

# Florian Boudin

Last update: January 19, 2022

CONTACT INFORMATION	LS2N - Université de Nantes 2 rue de la Houssinière 44322 Nantes, France	+33 2 51 12 59 98 florian.boudin@univ-nantes.fr <a href="http://florianboudin.org">http://florianboudin.org</a>
EMPLOYMENT HISTORY	<b>Associate Professor</b> Université de Nantes, France <b>Assistant Professor</b> Université d'Avignon et des Pays de Vaucluse, France <b>Postdoctoral Fellow</b> Université de Montréal, Canada <b>Research Assistant</b> Université d'Avignon et des Pays de Vaucluse, France	2011 to present 2010-2011 2009-2010 2006-2009
EDUCATION	<b>Ph.D. in Computer Science</b> Université d'Avignon et des Pays de Vaucluse, France Advisors: Juan-Manuel Torres-Moreno, Marc El-Bèze <b>M.S. in Computer Science</b> Université d'Avignon et des Pays de Vaucluse, France <b>B.S. in Mathematics and Computer Science</b> Université d'Avignon et des Pays de Vaucluse, France University of Birmingham, UK (ERASMUS Programme)	2006-2008 2004-2006 2000-2004
AWARDS AND HONORS	<b>Prime d'Encadrement Doctoral et de Recherche</b> Bonus for doctoral supervision and research activity <b>Atlantic 2020 sabbatical grant for 6 months</b> National Institute of Informatics (NII), Tokyo	2017-2021 2019
STUDENTS	<b>Ph.D. Students</b> <ul style="list-style-type: none"><li>Maël Houbre (expected graduation ~ 2025)</li><li>Ygor Gallina (expected graduation ~ 2022)</li><li>Adrien Bougouin (2015; now software engineer at WOVN.io)</li></ul> <b>M.S. Students</b> <ul style="list-style-type: none"><li>Timothée Poulain (2020)</li><li>Rémi Bois (2014; now data scientist at Ouest France)</li></ul>	
PUBLICATIONS	In Natural Language Processing and Information Retrieval, conferences are more competitive, more highly regarded, and more highly cited than journals. Top-tier NLP/IR conferences: ACL (1 paper), NAACL-HLT (4 papers), EMNLP (2 papers), COLING (3 papers), ECIR (2 papers) and JCDL (2 papers). According to Google Scholar, my five most cited papers are: [22] (289 citations), [20] (143 citations), [6] (132 citations), [14] (101 citations) and [21] (87 citations). <b>Journal Papers</b> <ul style="list-style-type: none"><li>[1] Kenichi Iwatsuki, Florian Boudin, and Akiko Aizawa. Extraction and evaluation of formulaic expressions used in scholarly papers. <i>Expert Systems with Applications</i>, 187. 2022.</li><li>[2] Béatrice Daille Adrien Bougouin Florian Boudin. Modélisation à base de graphe pour l'indexation en domaines de spécialité. <i>Recherche d'information, document et web sémantique</i>, 17:1. 2017.</li></ul>	

- [3] Adrien Bougouin, Florian Boudin, Béatrice Daille, Sabine Barreaux, Damien Cram, and Amir Hazem. Indexation d'articles scientifiques Présentation et résultats du défi fouille de textes DEFT 2016. *Recherche d'information, document et web sémantique*, 17:1. 2017.
- [4] Adrien Bougouin and Florian Boudin. TopicRank : ordonnancement de sujets pour l'extraction automatique de termes-clés. *Traitement Automatique des Langues*, 55:1. 2014.
- [5] Romain Deveaud and Florian Boudin. De quoi parle ce Tweet? Résumer Wikipédia pour contextualiser des microblogs. *The Information - Intelligence - Interaction (I3) Journal*. 2014.
- [6] Florian Boudin, Jian-Yun Nie, Joan Bartlett, Roland Grad, Pierre Pluye, and Martin Dawes. Combining classifiers for robust PICO element detection. *BMC Medical Informatics and Decision Making*, 10:29. 2010.

