Florian Boudin   Curriculum Vitæ

Last update: June 18, 2025

Nantes Université
2 rue de la Houssinière
44322 Nantes, France

boudinfl.github.io/
florian.boudin@univ-nantes.fr
+33 2 51 12 59 98

Employment History

09/2011 – **Associate Professor** - *Maître de Conférences*
present   Nantes Université, France

09/2010 – **Assistant Professor** - *Attaché Temporaire d'Enseignement et de Recherche*
08/2011   Université d'Avignon et des Pays de Vaucluse, France

02/2009 – **Postdoctoral Fellow**
08/2010   Université de Montréal, Canada

09/2006 – **Research Assistant**
01/2009   Université d'Avignon et des Pays de Vaucluse, France

Education

06/2023   **Habilitation à Diriger des Recherches**
Nantes Université, France

09/2006 – **Ph.D. in Computer Science**
12/2008   Université d'Avignon et des Pays de Vaucluse, France
Advisors: Juan-Manuel Torres-Moreno, Marc El-Bèze

2006   **M.Sc. in Computer Science**
Université d'Avignon et des Pays de Vaucluse, France

2004   **B.Sc. in Mathematics and Computer Science**
Université d'Avignon et des Pays de Vaucluse, France
University of Birmingham, UK (ERASMUS Programme)

Selected Awards, Grants and Fellowships

09/2023 – **CNRS research leave** – *Délégation CNRS*
08/2025   Japanese-French Laboratory for Informatics (JFLI), Tokyo

2020   **French Defense Innovation Agency and CNRS research grant** – NaviTerm project

2023   **Prime individuelle C3** – *Supervision and research activity bonus*

2020   **ANR JCJC research grant** – *Young researchers grant* – DELICES project

2020   **EMNLP 2020 Outstanding Reviewer**

07/2019 – **Atlanstic 2020 sabbatical grant for 6 months**
12/2019   National Institute of Informatics (NII), Tokyo

2017   **Prime d'Encadrement Doctoral et de Recherche** – *Supervision and research activity bonus*

2016   **NAACL-HLT 2016 Best Reviewer**

Research Interests

My research focuses on **Natural Language Processing (NLP)** and **Information Retrieval (IR)**, specializing in the analysis, understanding, and generation of natural language, as well as extracting information from document collections. I concentrate on applications involving **scholarly documents**, leveraging generative AI to assist researchers in critical tasks such as literature search, paper understanding and scientific writing.

Publications

In the fields of NLP and IR, conferences are more competitive, more highly regarded, and more highly cited than journals. According to Google Scholar, my five most cited papers are: (C22) (599 citations), (C16) (307 citations), (C23) (188 citations), (J2) (186 citations) and (C24) (105 citations). Top-tier NLP/IR conferences: ACL (2 papers), NAACL (4 papers), EMNLP (3 papers), COLING (5 papers), ECIR (2 papers) and JCDL (3 papers).

My research profiles: Google Scholar, Semantic Scholar, ACL Anthology, dblp

Journal Papers

(J1) Kenichi Iwatsuki, **Florian Boudin**, Akiko Aizawa. 2022. Extraction and evaluation of formulaic expressions used in scholarly papers. *Expert Systems with Applications.*

(J2) **Florian Boudin**, Jian-Yun Nie, Joan Bartlett, Roland Grad, Pierre Pluye, Martin Dawes. 2010. Combining classifiers for robust PICO element detection. *BMC Medical Informatics and Decision Making.*

Conference Papers

(C1) Leane Jourdan, Nicolas Hernandez, **Florian Boudin**, Richard Dufour. Identifying Reliable Evaluation Metrics for Scientific Text Revision. *Association for Computational Linguistics (ACL).*

(C2) Kon Woo Kim, Rezarta Islamaj, Jin-Dong Kim, **Florian Boudin**, Akiko Aizawa. Repurposing Annotation Guidelines to Instruct LLM Annotators: A Case Study. *International Conference on Natural Language & Information Systems (NLDB).*

(C3) Julien Aubert-Béduchaud, **Florian Boudin**, Béatrice Daille, Richard Dufour. 2025. ACL-rlg : A Dataset for Reading List Generation. *International Conference on Computational Linguistics (COLING).*

