



Ahmed Hamdi

Teacher researcher
La Rochelle University, L3i Laboratory
Office 013
ahmed.hamdi@univ-lr.fr
[webpage](#) | [linkedin](#)

EDUCATION

AIX-MARSEILLE UNIVERSITY PhD in Computational Linguistics (obtained with honors)	2011-2015
AIX-MARSEILLE UNIVERSITY Master's Degree in Computational Linguistics (Mention bien)	2010-2011
NATIONAL SCHOOL OF COMPUTER SCIENCE (TUNISIA) Engineering's Degree in Computer Science	2007-2010
PREPARATORY INSTITUTE FOR ENGINEERING STUDIES (TUNISIA) Bachelor's Degree in Mathematics - Physics	2005-2007

RESEARCH ACTIVITIES

TEACHER AND RESEARCHER 2022-today / La Rochelle University

- Co-organisation of [DocILE'23](#) competition on Document Information Localization and Extraction at [ICDAR 2023](#) and [CLEF 2023](#)
- Invited speaker at the Digital Humanities Research Seminar in [RGCL](#), University of Wolverhampton, on May 19, 2022: [Content Analysis of Digital Text with Special Focus on Named Entity Recognition and Linking](#) [SLIDES]
- Reviewer: journals (LRE, JOCCH, CCOS, JCCE), conferences ([CHR](#), [DAS](#))

POSTDOCTORAL RESEARCHER 2017-2021 / La Rochelle University

- **Research topics**
 - Named Entity Recognition and Linking on Historical Newspapers.
 - Stance Detection towards Named entities.
 - Building word embedding models for less-resourced languages.
 - Information Extraction from Business Documents.
 - Document Flow Separation.
- **Supervision**
 - Vinh-Nam Huynh in L3 computer science from the University of Hanoi. It is led to a publication at the ICADL'20 conference.
 - Joris Voerman in M2 computer science at the University of La Rochelle. It is led to a publication at the WML@ICDAR'17 workshop.

PHD STUDENT 2011-2015 / Aix-Marseille University

- **Supervision:** Alexis NASR and Núria GALA
- **Topic:** Tunisian dialect processing using NLP tools and resources of Standard Arabic.
- **Defense date:** December 04, 2015
- **Research internship** at Columbia University (New York - USA), Center for Computational Learning Systems (CCLS), under the supervision of Nizar Habash, Owen Rambow. The topic is "morphological analysis of Arabic dialects". May-Sept 2014
- **Supervision** of the student Haidi Serraye in M1 computer science at Aix-Marseille University. The topic is building language models using SRILM.

INTERNSHIP Feb 2011-Jun 2011 / Aix-Marseille University

- **Supervision:** Alexis NASR and Núria GALA
- **Topic:** Contribution of diacritisation on morphological analysis of Arabic texts.

TEACHING ACTIVITIES

TEACHER AND RESEARCHER

2022-today / La Rochelle University

- Teaching hours: 296h
- Teaching units: Big Data, Semantic web (text mining), Process mining, Networking and security, Service-Oriented Architecture, Modeling, Web development.
- **Responsibilities**
 - Responsible for the Master INFO (syllabus *Architecte Logiciel*) at NiortTech / La Rochelle University
 - Development of relationships between companies of Niort and La Rochelle University.
 - Training responsibility in the [AI4industry workshop](#) at La Rochelle university
 - Responsible for courses *Mise en oeuvre des systèmes big data* (M2, info), 5V-model, Hadoop, Spark, NoSQL databases, SQALA, KAFKA... [[CM](#), [TP](#)]
 - Responsible for courses *Technologies du web* (L2, Gestion), HTML and CSS basis, Wordpress... [[CM](#), [TP](#)],
 - Teaching at the European university (Eu-Conexus), Computer science for e-tourism (in English, All licenses) web scrapping and information extraction [[CM](#), [TP](#)].
 - Pedagogical tutoring (internships: L3, M1 and M2)

ASSISTANT TEACHER (VACATAIRE)

2017-2021 / La Rochelle University

- Teaching hours: 56h
- Teaching units: Computer science for e-tourism, Web development.

ASSISTANT TEACHER AND RESEARCHER (ATER)

2014-2016 / Aix-Marseille university

- Teaching hours: 384h
- Teaching units: Algorithmic, Introduction to computer science and programming, C language, Computer tools and C2i, Python language, Web development, Combinatorial algebra.

