

## Alaa ABI HAIDAR (1982-)

Fortia	LIP6 . UPMC . CNRS	(T): +33 1 42 17 65 15
Senior Data Scientist	Couloir 25-26 . Bureau 503	(F) +33 1 42 17 74 62
17 Ave George V	4 Place Jussieu	<a href="mailto:alahay@alahay.org">alahay@alahay.org</a>
75008 Paris	75005 Paris	<a href="http://www.alahay.org">www.alahay.org</a>

### Formation

**Ph.D.** en Informatique [informatics] des systèmes complexes, Indiana University, 2011  
Directeur de Thèse: Luis M. Rocha. Commité: Filippo Menczer, Predrag Radivojac et Alessandro Flammini

**Ph.D.** en Informatique [computer science] (Mineur), Indiana University, 2011

**M.S.** en Informatique [computer science], Indiana University, 2005

**B.S.** en Informatique [computer science], American University of Beirut, 2003

### Experience d'Enseignement et d'Encadrement

Chargé des TD/TME et encadrant des projets de M1 et de M2	2011-2014
Université Pierre et Marie Curie	UPMC

Chargé des TD/TME du cours Methodes pour l'Intelligence Artificielle (MIA) de la specialisation IAD du master de l'UPMC et de plusieurs projets de M1 et M2

Chargé d'un Projet IAD sur l'extraction des données de la litterature sur l'immunologie pour le cours methodes pour l'intelligence artificielle niveau de M1 (3 étudiants encadrés), un projet sur l'histoire de l'immunologie avec l'aide de la fouille de données pour le cours immunologie des systèmes et le niveau de M1 (3 étudiants encadrés), un Projet d'Ingenieur (PING) pour developper un outil d'interface pour l'annotation manuel des entités nommées (1 étudiant encadré), un projet de Data Mining and Knowledge Management  $D_M K_M$  (Erasmus Mundus) pour developper une methode d'annotation automatique et non-supervisé des entités nommées (1 étudiant encadré), un autre projet d'Ingenieur (PING) pour developper un autre outil d'interface pour paramétriser notre algorithme de reconnaissance des entités nommées (2 étudiants encadrés) et un Projet de fin d'études de M2 pour améliorer notre methode d'annotation automatique et non-supervisé des entités nommées et la comparer a d'autres méthodes sur les données du CoNLL (1 étudiant encadré).

Attaché d'enseignement (Assistant Instructor)	2008–2011
Indiana University	Indiana University
Attaché d'enseignement pour les cours magistraux suivants: Simplicité de la complexité (I400), la gestion de l'information (I502), Introduction à la programmation en Python (I210) et la représentation de l'information (I308)	

### Experience de Recherche

Senior Data Scientist	02/2017–RIP
Fortia Financial Solutions	17 ave George V, 75008 Paris
Ma recherche porte sur la fouille, l'analyse et la classification de données textuelles et le traitement de la langue. Cela porte sur la reconnaissance d'entités nommées pour le developpement des applications	

mobiles avec l'aide de l'apprentissage non/semi/supervisé et le deeplearning dans les domaines de finances et de l'analyse des fonds et des réglementaires fintech regtech . [Fortia Financial Solutions](#).

Senior Data Scientist 07/2016–2017  
:snips 18 rue Saint Marc, 75002 Paris  
Ma recherche porte sur la fouille, l'analyse et la classification de données textuelles et le traitement de la langue. Cela porte sur la reconnaissance d'entités nommées pour le développement des applications mobiles avec l'aide de l'apprentissage non/semi/supervisé et le deeplearning. [:snips](#).

Chercheur Postdoctorant 2012 - 2016  
Jean-Gabriel Ganascia ACASA . LIP6 . CNRS UMR 7606 . UPMC .  
Paris  
Ma recherche porte sur la fouille, l'analyse et la classification de données textuelles. Actuellement, cela porte sur la reconnaissance des entités nommées dans la littérature française pour un projet de Labex OBVIL, dans les livres jeunesse avec plusieurs partenaires comme JOUVES et LeROBERT pour le projet LOCUPLETO, dans les bandes dessinées et les mangas pour la recommandations automatique des livres et des libraires avec ACTIALUNA, dans des milliers de revus du 19-20ème siècle avec la BnF et Europeana, et dans les articles scientifiques d'immunologie en PubMed pour mieux comprendre le comportement des systèmes et des sous-systèmes immunitaires avec l'équipe d'immunologie I2 à la Pitié Salpêtrière. Nous avons plusieurs partenaires qui financent indirectement ma recherche et celle de notre équipe [ACASA](#).