(C4) Maël Houbre, **Florian Boudin**, Béatrice Daille, Akiko Aizawa. 2024. Self-Compositional Data Augmentation for Scientific Keyphrase Generation. *Joint Conference on Digital Libraries (JCDL).*

(C5) **Florian Boudin**, Akiko Aizawa. 2024. Unsupervised Domain Adaptation for Keyphrase Generation using Citation Contexts. *In Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*

(C6) Léane Jourdan, **Florian Boudin**, Nicolas Hernandez, Richard Dufour. 2024. CASIMIR: A Corpus of Scientific Articles enhanced with Multiple Author-Integrated Revisions. *Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING).*

(C7) Amir Hazem, Mérieme Bouhandi, **Florian Boudin**, Béatrice Daille. 2022. Cross-lingual and Cross-domain Transfer Learning for Automatic Term Extraction from Low Resource Data. *Language Resources and Evaluation Conference (LREC).*

(C8) Rui Meng, Debanjan Mahata, **Florian Boudin**. 2022. From Fundamentals to Recent Advances: A Tutorial on Keyphrasification. *European Conference on Information Retrieval (ECIR).*

(C9) **Florian Boudin**, Ygor Gallina. 2021. Redefining Absent Keyphrases and their Effect on Retrieval Effectiveness. *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).*

(C10) **Florian Boudin**. 2021. ACM-CR: A Manually Annotated Test Collection for Citation Recommendation. *Joint Conference on Digital Libraries (JCDL).*

(C11) **Florian Boudin**, Ygor Gallina, Akiko Aizawa. 2020. Keyphrase Generation for Scientific Document Retrieval. *Association for Computational Linguistics (ACL).*

(C12) Ygor Gallina, **Florian Boudin**, Béatrice Daille. 2020. Large-Scale Evaluation of Keyphrase Extraction Models. *Joint Conference on Digital Libraries (JCDL).*

(C13) **Florian Boudin**, Béatrice Daille, Evelyne Jacquey, Jian-Yun Nie. 2020. The DELICES Project: Indexing Scientific Literature Through Semantic Expansion. *Joint Conference of the Information Retrieval Communities in Europe (CIRCLE).*

(C14) Kenichi Iwatsuki, **Florian Boudin**, Akiko Aizawa. 2020. An Evaluation Dataset for Identifying Communicative Functions of Sentences in English Scholarly Papers. *Language Resources and Evaluation Conference (LREC).*

(C15) Ygor Gallina, **Florian Boudin**, Béatrice Daille. 2019. KPTimes: A Large-Scale Dataset for Keyphrase Generation on News Documents. *International Conference on Natural Language Generation (INLG).*

(C16) **Florian Boudin**. 2018. Unsupervised Keyphrase Extraction with Multipartite Graphs. *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).*

(C17) **Florian Boudin**. 2016. pke: an open source python-based keyphrase extraction toolkit. *International Conference on Computational Linguistics (COLING).*

(C18) Adrien Bougouin, **Florian Boudin**, Béatrice Daille. 2016. Keyphrase Annotation with Graph Co-Ranking. *International Conference on Computational Linguistics (COLING).*

(C19) Adrien Bougouin, Sabine Barreaux, Laurent Romary, **Florian Boudin**, Béatrice Daille. 2016. TermITH-Eval: a French Standard-Based Resource for Keyphrase Extraction Evaluation. *Language Resources and Evaluation Conference (LREC).*

(C20) **Florian Boudin**, Hugo Mougard, Benoit Favre. 2015. Concept-based Summarization using Integer Linear Programming: From Concept Pruning to Multiple Optimal Solutions. *Conference on Empirical Methods in Natural Language Processing (EMNLP).*

(C21) Ophélie Lacroix, Denis Béchet, **Florian Boudin**. 2014. Label Pre-annotation for Building Non-projective Dependency Treebanks for French. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing).*

(C22) Adrien Bougouin, **Florian Boudin**, Béatrice Daille. 2013. TopicRank: Graph-Based Topic Ranking for Keyphrase Extraction. *International Joint Conference on Natural Language Processing (IJCNLP).*

