Ahmed Hamdi

Enseignant chercheur
La Rochelle Université, Laboratoire L3i
Bureau 013
ahmed.hamdi@univ-lr.fr
webpage | linkedin

Diplômes

Aix-Marseille université *2011-2015*
Doctorat en sciences du langage (très honorable)

Aix-Marseille université *2010-2011*
Master en sciences du langage spécialité traitement automatique des langues (mention bien)

École Nationale des sciences de l'informatique (ENSI - Tunisie) *2007-2010*
Ingénieur en informatique

Institut préparatoire aux études d'ingénieurs d'El Monastir (IPEIMo - Tunisie) *2005-2007*
Premier cycle en Mathématique - Physique

Activités de recherche

Enseignant-chercheur *2022-aujourd'hui / La Rochelle Université*

- Co-organisation de la compétition DocILE'23 dans l'extraction et la localisation d'information dans des documents commerciaux à ICDAR 2023 and CLEF 2023
- Conférencier invité au séminaire "*Digital Humanities Research Seminar*" in RGCL, Université de Wolverhampton, 19 May 2022: *Content Analysis of Digital Text with Special Focus on Named Entity Recognition and Linking* [Transparents]
- Reviewer: revues (LRE, JOCCH, CCOS, JCCE), conférences (CHR, DAS)

Post-doctorant *2017-2021 / La Rochelle Université*

- **Thèmes de recherche**
  - Reconnaissance et désambiguïsation des entités nommées dans la presse ancienne.
  - Détection des opinions envers les entités nommées.
  - Développement de modèles de plongement lexical de mots (*word embedding*) pour les langues peu dotées.
  - Extraction d'information à partir de documents commerciaux.
  - Segmentation de flux documentaires.
- **Encadrement**
  - Vinh-Nam Huynh en L3 INFO à l'université de Hanoi. Cet encadrement a mené à une publication à la conférence internationale ICADL en 2020.
  - Joris Voerman en M2 INFO à La Rochelle Université. Cet encadrement a mené à une publication au workshop WML@ICDAR en 2017.

Doctorant *2011-2015 / Aix-Marseille Université*

- **Co-directeurs de thèse** : Alexis Nasr et Núria Gala
- **Sujet** : Traitement du dialecte tunisien à l'aide d'outils et des ressources de l'arabe standard.
- **Date de soutenance** : 04/12/2015
- **Séjour scientifique** à l'université de Columbia (New York - USA), Center for Computational Learning Systems (CCLS), sous la supervision de Nizar Habash, Owen Rambow. **Sujet** : analyse morphologique des dialectes arabes. Mai-Sept 2014
- **Encadrement** de l'étudiant Haiddi Serraye en M1 INFO à Aix-Marseille Université. **Sujet** Développement de modèle de langage à l'aide de SRILM.

Stage de recherche *Fév 2011-Juin 2011 / Aix-Marseille Université*

- **Co-encadrement**: Alexis Nasr and Núria Gala
- Sujet: Apport de la diacritisation dans l'analyse morphologique de l'arabe standard.

Activités d'enseignement

Enseignant-chercheur
2022-aujourd'hui / La Rochelle Université

- Volume horaire: **296h**
- Unités d'enseignement: Mise en œuvre des systèmes Big Data, web sémantique (fouille de texte), fouille de processus, réseau et sécurité, architecture orientée service, Modélisation, développement web.
- **Responsabilités**
  - Responsable du Master INFO en alternance (syllabus *Architecte Logiciel*) à NiortTech / La Rochelle Université
  - Développement de relations entre les entreprises à Niort et La Rochelle Université.
  - Responsabilité de formation AI4industry à La Rochelle université
  - Responsable de cours *Mise en œuvre des systèmes big data* (M2, info), 5V-model, Hadoop, Spark, NoSQL databases, SQALA, KAFKA...
  - Responsable de cours *Technologies du web* (L2, Gestion), HTML and CSS basis, Wordpress...
  - Enseignement à l'université européenne (Eu-Conexus), Computer science for e-tourism (cours en anglais pour toutes les licences) web scrapping et extraction d'information...
  - Tutorat pédagogique (stages: L3, M1 et M2)

Vacataire
2017-2021 / La Rochelle Université

- Volume horaire: **56h**
- Unités d'enseignement: Computer science for e-tourism (cours en anglais), développement web.

Attaché temporaire d'enseignement et de recherche (ATER)
2014-2016 / Aix-Marseille université

- Volume horaire: **384h**
- Unités d'enseignement: Algorithmique, Introduction à l'informatique et la programmation, Programmation avancée en langage C, Outils informatiques et C2i, Programmation en langage Python, développement web, Automates et circuits.

Moniteur
2012-2014 / Aix-Marseille université

- Volume horaire: **128h**
- Unités d'enseignement: Introduction à l'informatique et la programmation, Programmation avancée en langage C, Outils informatiques et C2i, Automates et circuits.