ASSISTANT TEACHER (MONITEUR)

2012-2014 / Aix-Marseille university

- Teaching hours: 128h
- Teaching units: Introduction to computer science and programming, C language, Computer tools and C2i, Combinatorial algebra.

QUALIFICATION AND AWARDS

- Qualification French CNU 27 (computer science) in 2021
- CLEF-HIPE 2022: Named Entity Recognition (NER) and Entity Linking (EL) on Historical Newspapers,
 - NER: **1st place for our L3i-team in 20/26 leaderboards**
 - EL: **1st place for our L3i-team in 6/8 leaderboards**
<https://hipe-eval.github.io/HIPE-2022/results>
- CLEF-HIPE 2020: Named Entity Recognition (NER) and Entity Linking (EL) on Historical Newspapers NER
 - NER: **1st place for our L3i-team in all leaderboards.**
 - EL: **1st place for our L3i-team in 50/52 leaderboards and 2nd place in 2 leaderboards.**
https://github.com/impresso/CLEF-HIPE-2020/blob/master/evaluation-results/ranking_summary_final.md
- ICADL 2019: "Impact of OCR quality on named entity linking", **best paper**
- WANLP@ACL 2015: "POS-tagging of Tunisian dialect using standard Arabic resources and tools", **best paper**
- Hackathon "La Rochelle ville connectée" 2017, **Hackathon price**
<https://opendata.agglo-larochelle.fr/>

OTHER ACTIVITIES AND RESPONSIBILITIES

- Member of the organizing committee of the **INFORSID** (INFormatique des ORganisations and Systèmes d'Information et de Décision) congress 2023 at La Rochelle University.
- Member of the organizing committee of the international conference **CICLING** (*International Conference on Intelligent Text Processing and Computational Linguistics*) 2019 at La Rochelle university.
- Participation in organizing the local event at the L3i laboratory "Journée des Ingénieurs, Doctorants, ATER et Post-docs" (JIDAP) en 2018 et 2019.
- Participation to the national event "fêtes de la science" at the university of La Rochelle in 2018, 2019, 2020 and 2021. <https://www.univ-larochelle.fr/wp-content/uploads/pdf/FDLS-2020-PROGRAMME-PUBLIC-20200923.pdf>
- Participation to the national event "journées portes ouvertes" at the university of La Rochelle in 2019 and 2020. [SLIDES JPO 2020]

NLP RESOURCES

- Multilingual Dataset for Named Entity Recognition, Entity Linking and Stance Detection in Historical Newspapers (**NewsEye**).
- Benchmark for the evaluation of named entity recognition over ancient documents (**OCRed_NER**).
- Benchmark for the evaluation of entity linking over ancient documents (**OCRed_EL**).

SUMMARY OF TEACHING ACTIVITIES:

Year	Entitled	Public	Discipline	CM	TD	TP
2022-2023	Big Data	M2	INFO	30	-	66
	Process Mining	M2	INFO	10,5	10,5	13,5
	Semantic Web	M2	INFO	6	3	3
	Experiencing Professional Students - Project	M2	INFO	9	-	24
	Operating Systems	L1	INFO	-	-	27
	Web Technology	L2	Management	12	-	12
2021-2022	Big Data	M2	Computer Science	4	-	6
	Service Oriented Architecture	M1	Computer Science	3	9	-
	Networking and security	M1	Computer Science	18	12	-
	Computer Science for e-Tourism		All licences	6	6	-
	Information Technology	L2	Management	-	28	-
	Modeling	L1	Computer Science	-	7,5	15
2020-2021	Web Basics	L1	Computer Science	-	30	-
	Big Data	M1	Computer Science	4	-	6
	Computer Science for e-Tourism		All licences	6	6	-
2019-2020	Information Technology	L2	Management	-	28	-
	Information Technology	L1	Management	-	28	-
2018-2019	Information Technology	L1	Management	-	28	-
2017-2018	Information Technology	L1	Management	-	14	-
2016-2017	Information Technology	L1	Management	-	14	-
2015-2016	Algorithmic	L2	Mathematics	-	24	20
	Introduction to Computer Science and Programming	L1	Computer Science	20	24	21
	Computer tools and C2I	L1	Computer Science	-	24	-
	Python II	Licence 1	Computer Science	-	24	-
2014-2015	Combinatorial Algebra	L1	Computer Science	-	24	18
	Algorithmic	L2	Computer Science	-	24	20
	Computer tools and C2I	L1	Computer Science	-	48	-
	Introduction to Computer Science and Programming	L1	Computer Science	-	-	63
	C language	L1	Computer Science	-	18	21
2013-2014	Web development I	L1	Computer Science	-	15	-
	Introduction to computer science and programming	L1	Computer Science	-	-	42
2012-2013	Computer Tools and C2I	L1	Computer Science	-	24	-
	Introduction to Computer Science and Programming	L1	Computer Science	-	-	21
	Computer Tools and C2I	L1	Computer Science	-	24	-
	Combinatorial Algebra	L1	Computer Science	-	24	-