(C23) **Florian Boudin**. 2013. A Comparison of Centrality Measures for Graph-Based Keyphrase Extraction. *International Joint Conference on Natural Language Processing (IJCNLP).*

(C24) **Florian Boudin**, Emmanuel Morin. 2013. Keyphrase Extraction for N-best Reranking in Multi-Sentence Compression. *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).*

(C25) **Florian Boudin**, Jian-Yun Nie, Martin Dawes. 2012. Using a Medical Thesaurus to Predict Query Difficulty. *European Conference on Information Retrieval (ECIR).*

(C26) **Florian Boudin**, Stéphane Huet, Juan-Manuel Torres-Moreno. 2011. A Graph-based Approach to Cross-language Multi-document Summarization. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing).*

(C27) **Florian Boudin**, Jian-Yun Nie, Martin Dawes. 2010. Positional Language Models for Clinical Information Retrieval. *Conference on Empirical Methods in Natural Language Processing (EMNLP).*

(C28) **Florian Boudin**, Jian-Yun Nie, Martin Dawes. 2010. Clinical Information Retrieval using Document and PICO Structure. *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).*

(C29) **Florian Boudin**, Lixin Shi, Jian-Yun Nie. 2010. Improving Medical Information Retrieval with PICO Element Detection. *European Conference on Information Retrieval (ECIR).*

(C30) **Florian Boudin**, Marc El-Bèze, Juan-Manuel Torres-Moreno. 2008. A Scalable MMR Approach to Sentence Scoring for Multi-Document Update Summarization. *International Conference on Computational Linguistics (COLING).*

(C31) **Florian Boudin**, Juan Torres-Moreno, Marc El-Bèze. 2008. Mixing Statistical and Symbolic Approaches for Chemical Names Recognition. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing).*

(C32) **Florian Boudin**, Juan-Manuel Torres-Moreno, Patricia Velázquez-Morales. 2008. An Efficient Statistical Approach for Automatic Organic Chemistry Summarization. *International Conference on Natural Language Processing (GoTAL).*

(C33) **Florian Boudin**, Juan-Manuel Torres-Moreno. 2007. A Cosine Maximization Minimization approach for User Oriented Multi-Document Update Summarization. *Recent Advances in Natural Language Processing (RANLP).*

(C34) **Florian Boudin**, Juan Torres Moreno. 2007. NEO-CORTEX: A Performant User-Oriented Multi-Document Summarization System. *Conference on Intelligent Text Processing and Computational Linguistics (CICLing).*

Workshop Papers

(W1) **Florian Boudin**, Akiko Aizawa. An Analysis of Datasets, Metrics and Models in Keyphrase Generation. *Workshop on Natural Language Generation, Evaluation, and Metrics (GEM²).*

(W2) An Dao, Hiroki Teranishi, Yuji Matsumoto, **Florian Boudin**, Akiko Aizawa. Overcoming Data Scarcity in Named Entity Recognition: Synthetic Data Generation with Large Language Models. *Biomedical Natural Language Processing Workshop (BioNLP).*

(W3) Léane Jourdan, Nicolas Hernandez, Richard Dufour, **Florian Boudin**, Akiko Aizawa. 2025. ParaRev: Building a dataset for Scientific Paragraph Revision annotated with revision instruction. *International Workshop on Writing Aids at the Crossroads of AI, Cognitive Science and NLP (WR-AI-CogS).*

(W4) Enrico Benedetti, Akiko Aizawa, **Florian Boudin**. 2024. Automatically Suggesting Diverse Example Sentences for L2 Japanese Learners Using Pre-Trained Language Models. *Association for Computational Linguistics (ACL): Student Research Workshop.*

(W5) Léane Jourdan, **Florian Boudin**, Nicolas Hernandez, Richard Dufour. 2023. Text revision in Scientific Writing Assistance: A Review. *International Workshop on Bibliometric-enhanced Information Retrieval (BIR).*

(W6) Maël Houbre, **Florian Boudin**, Béatrice Daille. 2022. A large-scale dataset for biomedical keyphrase generation. *International Workshop on Health Text Mining and Information Analysis (LOUHI).*

(W7) Amir Hazem, Mérieme Bouhandi, **Florian Boudin**, Beatrice Daille. 2020. TermEval 2020: TALN-LS2N System for Automatic Term Extraction. *6th International Workshop on Computational Terminology (CompuTerm).*