Chercheur Oct 2014 - Jan 2015  
Laurant Duploux ACASA . LIP6 . CNRS UMR 7606 . UPMC .  
Paris  
Ma recherche porte sur la reconnaissance des entités nommées supervisée et non=supervisée sur une masse de données en format ALTO XML avec plusieurs mées que nous et nos partenaires on développé.

Chercheur Postdoctorant 2011 - 2012  
Veronique Thomas-Vaslin et Adrien Six ACASA . LIP6 et I2D3 . CNRS . UPMC . UMR  
7211. Paris  
Ma recherche porte sur la fouille et la classification automatique de données, afin d'organiser et visualiser les connaissances de la littérature biomédicale, permettant de valider, de tester et de visualiser de nouvelles hypothèses concernant le système immunitaire et ses sous-systèmes [I3-Immuno.fr](#).

Assistant de Recherche (Research Assistant) 2006–2008  
Luis M. Rocha, Professeur Indiana University  
Classificateur de spam automatique inspiré par le système immunitaire, apprentissage automatique et classement de texte, intelligence artificielle et inspiré par des systèmes biologiques, systèmes complexes. [CASCI](#).

Assistant de Recherche (Research Assistant) Les étés de 2006–2010  
Luis M. Rocha, Professeur Instituto Gulbenkian de Ciência  
Modélisation multi-agent des lymphocytes régulateurs croisés et apprentissage automatique et classement des articles scientifiques selon leurs pertinence a l'interaction des protéines. [FLAD Collaboratorium](#) .

Assistant de Recherche (Research Assistant) 2005  
Katy Börner, Professeur Indiana University  
Visualization de données de taxonomie.

Assistant de Recherche (Research Assistant) 2004–2006  
Peter J. Ortoleva, Professeur é Indiana University  
développement de PHP and SQL, visualization de données, fouille et organisation de données,

réseaux de régulation transcriptionnelle et Gene Ontology au [CCVT](#).

Assistant de Recherche (Research Assistant)

2003

Rabih Sultan, Professeur

American University of Beirut

Traitement d'image, conception de logiciels et de web design [Page du département de Chimie](#) au AUB.

### Connaissances en Informatique

Expert: R, PHP, Python, SQL, HTML,  $\LaTeX$ , Linux, Windows, Flash, VI, NetLogo

Intermediat: Matlab, OpenGL, VTK, JavaScript, Java, Perl, CSS, Bash

### Langues

Parlées lues et écrites couramment: anglais, italien, français, arabe levantin (e.g. libanais), portugais, espagnol, arabe classique

Notions de base: hébreu biblique, arménien, grec moderne

### Prix

Vainqueur du prix du meilleur poster au Forum portugais sur Computational Biology (FPBC 2008)

Bourse de stage à l'Instituto Gulbenkian de Ciência (CIG) en 2006, 2007, 2008 et 2010

Bourses pour assister à AlifeX et ICARIS2010.

Vainqueur du prix du meilleur article à la conférence internationale des systèmes immunitaires artificiels (ICARIS 2010)

Vainqueur 1000 Euros de CFCAM et 2000 Euros from the RNSC pour une proposition à organiser un congrès [ImmunocomplexiT](#)

### Affiliations

Membre de la comité scientifique: Labex OBVIL à la Sorbonne.

Membre des comité: International Conference of Artificial Immune Systems (ICARIS) 2012 and ICARIS 2013

Membre des comité: European Conference of Artificial Life (ECAL) 2013, Taormina, Italy

Co-organisateur du Workshop Massive Data Flows a l'institut de Systemes Complexes a Paris.

Membre des comité: International Society of Molecular Biology (ISMB) 2014 et co-organisateur du Art and Science Exhibit, Boston