PUBLICATION STATS

30 publications in total – **h-index=11** (Google Scholar, 2023)

- 3 peer-reviewed international journals
- 13 peer-reviewed international conferences
- 1 book
- 3 papers in international workshops
- 2 working notes in international labs
- 3 peer-reviewed national conferences
- 5 non-peer-reviewed papers

FULL LIST OF PUBLICATIONS

1. Journal articles

- ACM Computing Surveys 2023 (Q1 – impact factor = 7.99).
Named Entity Recognition and Classification on Historical Documents: A Survey. Maud Ehrmann, **Ahmed Hamdi**, Elvys Linhares Pontes, Matteo Romanello and Antoine Doucet.
- NLE 2022 (Q1).
In-depth analysis of the impact of OCR errors on named entity recognition and linking. **Ahmed Hamdi**, Elvys Linhares Pontes, Nicolas Sidère, Mickaël Coustaty and Antoine Doucet. *Natural Language Engineering Journal*, pp 1 – 24.
- IJDL 2021 (Q2).
MELHISSA: A Multilingual Entity Linking Architecture for Historical Press Articles. Elvys Linhares Pontes, Luis Adrian Cabrera-Diego, Jose G. Moreno, Emanuela Boros, **Ahmed Hamdi**, Nicolas Sidère, Mickaël Coustaty and Antoine Doucet. *International Journal on Digital Libraries, IJDL*. pp 1 – 28.

2. International conferences

- ECIR 2023 (Core A).
Knowledge-based Contexts for Historical Named Entity Recognition Linking Documents. Emanuela Boros, Carlos-Emiliano González-Gallardo, Edward Giamphy, **Ahmed Hamdi**, Jose G. Moreno and Antoine Doucet. The 45th European Conference on Information Retrieval (ECIR 2023).
- ECIR 2023 (Core A).
DocILE 2023 Teaser: Document Information Localization and Extraction. Štěpán Šimsa, Milan Šulc, Matyas Skalický, Yash Patel and **Ahmed Hamdi**. The 45th European Conference on Information Retrieval (ECIR 2023).
- ICCCI 2022 (Core B)
BERT and Word Embedding for Interest Mining of Instagram Users. Sana Hamdi, **Ahmed Hamdi** et Sadok Ben Yahia. 14th International Conference on Computational Collective Intelligence. Springer, 2022, pp. 123-136.
- SIGIR 2021 (Core A*^{*}).
A Multilingual Dataset for Named Entity Recognition, Entity Linking and Stance Detection in Historical Newspapers. **Ahmed Hamdi**, Elvys Linhares Pontes, Emanuela Boros, Tuyet Hai Nguyen Thi, Hackl Günter, Jose G. Moreno and Antoine Doucet. The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021).