(W8) **Florian Boudin**, Hugo Mougard and Damien Cram. 2016. How Document Pre-processing affects Keyphrase Extraction Performance. *Workshop on Noisy User-generated Text (WNUT).*

(W9) **Florian Boudin**. 2015. Reducing Over-generation Errors for Automatic Keyphrase Extraction using Integer Linear Programming. *Workshop on Novel Computational Approaches to Keyphrase Extraction.*

(W10) Emmanuel Morin, Amir Hazem, **Florian Boudin**, Elizaveta Loginova-Clouet. 2015. LINA: Identifying Comparable Documents from Wikipedia. *Eighth Workshop on Building and Using Comparable Corpora (BUCC).*

(W11) Romain Deveaud, **Florian Boudin**. 2013. Effective Tweet Contextualization with Hashtags Performance Prediction and Multi-Document Summarization. *INitiative for the Evaluation of XML Retrieval (INEX).*

(W12) Romain Deveaud, **Florian Boudin**. 2012. LIA/LINA at the INEX 2012 Tweet Contextualization track. *INitiative for the Evaluation of XML Retrieval (INEX).*

(W13) Romain Deveaud, **Florian Boudin**, Patrice Bellot. 2011. LIA at INEX 2010 Book Track *INitiative for the Evaluation of XML Retrieval (INEX)*.

(W14) **Florian Boudin**, Jian-Yun Nie, Martin Dawes. 2010. Deriving a test collection for clinical information retrieval from systematic reviews. *Data and Text Mining in Biomedical Informatics (DTMBIO)*.

(W15) **Florian Boudin**, Marc El-Bèze, Juan-Manuel Torres-Moreno. 2008. The LIA Update Summarization system at TAC-2008. *Text Analysis Conference (TAC)*.

(W16) **Florian Boudin**, Benoit Favre, Frederic Béchet, Marc El-Bèze, Laurent Gillard, Juan-Manuel Torres-Moreno. 2007. The LIA-Thales summarization system at DUC-2007. *Document Understanding Conference (DUC)*.

(W17) Benoit Favre, Frederic Béchet, Patrice Bellot, **Florian Boudin**, Marc El-Beze, Laurent Gillard, Guy Lapalme, Juan-Manuel Torres-Moreno. 2006. The LIA-Thales summarization system at DUC-2006. *Document Understanding Conference (DUC)*.

National Journal Papers

(R1) **Florian Boudin**, Akiko Aizawa. 2024. Preface to the Special Issue of the TAL Journal on Scholarly Document Processing. *Traitement Automatique des Langues*.

(R2) Adrien Bougouin, **Florian Boudin**, Béatrice Daille. 2017. Modélisation à base de graphe pour l'indexation en domaines de spécialité. *Recherche d'information, document et web sémantique*.

(R3) Adrien Bougouin, **Florian Boudin**, Béatrice Daille, Sabine Barreaux, Damien Cram, Amir Hazem. 2017. Indexation d'articles scientifiques Présentation et résultats du défi fouille de textes DEFT 2016. *Recherche d'information, document et web sémantique*.

(R4) Adrien Bougouin, **Florian Boudin**. 2014. TopicRank: ordonnancement de sujets pour l'extraction automatique de termes-clés. *Traitement Automatique des Langues*.

(R5) Romain Deveaud, **Florian Boudin**. 2014. De quoi parle ce Tweet? Résumer Wikipédia pour contextualiser des microblogs. *The Information-Intelligence-Interaction (I3) Journal*.

National Conference Papers

(N1) Léane Jourdan, **Florian Boudin**, Richard Dufour, Nicolas Hernandez. Identification de mesures d'évaluation fiables pour la révision de textes scientifiques. *Traitement Automatique des Langues Naturelles (TALN)*.

(N2) Julien Aubert-Béduchaud, **Florian Boudin**, Béatrice Daille, Richard Dufour. ACL-rlg: Un dataset pour la génération de listes de lecture. *Traitement Automatique des Langues Naturelles (TALN)*.

