Grenoble 3
Juan-Manuel Torres-Moreno	LIA, Université d'Avignon et des Pays de Vaucluse
Mathieu Lafourcade	LIRMM, Université de Montpellier 2
Mathieu Roche	LIRMM, Université de Montpellier 2
Nicolas Hernandez	LINA, Université de Nantes
Nuria Gala	LIF, Université d'Aix-Marseille
Olivier Kraif	LIDILEM, Université Grenoble 3
Philippe Langlais	RALI, Université de Montréal, Canada
Raphael Rubino	School of Computing, Dublin City University, Irlande
Romain Deveaud	LIA, Université d'Avignon et des Pays de Vaucluse
Solen Quiniou	LINA, Université de Nantes
Stéphane Huet	LIA, Université d'Avignon et des Pays de Vaucluse
Thierry Hamon	LIM&BIO, Université Paris 13
Violaine Prince	LIRMM, Université Montpellier 2
Yayoi Nakamura-Delloye	LCAO, Université Paris 7

Partenaires

Or



Argent



Bronze



Institutionnels



Table des matières

Articles longs	1
<i>Acquisition de lexique bilingue d'expressions polylexicales : Une application à la traduction automatique statistique</i> Dhouha Bouamor	1
<i>Quelques variations sur les mesures de comparabilité quantitatives et évaluations sur des corpus comparables Français-Anglais synthétiques</i> Guiyao Ke.....	15
<i>Inférence grammaticale guidée par clustering</i> Noémie-Fleur Sandillon-Rezer.....	28
<i>Améliorer l'extraction et la description d'expressions polylexicales grâce aux règles transformationnelles</i> Aurélie Joseph	42
<i>Construction de corpus multilingues : état de l'art</i> Manuela Yapomo	56
<i>Développement de ressources pour l'entraînement et l'utilisation de l'étiqueteur morpho-syntaxique TreeTagger sur l'arabe</i> Dhaou Ghoul	69
<i>Détection de polarité d'opinions dans les forums en langue arabe par fusion de plusieurs SVM</i> Amel Ziani, Nabih Azizi et Yamina Tlili Guiassa	83
<i>État de l'art des méthodes d'extraction automatique de termes-clés</i> Adrien Bougouin.....	96
<i>Influence de l'étiquetage syntaxique des têtes sur l'analyse en dépendances discontinues du français</i> Ophélie Lacroix	110
<i>Une approche linguistique pour l'extraction des connaissances dans un texte arabe</i> Houda Saadane	124
<i>Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines : une première expérimentation</i> Sylvain Hatier	138
<i>Vers une identification automatique du chiasme de mots</i> Marie Dubremetz	150

<i>Représentation des connaissances du DEC : Concepts fondamentaux du formalisme des Graphes d'Unités</i>	
Maxime Lefrançois.....	164
<i>Vers un système générique de réécriture de graphes pour l'enrichissement de structures syntaxiques</i>	
Corentin Ribeyre	178
<i>État de l'art de l'induction de sens : une voie vers la désambiguïisation lexicale pour les langues peu dotées</i>	
Mohammad Nasiruddin	192
<i>Génération des corpus en dialecte tunisien pour la modélisation de langage d'un système de reconnaissance</i>	
Rahma Boujelbane	206
<i>Détection automatique des sessions de recherche par similarité des résultats provenant d'une collection de documents externe</i>	
Simon Leva et Nicolas Faessel	217
<i>Une approche mixte morpho-syntaxique et statistique pour la reconnaissance d'entités nommées en langue chinoise</i>	
Zhen Wang	231
Liste des auteurs	244
Liste des mots clés	245