- ICDAR 2020 (Core A).
Information extraction from invoices. **Ahmed Hamdi**, Elodie Carel, Aurélie Joseph, Mickaël Coustaty and Antoine Doucet. The 16th International Conference on Document Analysis and Recognition.
- TPDF 2020 (Core B).
Assessing and Minimising the Impact of OCR Quality on Named Entity Recognition. **Ahmed Hamdi**, Axel-Jean Caurant, Nicolas Sidère, Mickaël Coustaty and Antoine Doucet. Digital Libraries for Open Knowledge - 24th International Conference on Theory and Practice of Digital Libraries. Springer, 2020, pp. 87–101. DOI: 10.1007/978-3-030-54956-5_7. https://doi.org/10.1007/978-3-030-54956-5_7 [SLIDES]
- CONLL 2020 (Core A).
Alleviating Digitization Errors in Named Entity Recognition for Historical Documents. Emanuela Boros, **Ahmed Hamdi**, Elvys Linhares Pontes, Luis Adrian Cabrera-Diego, Jose G. Moreno, Nicolas Sidère and Antoine Doucet. Proceedings of the 24th Conference on Computational Natural Language Learning, pp. 431–441. <https://www.aclweb.org/anthology/2020.conll-1.35>
- ICADL 2020 (Core A).
When to Use OCR Post-correction for Named Entity Recognition? Vinh-Nam Huynh, **Ahmed Hamdi** and Antoine Doucet. Digital Libraries at Times of Massive Societal Transition - 22nd International Conference on Asia-Pacific Digital Libraries. Lecture Notes in Computer Science 12504 (2020), pp, 33-42. DOI: 10.1007/978-3-030-64452-9_3. https://doi.org/10.1007/978-3-030-64452-9_3. [SLIDES]
- ICADL 2020 (Core A).
Entity Linking for Historical Documents: Challenges and Solutions. Elvys Linhares Pontes, Luis Adrian Cabrera-Diego, Jose G. Moreno, Emanuela Boros, **Ahmed Hamdi**, Nicolas Sidère, Mickaël Coustaty and Antoine Doucet. Digital Libraries at Times of Massive Societal Transition - 22nd International Conference on Asia-Pacific Digital Libraries. Lecture Notes in Computer Science 12504 (2020), pp, 215-231. DOI: 10.1007/978-3-030-64452-9_19. https://doi.org/10.1007/978-3-030-64452-9_19.
- JCDL 2019 (Core A*)
An Analysis of the Performance of Named Entity Recognition over OCRed Documents. Ahmed Hamdi, Axel Jean-Caurant, Nicolas Sidere, Mickaël Coustaty and Antoine Doucet. 19th ACM/IEEE Joint Conference on Digital Libraries, pp, 333-334. DOI: 10.1109/JCDL.2019.00057. <https://doi.org/10.1109/JCDL.2019.00057>.
- ICADL 2019 (Core A).
Impact of OCR Quality on Named Entity Linking. Elvys Linhares Pontes, **Ahmed Hamdi**, Nicolas Sidere and Antoine Doucet. Digital Libraries at the Crossroads of Digital Information for the Future - 21st International Conference on Asia-Pacific Digital Libraries. Lecture Notes in Computer Science 11853 (2019) pp, 102-111. DOI: 10.1007/978-3-030-34058-2_11. https://doi.org/10.1007/978-3-030-34058-2_11,
- DAS 2018 (Core B).
Feature Selection for Document Flow Segmentation. **Ahmed Hamdi**, Mickaël Coustaty, Aurélie Joseph, Vincent Poulain D'Andecy, Antoine Doucet and Jean-Marc Ogier. 13th IAPR International Workshop on Document Analysis Systems. IEEE Computer Society, pp, 245–250. DOI:10.1109/DAS.2018.66. <https://doi.org/10.1109/DAS.2018.66>.
- MT SUMMIT 2013 (Core B).
The Effects of Factorizing Root and Pattern Mapping in Bidirectional Tunisian - Standard Arabic Machine Translation. **Ahmed Hamdi**, Rahma Boujelbane, Nizar Habash and Alexis Nasr. MT Summit 2013. <https://hal.archives-ouvertes.fr/hal-00908761>.

3. International labs

- CEUR@CLEF 2022
Knowledge-based Contexts for Historical Named Entity Recognition Linking Documents. Emanuela Boros, Carlos-Emiliano González-Gallardo, Edward Giamphy, **Ahmed Hamdi**, Jose G. Moreno and Antoine Doucet. Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, 2022. <https://ceur-ws.org/Vol-3180/paper-84.pdf>.
- CEUR@CLEF 2020
Robust Named Entity Recognition and Linking on Historical Multilingual Documents. Emanuela Boros, Elvys Linhares Pontes, Luis Adrian Cabrera-Diego, **Ahmed Hamdi**, Jose G. Moreno, Nicolas Sidère and Antoine Doucet. Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, 2020. http://ceur-ws.org/Vol-2696/paper_171.pdf.