(N3) Adrien Bougouin, **Florian Boudin**, Béatrice Daille. 2016. Modélisation unifiée du document et de son domaine pour une indexation par termes-clés libre et contrôlée. *Traitement Automatique des Langues Naturelles (TALN)*.

(N4) Adrien Bougouin, **Florian Boudin**, Béatrice Daille. 2014. Influence des domaines de spécialité dans l'extraction de termes-clés. *Traitement Automatique des Langues Naturelles (TALN)*.

(N5) **Florian Boudin**. 2013. TALN Archives: une archive numérique francophone des articles de recherche en Traitement Automatique de la Langue. *Traitement Automatique des Langues Naturelles (TALN)*.

(N6) Nicolas Hernandez, **Florian Boudin**. 2013. Construction d'un large corpus écrit libre annoté morpho-syntaxiquement en français. *Traitement Automatique des Langues Naturelles (TALN)*.

(N7) Romain Deveaud, **Florian Boudin**. 2013. Contextualisation automatique de Tweets à partir de Wikipédia. *Conférence en Recherche d'Information et Applications (CORIA)*.

(N8) **Florian Boudin**, Nicolas Hernandez. 2012. Détection et correction automatique d'erreurs d'annotation morpho-syntaxique du French TreeBank. *Traitement Automatique des Langues Naturelles (TALN)*.

(N9) Stéphane Huet, **Florian Boudin**, Juan-Manuel Torres-Moreno. 2011. Utilisation d'un score de qualité de traduction pour le résumé multi-document cross-lingue. *Traitement Automatique des Langues Naturelles (TALN)*.

(N10) Romain Deveaud, **Florian Boudin**, Eric SanJuan, Patrice Bellot. 2011. Correction de césures et enrichissement de requêtes pour la recherche de livres. *Conférence en Recherche d'Information et Applications (CORIA)*.

(N11) **Florian Boudin**, Juan-Manuel Torres-Moreno. 2009. Résumé automatique multi-document et indépendance de la langue: une première évaluation en français. *Traitement Automatique des Langues Naturelles (TALN)*.

National Workshop Papers

(A1) Léane Jourdan, **Florian Boudin**, Nicolas Hernandez, Richard Dufour. 2023. CASIMIR: un Corpus d'Articles Scientifiques Intégrant les ModIfications et Révisions des auteurs. *Atelier sur l'Analyse et Recherche de Textes Scientifiques (ARTS)*.

(A2) Maël Houbre, **Florian Boudin**, Béatrice Daille. 2023. Classification de relation pour la génération de mots-clés absents. *Atelier sur l'Analyse et Recherche de Textes Scientifiques (ARTS)*.

(A3) **Florian Boudin**, Richard Dufour, Béatrice Daille. 2023. Projet NaviTerm: navigation terminologique pour une montée en compétence rapide et personnalisée sur un domaine de recherche. *Atelier sur l'Analyse et Recherche de Textes Scientifiques (ARTS)*.

(A4) Mérième Bouhandi, **Florian Boudin**, Ygor Gallina. 2019. DeFT 2019: Auto-encodeurs, Gradient Boosting et combinaisons de modèles pour l'identification automatique de mots-clés. *Défi Fouille de Textes (DEFT)*.

(A5) Béatrice Daille, Sabine Barreaux, **Florian Boudin**, Adrien Bougouin, Damien Cram, Amir Hazem. 2016. Indexation d'articles scientifiques: Présentation et résultats du défi fouille de textes DEFT 2016. *Défi Fouille de Textes (DEFT)*.

(A6) Adrien Bougouin, **Florian Boudin**, Béatrice Daille. 2016. TopicRank en domaines de spécialité: participation du LINA à DEFT 2016. *Défi Fouille de Textes (DEFT)*.

(A7) **Florian Boudin**, Amir Hazem, Nicolas Hernandez, Prajol Shrestha. 2012. Participation du LINA à DEFT 2012. *Défi Fouille de Textes (DEFT)*.

Book Chapters

(B1) **Florian Boudin**, Juan-Manuel Torres-Moreno. 2009. A Maximization-Minimization Approach for Update Text Summarization. *Current Issues in Linguistic Theory: Recent Advances in Natural Language Processing*.