## Conference Papers

- [7] Florian Boudin. ACM-CR: A Manually Annotated Test Collection for Citation Recommendation. *Joint Conference on Digital Libraries (JCDL)*. 2021.
- [8] Florian Boudin and Ygor Gallina. Redefining Absent Keyphrases and their Effect on Retrieval Effectiveness. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 2021.
- [9] Florian Boudin, Béatrice Daille, Evelyne Jacquey, and Jian-Yun Nie. The DELICES Project: Indexing Scientific Literature Through Semantic Expansion. *Joint Conference of the Information Retrieval Communities in Europe (CIRCLE)*. 2020.
- [10] Florian Boudin, Ygor Gallina, and Akiko Aizawa. Keyphrase Generation for Scientific Document Retrieval. *Association for Computational Linguistics (ACL)*. 2020.
- [11] Ygor Gallina, Florian Boudin, and Béatrice Daille. Large-Scale Evaluation of Keyphrase Extraction Models. *Joint Conference on Digital Libraries (JCDL)*. 2020.
- [12] Kenichi Iwatsuki, Florian Boudin, and Akiko Aizawa. An Evaluation Dataset for Identifying Communicative Functions of Sentences in English Scholarly Papers. *Language Resources and Evaluation Conference (LREC)*. 2020.
- [13] Ygor Gallina, Florian Boudin, and Béatrice Daille. KPTime: A Large-Scale Dataset for Keyphrase Generation on News Documents. *International Conference on Natural Language Generation (INLG)*. 2019.
- [14] Florian Boudin. Unsupervised Keyphrase Extraction with Multipartite Graphs. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 2018.
- [15] Florian Boudin. pke: an open source python-based keyphrase extraction toolkit. *International Conference on Computational Linguistics (COLING)*. 2016.
- [16] Adrien Bougouin, Sabine Barreaux, Laurent Romary, Florian Boudin, and Béatrice Daille. TermITH-Eval: a French Standard-Based Resource for Keyphrase Extraction Evaluation. *Language Resources and Evaluation Conference (LREC)*. 2016.
- [17] Adrien Bougouin, Florian Boudin, and Béatrice Daille. Keyphrase Annotation with Graph Co-Ranking. *International Conference on Computational Linguistics (COLING)*. 2016.
- [18] Florian Boudin, Hugo Mougard, and Benoit Favre. Concept-based Summarization using Integer Linear Programming: From Concept Pruning to Multiple Optimal Solutions. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2015.
- [19] Ophélie Lacroix, Denis Béchet, and Florian Boudin. Label Pre-annotation for Building Non-projective Dependency Treebanks for French. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*. 2014.
- [20] Florian Boudin. A Comparison of Centrality Measures for Graph-Based Keyphrase Extraction. *International Joint Conference on Natural Language Processing (IJCNLP)*. 2013.
- [21] Florian Boudin and Emmanuel Morin. Keyphrase Extraction for N-best Reranking in Multi-Sentence Compression. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 2013.

- [22] Adrien Bougouin, Florian Boudin, and Béatrice Daille. TopicRank: Graph-Based Topic Ranking for Keyphrase Extraction. *International Joint Conference on Natural Language Processing (IJCNLP)*. 2013.
- [23] Florian Boudin, Jian-Yun Nie, and Martin Dawes. Using a Medical Thesaurus to Predict Query Difficulty. *European Conference on Information Retrieval (ECIR)*. 2012.
- [24] Florian Boudin, Stéphane Huet, and Juan-Manuel Torres-Moreno. A Graph-based Approach to Cross-language Multi-document Summarization. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*. 2011.
- [25] Florian Boudin, Jian-Yun Nie, and Martin Dawes. Clinical Information Retrieval using Document and PICO Structure. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 2010.
- [26] Florian Boudin, Jian-Yun Nie, and Martin Dawes. Positional Language Models for Clinical Information Retrieval. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2010.
- [27] Florian Boudin, Lixin Shi, and Jian-Yun Nie. Improving Medical Information Retrieval with PICO Element Detection. *European Conference on Information Retrieval (ECIR)*. 2010.
- [28] Florian Boudin, Marc El-Bèze, and Juan-Manuel Torres-Moreno. A Scalable MMR Approach to Sentence Scoring for Multi-Document Update Summarization. *International Conference on Computational Linguistics (COLING)*. 2008.
- [29] Florian Boudin, Juan Torres-Moreno, and Marc El-Bèze. Mixing Statistical and Symbolic Approaches for Chemical Names Recognition. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*. 2008.
- [30] Florian Boudin, Juan-Manuel Torres-Moreno, and Patricia Velázquez-Morales. An Efficient Statistical Approach for Automatic Organic Chemistry Summarization. *International Conference on Natural Language Processing (GoTAL)*. 2008.
- [31] Florian Boudin and Juan Torres Moreno. NEO-CORTEX: A Performant User-Oriented Multi-Document Summarization System. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*. 2007.
- [32] Florian Boudin and Juan-Manuel Torres-Moreno. A Cosine Maximization Minimization approach for User Oriented Multi-Document Update Summarization. *Recent Advances in Natural Language Processing (RANLP)*. 2007.