4. International workshops

- WML@ICDAR 2017
Machine Learning vs Deterministic Rule-Based System for Document Stream Segmentation. **Ahmed Hamdi**, Joris Voerman, Mickaël Coustaty, Aurélie Joseph, Vincent Poulain D'Andecy and Jean-Marc Ogier. First Workshop of Machine Learning, 14th IAPR International Conference on Document Analysis and Recognition, WML@ICDAR 2017. IEEE, pp, 77-82. DOI: 10.1109/ICDAR.2017.332 <https://doi.org/10.1109/ICDAR.2017.332>.
- WANLP@ACL 2015
POS-tagging of Tunisian Dialect Using Standard Arabic Resources and Tools. **Ahmed Hamdi**, Alexis Nasr, Nizar Habash and Nuria Gala. Proceedings of the 2nd Workshop on Arabic Natural Language Processing Processing, ANLP@ACL 2015. Association for Computational Linguistics. DOI: 10.18653/v1/W15-3207. <https://doi.org/10.18653/v1/W15-3207>.
- VarDial@COLING 2014
Automatically building a Tunisian Lexicon for Deverbal Nouns. **Ahmed Hamdi**, Nuria Gala and Alexis Nasr. Proceedings of the 1st Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects, VarDial@COLING 2014. Association for Computational Linguistics and Dublin City University, pp, 95-102. DOI: 10.3115/v1/W14-5311. <https://doi.org/10.3115/v1/W14-5311>.

5. non-peer reviewed

- arXiv preprint 2023.
DocILE Benchmark for Document Information Localization and Extraction. Štěpán Šimsa, Milan Šulc, Michal Uříčář, Yash Patel, **Ahmed Hamdi**, Matěj Kocián, Matyáš Skalický, Jiří Matas, Antoine Doucet Mickaël Coustaty and Dimosthenis Karatzas. <https://arxiv.org/abs/2302.05658>
- arXiv preprint 2023.
DocILE 2023 Teaser: Document Information Localization and Extraction. Štěpán Šimsa, Milan Šulc, Matyáš Skalický, Yash Patel and **Ahmed Hamdi**. <https://arxiv.org/abs/2301.12394>
- arXiv preprint 2021.
Named Entity Recognition and Classification on Historical Documents: A Survey. Maud Ehrmann, **Ahmed Hamdi**, Elvys Linhares Pontes, Matteo Romanello and Antoine Doucet. <https://arxiv.org/abs/2109.11406>
- *Named Entity Recognition and Linking*. **Ahmed Hamdi**, Elvys Linhares Pontes and Antoine Doucet. deliverable at NewsEye 2020.
- *Stance detection*. **Ahmed Hamdi**, Thi Tuyet Hai Nguyen and Antoine Doucet. deliverable at NewsEye 2020.

6. Thesis Report

- Ahmed Hamdi, 2015. *Traitement automatique du dialecte tunisien à l'aide d'outils et de ressources de l'arabe standard : application à l'étiquetage morphosyntaxique*. Aix-Marseille Université. <https://www.theses.fr/en/2015AIXM4089>.

7. National Conferences

- CORIA 2021
Atténuer les erreurs de numérisation dans la reconnaissance d'entités nommées pour les documents historiques. Emanuela Boros, **Ahmed Hamdi**, Elvys Linhares Pontes, Luis Adrian Cabrera-Diego, Jose G. Moreno, Nicolas Sidère and Antoine Doucet. Conférence franco-phone en Recherche d'Information et Application. <https://hal.science/hal-03320332/>
- TALN 2013
Un système de traduction de verbes entre arabe standard et arabe dialectal par analyse morphologique profonde. **Ahmed Hamdi**, Rahma Boujelbane, Nizar Habash and Alexis Nasr. Traitement Automatique des Langues Naturelles, TALN 2013. The Association for Computer Linguistics, pp, 395-406. <https://www.aclweb.org/anthology/F13-1029/>.
- RECITAL 2012
Apport de la diacritisation de l'analyse morphosyntaxique de l'arabe. **Ahmed Hamdi**. Proceedings of the Joint Conference JEP-TALN-RECITAL 2012. ATALA/AFCP pp, 247-254. <https://www.aclweb.org/anthology/F12-3019/>.