Qualification et prix

- Qualification au poste de maître de conférence CNU 27 (informatique) en 2021
- CLEF-HIPE 2022: Named Entity Recognition (NER) and Entity Linking (EL) on Historical Newspapers,
  - NER: 1^{re} **place dans 20/26 classements**
  - EL: 1^{re} **place dans 6/8 classements**
    https://hipe-eval.github.io/HIPE-2022/results
- CLEF-HIPE 2020: Named Entity Recognition (NER) and Entity Linking (EL) on Historical Newspapers NER
  - NER: 1^{re} **place dans tous les classements.**
  - EL: 1^{re} **place dans 50/52 classements and** 2^{nd} **place in 2 classements.**
    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, 1^{er} **prix du Hackathon**
  https://opendata.agglo-larochelle.fr/

Autres activités et responsabilités

- Membre du comité d'organisation du congrès INFORSID (INFormatique des ORganisations and Systèmes d'Information et de Décision) à La Rochelle université.
- Membre du comité d'organisation de la conférence internationale CICLING (*International Conference on Intelligent Text Processing and Computational Linguistics*) 2019 à La Rochelle université.
- Co-organisateur de l'évènement local au laboratoire L3i "Journée des Ingénieurs, Doctorants, ATER et Post-docs" (JIDAP) en 2018 et 2019.
- Participation à l'évènement national "fêtes de la science" (La Rochelle université / L3i) en 2018, 2019, 2020 et 2021. https://www.univ-larochelle.fr/wp-content/uploads/pdf/FDLS-2020-PROGRAMME-PUBLIC-20200923.pdf
- Participation à l'évènement national "journées portes ouvertes" (La Rochelle université / L3i) en 2019 and 2020. [Transparents 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).

Récapitulatif des activités d'enseignement:

Année	Intitulé	Public	Discipline	CM	TD	TP
2022-2023	Mise en œuvre des systèmes big data	M2	INFO	30	–	66
	Fouille de processus	M2	INFO	10,5	10,5	13,5
	Web sémantique	M2	INFO	6	3	3
	Mise en situation professionnelle - Projet	M2	INFO	9	–	24
	Système d'exploitation	L1	INFO	–	–	27
	Technologies du web	L2	Gestion	12	–	12
2021-2022	Mise en œuvre des systèmes big data	M2	INFO	4	–	6
	Architecture orientée service	M1	INFO	3	9	–
	Réseaux et sécurité	M1	INFO	18	12	–
	Informatique pour le e-Tourisme	Toutes les licences		6	6	–
	Technologies du web	L2	Gestion	–	28	–
	Modélisation	L1	INFO	–	7,5	15
	Bases du web	L1	INFO	–	30	–
2020-2021	Big Data	M1	INFO	4	–	6
	Informatique pour le e-Tourisme	Toutes les licences		6	6	–
	Technologies du web	L2	Gestion	–	28	–
2019-2020	Technologies du web	L2	Gestion	–	28	–
2018-2019	Technologies du web	L2	Gestion	–	28	–
2017-2018	Technologies du web	L2	Gestion	–	14	–
2016-2017	Technologies du web	L2	Gestion	–	14	–
2015-2016	Algorithmique	L2	MATHS	–	24	20
	Introduction à l'informatique et la programmation	L1	INFO	20	24	21
	Outils informatiques et C2i	L1	INFO	–	24	–
	Python II	Licence 1	INFO	–	24	–
	Automates et circuits	L1	INFO	–	24	18
2014-2015	Algorithmique	L2	INFO	–	24	20
	Outils informatiques et C2i	L1	INFO	–	48	–
	Introduction à l'informatique et la programmation	L1	INFO	–	–	63
	C language	L1	INFO	–	18	21
	Web development I	L1	INFO	–	15	–
2013-2014	Introduction à l'informatique et la programmation	L1	INFO	–	–	42
	Outils informatiques et C2i	L1	INFO	–	24	–
2012-2013	Introduction à l'informatique et la programmation	L1	INFO	–	–	21
	Outils informatiques et C2i	L1	INFO	–	24	–
	Automates et circuits	L1	INFO	–	24	–

Statistiques sur mes publications

30 publications en total – **h-index=12** (Google Scholar, 2023)

- 3 articles de revues internationales
- 13 articles de conférences internationales
- 1 livre
- 3 articles dans des workshops internationaux
- 2 notes de travail dans des compétions internationales
- 3 articles de conférences nationales
- 5 articles publiés sur arXiv

Liste complète de publications

1. **Articles de revues internationales**

   - 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. **Articles de conférences internationales**

   - 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 44^{th} 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 16^{th} International Conference on Document Analysis and Recognition (ICDAR 2021).
- TPDL 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 - 24^{th} 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
- 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 24^{th} 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 - 22^{nd} 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.
- 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 - 22^{nd} 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 (Croe 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. 19^{th} 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 - 21^{st} International Conference on Asia-Pacific Digital Libraries.Lecture Notes in Computer Science 11853 (2019) pp, 102-11. 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. 13^{th} 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. **Articles de compétitions internationales**

- CEUR@CLEF 2022
  Knowledge-based Contexts for Historical Named Entity Recognition Linking. 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. **Articles d'ateliers internationaux**

- 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, 14^{th} 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 2^{nd} 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 1^{st} 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. **Articles non-peer reviewed**

- arXiv preprint 2023.
  DocILE Benchmark for Document Information Localization and Extraction. Štěpán Šimsa, Milan Šulc, Michal Uřičář, 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 Haï Nguyen and Antoine Doucet. deliverable at NewsEye 2020.

6. **Rapport de thèse**

- 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. **Articles de conférences nationales**

- 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 francophone 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/.