Pre-print Papers

(P1) Xanh Ho, Anh Khoa Duong Nguyen, An Tuan Dao, Junfeng Jiang, Yuki Chida, Kaito Sugimoto, Huy Quoc To, **Florian Boudin**, Akiko Aizawa. 2024. A Survey of Pre-trained Language Models for Processing Scientific Text.

Others

(T1) **Florian Boudin**. 2023. Analyse et indexation de textes scientifiques. *Habilitation à Diriger les Recherches (HDR) – Nantes Université*.

(T2) **Florian Boudin**. 2008. Exploration d'approches statistiques pour le résumé automatique de texte. *Thèse de Doctorat – Université d'Avignon*.

Mentoring and Student Supervision

Ph.D. Students

10/2023 – **Julien Aubert-Béduchaud** (co-supervised with Béatrice Daille and Richard Dufour)
present Working on scientific reading list generation

10/2022 – **Léane Jourdan** (co-supervised with Nicolas Hernandez and Richard Dufour)
present Working on scientific writing assistance using LLMs

03/2022 – **Maël Houbre** (co-supervised with Béatrice Daille)
present Working on generative models for keyphrase generation

10/2018 – **Ygor Gallina** (co-supervised with Béatrice Daille)
03/2022 Worked on neural models for keyphrase generation – thesis 🗎

10/2012 – **Adrien Bougouin** (co-supervised with Béatrice Daille)
10/2015 Worked on graph-based keyphrase extraction – thesis 🗎

M.Sc. Students

2022 **Rima Boubekeur** (co-supervised with Richard Dufour)
Worked on automatic hashtag generation for short text messages

2020 **Timothée Poulain**
Worked on scientific document indexing

2019 **Carol Couillerot**
Worked on phishing email detection

2014 **Rémi Bois**
Worked on multi-document summarization through sentence fusion

SERVICE

Area Chairing

I am Area Chair for the ACL Rolling Review (since 2021).

Annual Meeting of the Association for Computational Linguistics (ACL 2021, 2022, 2024), Conference on Empirical Methods in Natural Language Processing (EMNLP 2022, 2023, 2024), North American Chapter of the Association for Computational Linguistics (NAACL 2022, 2024), International Conference on Computational Linguistics (COLING 2024).

Programme Committee & Paper Reviewing

Conferences

International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2023, 2024, 2025), Conference on Information Retrieval in the Asia Pacific (SIGIR-AP 2024), Conference on Language Modeling (COLM 2024), International Conference on Theory and Practice of Digital Libraries (TPDL 2024), Annual Meeting of the Association for Computational Linguistics (ACL 2017, 2019, 2020, 2021, 2022, 2023), European Conference on Information Retrieval (ECIR 2023), International Joint Conference on Natural Language Processing and Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL 2020, 2022, 2023), International Conference on Computational Linguistics (COLING 2012, 2016, 2018, 2020, 2022), (LREC 2012, 2018, 2020, 2022), Conference on Empirical Methods in Natural Language Processing (EMNLP 2019, 2020, 2021), North American Chapter of the Association for Computational Linguistics (NAACL 2016, 2018, 2019, 2021), Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021), International Joint Conference on Natural Language Processing (IJCNLP 2017), Conference on Information and Knowledge Management (CIKM 2011).

Workshops

Workshop on Future of Event Detection (FuturED 2024), Evaluating Text Difficulty in a Multilingual Context (DeTermIt! 2024), Workshop on New Frontiers in Summarization (NewSum 2019, 2021, 2023), Workshop on

Automatic Summarization for Creative Writing (Creative-sum 2022), Workshop on Digital Infrastructures for Scholarly Content Objects (DISCO 2021), International Workshop on Narrative Extraction from Texts (Text2Story 2019, 2021, 2022), Joint Workshop on Narrative Understanding, Storylines, and Events (NUSE 2020), Artificial Intelligence for Narratives Workshop (AI4Narratives 2020), Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation (FNP-FNS 2020), Workshop on Summarization Across Languages, Genres and Sources (MultiLing 2019).

Journals

Traitement Automatique des Langues, Information Retrieval, Natural Language Engineering, Information Processing & Management, Language Resources and Evaluation, Artificial Intelligence, BMC Bioinformatics.