### Workshop Papers

- [33] Amir Hazem, Mérième Bouhandi, Florian Boudin, and Béatrice Daille. TermEval 2020: TALN-LS2N System for Automatic Term Extraction. *6th International Workshop on Computational Terminology (CompuTerm)*. 2020.
- [34] Mérième Bouhandi, Florian Boudin, and Ygor Gallina. DeFT 2019: Auto-encodeurs, Gradient Boosting et combinaisons de modèles pour l’identification automatique de mots-clés. *Défi Fouille de Textes (DEFT)*. 2019.
- [35] Florian Boudin, Hugo Mougard, and Damien Cram. How Document Pre-processing affects Keyphrase Extraction Performance. *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. 2016.
- [36] Adrien Bougouin, Florian Boudin, and Béatrice Daille. TopicRank en domaines de spécialité : participation du LINA à DEFT 2016. *Défi Fouille de Textes (DEFT)*. 2016.
- [37] Béatrice Daille, Sabine Barreaux, Florian Boudin, Adrien Bougouin, Damien Cram, and Amir Hazem. Indexation d’articles scientifiques : Présentation et résultats du défi fouille de textes DEFT 2016. *Défi Fouille de Textes (DEFT)*. 2016.
- [38] Florian Boudin. Reducing Over-generation Errors for Automatic Keyphrase Extraction using Integer Linear Programming. *Workshop on Novel Computational Approaches to Keyphrase Extraction*. 2015.

- [39] Emmanuel Morin, Amir Hazem, Florian Boudin, and Elizaveta Loginova-Clouet. LINA: Identifying Comparable Documents from Wikipedia. *Eighth Workshop on Building and Using Comparable Corpora*. 2015.
- [40] Romain Deveaud and Florian Boudin. Effective Tweet Contextualization with Hashtags Performance Prediction and Multi-Document Summarization. *INitiative for the Evaluation of XML Retrieval (INEX)*. 2013.
- [41] Florian Boudin, Amir Hazem, Nicolas Hernandez, and Prajol Shrestha. Participation du LINA à DEFT2012 (LINA at DEFT2012). *Défi Fouille de Textes (DEFT)*. 2012.
- [42] Romain Deveaud and Florian Boudin. LIA/LINA at the INEX 2012 Tweet Contextualization track. *INitiative for the Evaluation of XML Retrieval (INEX)*. 2012.
- [43] Romain Deveaud, Florian Boudin, and Patrice Bellot. LIA at INEX 2010 Book Track. *INitiative for the Evaluation of XML Retrieval (INEX)*. 2011.
- [44] Florian Boudin, Jian-Yun Nie, and Martin Dawes. Deriving a test collection for clinical information retrieval from systematic reviews. *Data and Text Mining in Biomedical Informatics (DTMBIO)*. 2010.
- [45] Florian Boudin, Marc El-Bèze, and Juan-Manuel Torres-Moreno. The LIA Update Summarization system at TAC-2008. *Text Analysis Conference (TAC)*. 2008.
- [46] Florian Boudin, Benoit Favre, Frederic Béchet, Marc El-Bèze, Laurent Gillard, and Juan-Manuel Torres-Moreno. The LIA-Thales summarization system at DUC-2007. *Document Understanding Conference (DUC)*. 2007.
- [47] B. Favre, F. Béchet, P. Bellot, F. Boudin, M. El-Beze, L. Gillard, G. Lapalme, and J.M. Torres-Moreno. The LIA-Thales summarization system at DUC-2006. *Document Understanding Conference (DUC)*. 2006.