Membre du Panel: Artificial Life Alife 2014 Art and Science Exhibit, New York

Editorial board: Journal of Pattern Recognition and Intelligent Systems

## Publications Récentes

- A. Abi-Haidar, O. Albertini and J.G. Ganascia. A simple method for named entity recognition for English and French. ADAC. [in press]
- M. Riguet and A. Abi-Haidar. 2017. Faire figure d'autorité : l'analyse de réseaux appliquée au discours Authority Figures: Social Network Analysis applied to Literature <https://www.openscience.fr/Faire-figure-d-autorite-l-analyse-de-reseaux-appliquee-au-discours>
- Marine Riguet and Alaa Abi Haidar. Autorités et réseaux d'influence dans la critique littéraire de la seconde moitié du XIXe siècle, Estrema : interdisciplinaire magazine of humanities, n8, 2016, p. 195-220. URL : <http://www.estrema-cec.com/wp/?p=562>.
- Alaa Abi-Haidar, Bin Yang, Jean-Gabriel Ganascia [2016]. Mapping the First World War Using Interactive Streamgraphs. Sociology and Anthropology Vol. 4(1), pp. 12 - 16 DOI: 10.13189/sa.2016.040103
- Abi Haidar, Alaa; Ganascia, Jean-Gabriel [2016]. Automatic Detection of Characters in Case Insensitive Text in Comics. Digital Humanities DH/ADHO July 2016 Krakow, Poland
- Abi Haidar, Alaa; Yang, Bin; Ganascia, Jean-Gabriel [2015]. Extracting and Visualizing Named Entities using Interactive Streamgraphs A Case Study on First World War Data. Digital Humanities CSDH / SCHN, Canada 2015.
- C. Brando, F. Frontini, A. Abi Haidar and J.G. Ganascia. [2015] *Reconnaissance d'entités nommées : adaptation au domaine de la littérature française du XIXe siècle*. JLC 2015 Orleans, 2-4 Sept 2015.
- A. Abi-Haidar, O. Albertini and J.G. Ganascia [2015]. *A simple method for named entity recognition for English and French*. WILEY's Data Mining and Knowledge Discovery. In Press
- Abi-Haidar, M. Tufis, Ganascia, J.G. *From Inter-Annotation to Intra-Publication Inconsistency*. Chapter in Inconsistency Robustness. Edited by Carl Hewitt and John Woods. College Publications 2014. ISBN 978-1-84890-159-9
- Y. Mosallem, A. Abi-Haidar and J.G. Ganascia [2014]. *Unsupervised Named Entity Recognition and Disambiguation: An Application to Old French Journals*. Proceedings of ICDM 2014. St. Petersburg, Russia.
- Abi-Haidar, A., Six, A., Thomas-Vaslin, V., Ganascia, J.G. *The Artificial Immune Systems Domain Identifying Progress and Main Contributors using Publication and Co-Authorship Analyses*. Advances in Artificial Life, ECAL. Vol. 12. 2013.
- Alaa Abi Haidar. *Adaptive Immune-Inspired Text Classification*. Adaptive Spam Detection and Bio-Medical Article Classification Inspired by T Cell Cross-Regulation in the Immune System. LAP Lambert Academic Publishing ( 2013-03-07 ). ISBN-13: 978-3-659-36068-8
- A. Abi-Haidar [2011]. *An adaptive document classifier inspired by T-Cell cross-regulation in the immune system* (PDF). PhD Dissertation, Indiana University.
- A. Abi-Haidar and Luis. M. Rocha [2011]. *Collective Classification of Textual Documents by Guided Self-Organization in T-Cell Cross-regulation Dynamics*. Journal of Evolutionary Intelligence. Evolutionary Intelligence, DOI: 10.1007/s12065-011-0052-5
- A. Loureno, M. Conover, A. Wong, F. Pan, Alaa Abi-Haidar, A. Nematzadeh, H. Shatkay, and L.M. Rocha [2010]. *Testing Extensive Use of NER tools in Article Classification and a Statistical Approach for Method Interaction Extraction in the Protein-Protein Interaction Literature*. Proceedings of the BioCreative III Workshop 2010, Bethesda, Maryland, September 13-15, 2010.

A. Abi-Haidar and L.M. Rocha [2010]. *Biomedical Article Classification Using an Agent-Based Model of T-Cell Cross-Regulation*. In: Artificial Immune Systems: 9th International Conference, (ICARIS 2010). Winner of Best Paper Award. E. Hart et al (Eds.) Lecture Notes in Computer Science. Springer-Verlag, 6209, 237-249.

A. Abi-Haidar and L.M. Rocha [2010]. *Collective Classification of Biomedical Articles using T-Cell Cross-regulation*. In: Artificial Life XII: Twelfth International Conference on the Simulation and Synthesis of Living Systems. H. Fellersmann et al et al (Eds.). MIT Press, pp. 706-713.

A. Kolchinsky, A. Abi-Haidar, J. Kaur, A.A. Hamed and L.M. Rocha [2010]. *Classification of protein-protein interaction full-text documents using text and citation network features*. IEEE/ACM Transactions On Computational Biology And Bioinformatics, 7(3):400-411.

A. Kolchinsky, A. Abi-Haidar, J. Kaur, A.A. Hamed and L.M. Rocha [2009]. *Classification of protein-protein interaction documents using text and citation network features*. Proceedings of the BioCreative II.5 Workshop 2009: Special Session on Digital Annotations, Madrid, Spain, October 7-9, 2009. pp 34.