National conferences

Conférence Traitement Automatique des Langues Naturelles (TALN 2010, 2011, 2013, 2014, 2015, 2016, 2017), COnférence en Recherche d'Information et Applications (CORIA) (2017, 2024), Rencontre des Étudiants Chercheurs en Informatique pour le TAL (RECITAL 2010, 2011, 2013, 2014, 2015, 2016, 2017).

Organizing Commitee

04/2022, **Member of the organizing commitee** of two one-day symposiums organized by the French
12/2024 Association for Artificial Intelligence (AFIA), through its College of Human Language Technologies (TLH), and the Association for Information Retrieval and Applications (ARIA). These events focused on the theme of interactive access to information (JAII 2021, 2024)

06/2023 **Head of the organizing committee and program chair** of the workshop on the Analysis and Retrieval of Scientific Texts (ARTS) held at the joint CORIA-TALN 2023 conference in Paris, France

04/2022 **Organizer and instructor** for the tutorial "From Fundamentals to Recent Advances: A Tutorial on Keyphrasification" given at the 44th European Conference on Information Retrieval (ECIR) in Stavanger, Norway.
Co-instructors: Rui Meng, Debanjan Mahata

07/2016 **Member of the organizing commitee** of the Défi fouille de textes (DEFT) shared task on scientific paper indexing held at the joint JEP-TALN-RECITAL 2016 in Paris, France

06/2013 **Head of the organizing committee and program chair** of the Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL) (*Meeting of Young Researchers in NLP*), held at the TALN 2013 conference in Les Sables d'Olonne, France

FUNDING

2023-2027 NaviTerm project – *Terminological navigation through the scientific literature*
Funded by the French Defense Innovation Agency and the CNRS, 213k€, sole PI
This project aims to accelerate how researchers familiarize themselves with a new field by automatically generating navigable representations of scientific knowledge
Collaborators: Béatrice Daille (Nantes Université), Richard Dufour (Nantes Université)

2020-2025 DELICES project – *Indexing scientific literature through semantic expansion*
Young Researcher project (JCJC) funded by the ANR, 200k€, sole PI
This project aims to improve the retrieval effectiveness of search engines in scientific digital libraries by enriching their indexing with automatically generated keyphrases
Collaborators: Béatrice Daille (Nantes Université), Evelyne Jacquey (ATILF), Jian-Yun Nie (Université de Montréal, Canada)

2020 WASP project – *Building a writing assistance system for scientific papers*
Atlanstic2020 grant for a visiting researcher from the University of Tokyo (canceled by COVID-19)
This project aims to design a writing assistance tool to help researchers select the most appropriate formulaic expressions for their articles
Collaborator: Akiko Aizawa (National Institute of Informatics, Japan)

2019 IKEBANA project – *Improving keyphrase extraction by adopting neural architectures*
Atlanstic2020 sabbatical grant, research stay at the National Institute of Informatics, 20k€
This project aims to improve the effectiveness of keyphrase extraction methods by adopting recurrent neural network models
Collaborator: Akiko Aizawa (National Institute of Informatics, Japan)

2016 TALIAS project – *Natural Language Processing for indexing scientific articles*
PEPS INS2I project funded by the CNRS, 5k€
This project aims to evaluate the robustness of keyphrase extraction models across varying levels of noise in documents

2015 GOLEM project – *An Optimization-Based Approach for Keyphrase Extraction*
PEPS INS2I/INSMI project funded by the CNRS, 6k€
Collaborator: Evgeny Gurevsky (Nantes Université)
This project aims to explore the use of Integer Linear Programming for automatic keyphrase extraction

INVITED TALKS

01/2024 Japanese-French Laboratory of Informatics (JFLI) seminar in Tokyo, Japan

10/2023 Aizawa Laboratory seminar at the National Institute of Informatics (NII), Tokyo, Japan

03/2023 Mots/Machines #5 seminar on the theme of terminology held at the Faculty of Arts and Humanities of the Université de Bretagne Occidentale in Brest, France

09/2021 "Defense Artificial Intelligence" seminar organized by the Defense Innovation Agency (AID) in Val-de-Grâce, France

07/2019 Aizawa Laboratory seminar at the National Institute of Informatics (NII), Tokyo, Japan