

Lundi, 25 mars 2019

EARIA'2019: Recherche d'Information et Apprentissage Automatique

8:30	OUVERTURE EARIA'19 Lynda Tamine, IRIT	14:30	Session 4 – Séance pratique
8:45	Session 1 – Introduction à la Recherche d'Information Mohand Boughanem, IRIT	15:00	Benjamin Piwowski, LIP6 José Moreno, IRIT
9:00		15:15	
9:15		15:30	
9:30		15:45	
9:45	Session 2 - Apprentissage pour la recherche d'information Eric Gaussier, LIG	16:00	Pause
10:00		16:15	
10:15		16:30	
10:30		16:45	
10:45	Pause	17:00	Session 5 – Présentation Hackathon et lancement d'activités Benjamin Piwowski, LIP6 José Moreno, IRIT
11:00		17:15	
11:15		17:30	
11:30		17:45	
11:45	Session 3 – Plongements lexicaux et réseaux de neurones Benjamin Piwowski, LIP6	18:00	Dîner
12:00		18:45	
12:15		19:00	
12:30		19:15	
12:45	Déjeuner	19:30	Session 6 – Hackathon Benjamin Piwowski, LIP6 José Moreno, IRIT
13:00		19:45	
13:30		20:00	
14:00		20:30	
14:15		21:00	
		21:30	
		22:00	
		22:30	

Mardi, 26 mars 2019

EARIA'2019: Recherche d'Information et Apprentissage Automatique

8:30	Session 7 – Machine learning for natural language understanding Marie-Francine Moens, Univ. Louven
8:45	
9:00	
9:15	
9:30	
9:45	Session 8 – Le Traitement Automatique des Langues appliqué à la recherche d'information Christophe Servan, Qwant
10:00	
10:15	
10:30	
10:45	
11:00	Pause
11:15	
11:30	Session 9 – Apprentissage profond d'agents conversationnels Pierre-Emmanuel Mazaré, Facebook
11:45	
12:00	
12:15	
12:30	
12:45	
13:00	Déjeuner
13:15	
13:30	
13:45	
14:00	
14:15	

14:30	Session 10 – Hackathon Benjamin Piwowski, LIP6 José Moreno, IRIT
14:45	
15:00	
15:15	
15:30	
15:45	
16:00	
16:15	Pause
16:30	
16:45	Session 11 – Hackathon Benjamin Piwowski, LIP6 José Moreno, IRIT
17:00	
17:15	
17:30	
17:45	
18:00	
18:45	
19:00	
19:15	Dîner
19:30	
20:00	
20:30	
21:00	



CORIA
EARIA
Lyon'19

Conférenciers/Orateurs



Marie-Francine Moens is full professor at the Department of Computer Science of KU Leuven and head of the Informatics section of this department. She is also the director of the Language Intelligence & Information Retrieval (LIIR) lab. She holds a M.Sc. and a Ph.D. degree in Computer Science from KU Leuven. She is author of more than 300 international peer-reviewed publications and of several books. In 2011 and 2012 she was appointed as chair of the European Chapter of the Association for Computational Linguistics (EACL). She was the scientific manager of the EU COST action iV&L Net (The European Network on Integrating Vision and Language, 2014-2018, IC1307). She is associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). She was appointed as Scottish Informatics and Computer Science Alliance (SICSA) Distinguished Visiting Fellow in 2014. She is holder of the 2017 ERC (European Research Council) Advanced Grant CALCULUS which researches the topic of language understanding.



Mohand Boughanem is a Professor at the University of Toulouse III he received his PhD from university of Toulouse in 1992. He was responsible of the Indexing and Information retrieval theme at IRIT. His current research focuses on IR (Information Retrieval) models, Contextual/personalized/Social based IR, aggregated search, Opinion mining. He has been involved in a number of projects and working groups on IR, and he has regularly participated in major IR evaluation forums (TREC, CLEF, INEX). He has served as a program committee member or chairman of the major IR conferences. He has been invited as a reviewer by several related-IR journals (IR journal, IP&M, JASIST, ACM-Tran. On the Web, ACM TOIS)). He founded ARIA (the French Association of Information Retrieval) and CORIA (the annual French conference in information retrieval), both were founded in 2004. He chaired the ARIA during 2004-2007. He is coordinator of the French-Brazilian working group on web science. He published more than 100 papers in international conferences and journals and he advised more than 25 Phd thesis



Benjamin Piwowarski est un chargé de recherche 1ère classe du CNRS au LIP6 (Sorbonne Université). Il a été fortement impliqué dans l'initiative INEX (2004-07), et a organisé un workshop (IRGM'07), deux tutoriels (ECIR 2012 et ICTIR 2013), a participé à l'organisation de CORIA 2015, et organise actuellement la conférence SIGIR 2019. Son domaine de recherche est celui de l'accès à l'information et de l'apprentissage statistique appliqué, en particulier l'apprentissage de représentations. Il a travaillé sur la recherche d'information XML (thèse 1999-2003), sur les bases de données XML (post-doctorat 2004-06 à l'université du Chili), sur l'analyse des interactions entre un utilisateur et un moteur de recherche (Yahoo! 2006-08) et sur l'application du formalisme mathématique de la physique quantique aux modèles d'accès à l'information (2008-11). Ses intérêts actuels portent sur (l'apprentissage de) représentation de données textuelles/sociales pour l'accès à l'information, et en particulier les applications à la représentations des mots ancrée dans le réel (grounding), le résumé automatique, l'extraction d'information et la recommandation.