#### National Conference Papers

- [48] Adrien Bougouin, Florian Boudin, and Béatrice Daille. Modélisation unifiée du document et de son domaine pour une indexation par termes-clés libre et contrôlée. *Traitement Automatique des Langues Naturelles (TALN)*. 2016.
- [49] Adrien Bougouin, Florian Boudin, and Béatrice Daille. Influence des domaines de spécialité dans l'extraction de termes-clés. *Traitement Automatique des Langues Naturelles (TALN)*. 2014.
- [50] Florian Boudin. TALN Archives : une archive numérique francophone des articles de recherche en Traitement Automatique de la Langue. *Traitement Automatique des Langues Naturelles (TALN)*. 2013.
- [51] Romain Deveaud and Florian Boudin. Contextualisation automatique de Tweets à partir de Wikipédia. *Conférence en Recherche d'Information et Applications (CORIA)*. 2013.
- [52] Nicolas Hernandez and Florian Boudin. Construction d'un large corpus écrit libre annoté morpho-syntaxiquement en français. *Traitement Automatique des Langues Naturelles (TALN)*. 2013.
- [53] Florian Boudin and Nicolas Hernandez. Détection et correction automatique d'erreurs d'annotation morpho-syntaxique du French TreeBank. *Traitement Automatique des Langues Naturelles (TALN)*. 2012.
- [54] Romain Deveaud, Florian Boudin, Eric SanJuan, and Patrice Bellot. Correction de césures et enrichissement de requêtes pour la recherche de livres. *Conférence en Recherche d'Information et Applications (CORIA)*. 2011.
- [55] Stéphane Huet, Florian Boudin, and Juan-Manuel Torres-Moreno. Utilisation d'un score de qualité de traduction pour le résumé multi-document cross-lingue. *Traitement Automatique des Langues Naturelles (TALN)*. 2011.
- [56] Florian Boudin and Juan-Manuel Torres-Moreno. Résumé automatique multi-document et indépendance de la langue : une première évaluation en français. *Traitement Automatique des Langues Naturelles (TALN)*. 2009.

## Book Chapters

- [57] Florian Boudin and Juan-Manuel Torres-Moreno. A Maximization-Minimization Approach for Update Text Summarization. *Current Issues in Linguistic Theory: Recent Advances in Natural Language Processing*. 2009.

## Ph.D. Thesis

- [58] Florian Boudin. Exploration d’approches statistiques pour le résumé automatique de texte. PhD thesis. Laboratoire Informatique d’Avignon – Université d’Avignon, 2008.

## FUNDING

<b>DELICES - Indexer la littérature scientifique par expansion sémantique</b>	2020
ANR - JCJC, 200k€ over four years. Sole PI.	
<b>IKEBANA - Improving Keyphrase Extraction By Adopting Neural Architectures</b>	2019
Atlantist 2020 Sabbatical, research stay at the National Institute of Informatics, 20k€.	
<b>TALIAS - LE TAL au service de l’Indexation des Articles Scientifiques</b>	2016
CNRS - PEPS INS2I, 5k€ over one year. Sole PI.	
<b>GOLEM - Approche par Optimisation pour l’Extraction de Mots-clés</b>	2015
CNRS - PEPS INS2I/INSMI, 6k€ over one year. Sole PI.	

## SERVICE

### Area Chairing

Action Editor for the ACL Rolling Review (oct. 2021, nov. 2021)  
ACL-IJCNLP 2021, Summarization track.

### Programme Committee & Paper Reviewing

Reviewer for the ACL Rolling Review (may 2021, sept. 2021)

#### Conferences

ACL (2017, 2019, 2020, 2021)	
NAACL-HLT (2016, 2018, 2019, 2021)	best reviewer in 2016
EMNLP (2019, 2020, 2021)	outstanding reviewer in 2020
EACL (2021)	
LREC (2012, 2018, 2020)	
COLING (2016, 2018, 2020)	
AACL-IJCNLP (2020)	
IJCNLP (2017)	
CIKM (2011)	

#### Workshops

NewSum at (EMNLP 2019, EMNLP 2021)  
Text2Story at (ECIR 2019, ECIR 2021)  
NUSE at ACL 2020  
AI4Narratives at IJCAI-PRICAI 2020  
FNP-FNS 2020  
MultiLing 2019 at RANLP 2019

### Journal Reviewing

Traitement Automatique des Langues, Information Retrieval, Natural Language Engineering, Information Processing & Management, Language Resources and Evaluation

## TEACHING

<b>Université de Nantes, France</b>	2000h+
Programming, Databases, Machine Learning	
<b>Université d’Avignon et des Pays de Vaucluse, France</b>	96h
Programming, Business Intelligence	
<b>Université de Montréal, Canada</b>	54h
Atelier en Informatique	

