

## INTRODUCTION

---

Ce numéro spécial est constitué d'articles issus de la 13<sup>e</sup> Conférence en Recherche d'Information et Applications (CORIA) et du 14<sup>e</sup> Colloque International Francophone sur l'Écrit et le Document (CIFED) organisés dans le cadre de la semaine du Document Numérique et de la Recherche d'Information (SDNRI) du 9 au 11 mars 2016 à Toulouse organisée par l'IRIT.

CORIA et CIFED sont les points de rassemblement des communautés francophones respectivement en recherche d'information et en analyse de l'écrit et des documents numérisés. Tout en préservant les spécificités de chaque conférence, cette édition constitue une opportunité pour les deux communautés de se retrouver autour de thématiques pour lesquelles il existe des synergies de recherche (recherche de documents multimédia, modèle d'interaction avec l'utilisateur, passage à l'échelle de système de recherche d'information, outils d'évaluation de performance pour la recherche d'information).

Les auteurs des papiers ont été invités à rédiger une version longue qui a été évaluée par un nouveau comité de lecture. Les sept articles retenus et publiés dans ce numéro spécial sont les suivants.

Dans le premier article, « Bandits contextuels pour la capture de données temps réel », Thibault Gisselbrecht, Patrick Gallinari et Sylvain Lamprier proposent une approche qui permet de sélectionner des profils de comptes susceptibles de publier des contenus pertinents. La tâche est formalisée comme un problème de bandit contextuel. Des expérimentations sont menées avec succès sur Twitter.

L'article intitulé « Apprentissage de représentations pour la modélisation de processus de diffusion dans les réseaux sociaux », proposé par Simon Bourigault, Sylvain Lamprier et Patrick Gallinari, modélise la diffusion d'information sur les réseaux sociaux pour prédire et pour expliquer. L'approche se base sur un modèle d'automate cellulaire probabiliste en cascade.

L'article intitulé « Un modèle probabiliste pour la détection de l'incertitude dans le langage naturel » est proposé par Pierre-Antoine Jean, Sébastien Harispe, Sylvie Ranwez, Patrice Bellot et Jacky Montmain. Il présente un modèle fondé sur l'apprentissage supervisé pour classer des phrases de textes en « incertaines/certaines ». Le modèle est testé sur des données de différents challenges.

L'article « Indiscriminabilité dans les espaces de représentation des termes et des documents » de Vincent Claveau, propose de revisiter la notion de dimension intrinsèque des documents. Il montre que sa proposition aide à caractériser les requêtes difficiles et à choisir au mieux les mots à étendre dans un contexte d'extension de requêtes.

L'article « Exploitation de l'échelle d'écriture pour améliorer la reconnaissance automatique des textes manuscrits arabes » est écrit par Edgard Chammas, Chafic Mokbel et Laurence Likforman-Sulem. Les auteurs présentent une étude de l'impact de la taille de l'écriture en entrée de systèmes de reconnaissance d'écriture manuscrite arabe à base de HMM et de BLSTM.

L'article intitulé « Un modèle syllabique du français et de l'anglais pour la reconnaissance de l'écriture » est proposé par Wassim Swaileh et Thierry Paquet. Cet article présente une méthode de syllabation orthographique supervisée pour la construction d'un vocabulaire de syllabes. Une évaluation est réalisée sur l'anglais et le français.

Le dernier article, « Using BLSTM for interpretation of 2-D languages. Case of handwritten mathematical expressions » est écrit par Ting Zhang, Harold Mouchère et Christian Viard-Gaudin. Les auteurs proposent une extension de l'utilisation des réseaux de type BLSTM pour traiter des données provenant de langages graphiques bidimensionnels tels que les formules mathématiques manuscrites saisies à l'aide d'un stylo numérique.

Concernant le processus de sélection, les articles soumis à ce numéro spécial de la revue Document Numérique ont fait l'objet d'une phase d'évaluation réalisée par au moins deux relecteurs qui, par leurs commentaires, ont participé à la qualité des articles retenus. Nous souhaitons donc les remercier pour leur participation à ce numéro spécial comme évaluateur. Nous remercions également les auteurs qui ont retravaillé leurs articles, les membres des comités de programmes de CORIA et CIFED pour la qualité de leurs relectures, ainsi que tous les contributeurs à ces deux conférences qui ont permis de faire de ces événements scientifiques nationaux un moment important de rencontre des communautés recherche d'information et analyse de l'écrit et du document.

Sylvie Calabretto  
LIRIS, INSA Lyon

Bertrand Couasnon  
IRISA, INSA Rennes

### **Comité de Lecture**

Nous remercions très chaleureusement le comité de lecture de ce numéro spécial qui a effectué un précieux travail de relecture.

### **Pour CORIA**

Max Chevalier – IRIT, Toulouse.

Eric Gaussier – LIG, Grenoble.

Mathias Géry – LHC, Saint-Etienne.

Léa Laporte – LIRIS, Lyon.

José Martinez – LINA, Nantes.

Philippe Mulhem – LIG, Grenoble.

Pierre-Edouard Portier – LIRIS, Lyon.

Jean-Michel Renders – Xerox Research Centre Europe

Isabelle Tellier – LaTTiCe, Paris.

Haïfa Zargayouna – LIPN, Paris.

### **Pour CIFED**

Clément Chatelain – LITIS, Rouen

Pierre Héroux – LITIS, Rouen

Slim Kanoun – ENIS, Sfax, Tunisie

Harold Mouchère – IRCCyN, Nantes

Thierry Paquet – LITIS, Rouen

Laurent Wendling – LIPADE - Paris