Eric Gaussier a fait sa thèse au Centre Scientifique d'IBM France et à l'Université Paris 7 sur les modèles probabilistes pour l'extraction de lexiques bilingues. Il a rejoint le Centre de Recherche Européen de Xerox en 1996 où il a dirigé une équipe de recherche sur l'Apprentissage Automatique pour l'Accès à l'Information. En septembre 2006, il a rejoint l'Université Joseph Fourier et le Laboratoire d'Informatique de Grenoble (LIG). Il est membre de l'équipe "dAta analysis, Modeling and mAchine learning" (AMA). Il est actuellement directeur du LIG. Ses recherches portent sur l'apprentissage automatique et la modélisation de l'information. Il a été membre, de 2007 à 2010, de l'Information Sciences panel for starting grants of the European Research Council et de l'Executive Board of the European Association for Computational Linguistics. Il est également membre, depuis 2005, de l'Advisory Board of SIGDAT.



José Moreno a fait sa thèse à l'Université Caen Normandie et est actuellement maître de conférence à l'Université Paul Sabatier de Toulouse. Il est membre de l'équipe "Information Retrieval and Information Synthesis" (IRIS) de l'Institut de Recherche en Informatique de Toulouse.



Pierre-Emmanuel Mazaré a rejoint Facebook en 2017. Depuis 2018, il est en charge des équipes d'ingénierie du centre de recherche en intelligence artificielle de Paris. Il y assure l'encadrement technique de projets allant de la vision assistée par ordinateur au traitement du langage naturel. Diplômé de l'École Polytechnique et de l'université de Californie à Berkeley, il était auparavant développeur puis chef d'équipe responsable des systèmes de recommandation à Criteo de 2012 à 2017.



CORIA
EARIA
Lyon'19

Résumés

Apprentissage pour la recherche d'information

Eric Gaussier, LIG

🕒 Lundi, 25 mars 2019, 9h45-11h00

Nous aborderons dans cet exposé les principes fondamentaux de l'apprentissage et leur mise en œuvre dans le cadre de la RI. Nous nous intéresserons en particulier à l'apprentissage d'ordonnancement et à son utilisation en RI et verrons quelques applications de l'apprentissage profond à la RI.

Plongements lexicaux et réseaux de neurones

Benjamin Piwowarski, LIP6

🕒 Lundi, 25 mars 2019, 11h30-12h45

Dans ce cours, nous nous intéresserons à l'étude des techniques liées au *deep learning*, et nous nous intéresserons donc aux modèles basés sur les réseaux de neurones pour l'accès à l'information. En particulier, nous aborderons les points suivants :

- 1) Représentation de mots : motivations et modèles ;
- 2) Comment représenter les textes dans un espace sémantique: modèles récurrents, convolutionnels, d'attention/interaction ;
- 3) Applications - Modèle neuronal pour l'accès à l'information: recherche, question-réponse, résumé automatique et analyse d'utilisateurs

Machine learning for natural language understanding

Marie-Francine Moens, Univ. Louven

🕒 Mardi, 26 mars 2019, 8h30-9h45

Human language understanding by a machine, for instance, when reading text, answering questions, following instructions, or conducting dialogues, is a long-standing goal of intelligent systems. Recent advances in machine learning have made this goal gradually achievable. In this tutorial we explain state-of-the-art approaches to semantic parsing, semantic role labeling, coreferent resolution and temporal and spatial recognitions and illustrate the methods in machine reading and question answering tasks. Important research questions are how to represent the meaning of language and how to translate between discrete and continuous meaning representations. In a second part, we go deeper into questions of how to ground the meaning representations in the real perceptual world, what properties the machine learning models should have, and how to measure the success of language understanding.

Le Traitement Automatique des Langues appliqué à la recherche d'information

Christophe Servan, Qwant

🕒 Mardi, 26 mars 2019, 9h45-11h00

Dans ce cours, nous aborderons sous forme d'une Etat-de-l'Art les principales techniques utilisées en recherche d'information issues du TAL.

Nous explorerons les dernières approches avec pour axes :

- le traitement de la requête ;
- le traitement des documents ;
- les traitements multilingues.

Le fil rouge de ce cours sera le « un passage à l'échelle respectant la vie privée des utilisateurs », la principale problématique chez Qwant Research.

Apprentissage profond d'agents conversationnels

Pierre-Emmanuel Mazaré, Facebook

🕒 Mardi, 26 mars 2019, 11h30-12h45

Dans cette présentation nous traitons des différentes tâches d'agents conversationnels : réponse automatisée aux questions, dialogue de "bavardage" et résumé automatique. Pour chacun de ces domaines nous faisons un état des lieux des jeux de données ouverts, des modèles et des méthodes d'évaluation, et discutons des applications potentielles dans des agents conversationnels autonomes ("chatbots").



CORIA
EARIA
Lyon'19

Sponsors

LIRIS



INSA INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
LYON



GRAND LYON
la métropole



SIGIR
Special Interest Group
on Information Retrieval



INFOLOGIC
ENGINEERING



Qwant



ED 512



Ecole doctorale



INSA Département
Informatique