A. Abi-Haidar and L.M. Rocha [2008]. *Adaptive Spam Detection Inspired by a Cross-Regulation Model of Immune Dynamics: A Study of Concept Drift*. In: Proceedings of 7th International Conference on Artificial Immune Systems (ICARIS 2008). Doheon Lee, Peter Bentley, Sungwon Jung (Eds.) Lecture Notes in Computer Science. Springer-Verlag, Volume 5132/2008 36-47 doi:10.1007/978-3-540-85072-4.

A. Abi-Haidar and L.M. Rocha [2008]. *Adaptive Spam Detection Inspired by the Immune System*. In: Artificial Life XI: Eleventh International Conference on the Simulation and Synthesis of Living Systems. S. Bullock, J. Noble, R. A. Watson, and M. A. Bedau (Eds.). MIT Press, pp. 1-8.

A. Abi-Haidar, J. Kaur, A. Maguitman, P. Radivojac, A. Retchsteiner, K. Verspoor, Z. Wang, and L.M. Rocha [2008]. *Uncovering protein interaction in abstracts and text using a novel linear model and word proximity networks*. Genome Biology 2008, 9(Suppl 2):S11doi:10.1186/gb-2008-9-s2-s11.

A. Abi-Haidar, J. Kaur, A. Maguitman, P. Radivojac, A. Retchsteiner, K. Verspoor, Z. Wang, and L.M. Rocha [2007]. *Uncovering Protein-Protein Interactions in the Bibliome*. Proceedings of the Second BioCreative Challenge Evaluation Workshop (ISBN 84-933255-6-2).

Qu, K., A.E. Abi Haidar, J. Fan, D. Basu, G. Lin, L. Ensman, M. Jolly, P. Ortoleva. [2007]. *Cancer Onset and Progression: A Genome-Wide, Nonlinear Dynamical Systems Perspective on Onconet-works*. Journal of Theoretical Biology. Volume 246, Issue 2, 21 Pages 234-244

Sun, J., K. Tuncay, A.Abi Haidar, F. Stanley, M. Trelinski, and P. Ortoleva. [2007]. *Transcriptional Regulatory Network Discovery via Multiple Method Integration: Application to E.coli K12*. Algorithms in Molecular Biology, 2:2 doi:10.1186/1748-7188-2-2

Kagan Tuncay, Lisa Ensman, Jingjun Sun, Alaa Abi Haidar, Frank Stanley, Michael Trelinski and Peter Ortoleva [2006]. *Transcriptional regulatory networks via gene ontology and expression data*. In Silico Biology 7, 0003

R.Sultan, Z. Shreif, Lara, A. Abi-Haydar [2004]. *Taming ring morphology in 2D Co(OH)<sub>2</sub> Liesegang patterns*. Phys. Chem. Chem. Phys., 6, 3461 - 3466

### Autres ouvrages

A. Abi-Haidar, M. Tufis, J.G. Ganascia [2014]. *From Inter-Annotation to Intra-Publication Inconsistency*. Proceedings of Inconsistency Robustness 2014. Stanford University, USA.

Alaa ABI HAIDAR, Jean-Gabriel GANASCIA [2014]. Reconnaissance des entites nommees, application sur la litterature Francaise. ACFAS. 13-15 Mai 2014. Montreal

Alaa ABI HAIDAR, Jean-Gabriel GANASCIA [2014]. Mapping French Press to the Digital Age. at Digital Humanities 2014 Conference

Alaa ABI HAIDAR [2012]. Data Mining: What Can we Learn from the Immune Literature ?. October 23th 2012. Institut des Systemes Complexes Paris Ile-de-France. Complexity of the Immune system: from multi-scale data mining and integration to lymphocyte dynamics and repertoires modeling. Presentation Orale

Alaa ABI HAIDAR, Julien VIBERT, Maxime CLEMENT, Romain Gugert, David KRIEF, Adrien SIX, Veronique THOMAS-VASLIN, Jean-Gabriel GANASCIA [2012]. Lymphocyte Dynamics Event Mining from Immune Literature. In proceedings of the International Conference of Artificial Immune Systems ICARIS 2012. Taormina, Italy 2012. Presentation orale

Alaa ABI HAIDAR, Adrien SIX, Veronique THOMAS-VASLIN, Jean-Gabriel GANASCIA [2012]. How are we doing in the AIS domain?. In proceedings of the International Conference of Artificial Immune Systems ICARIS 2012. Taormina, Italy 2012. Presentation orale

Julien VIBERT, Alaa ABI HAIDAR, Phuong PHAM, Adrien SIX, Veronique THOMAS-VASLIN [2012]. A mathematical model of thymocyte proliferation reveals differences in T cell dynamics according to age and genetic background. In proceedings of the International Conference of Artificial Immune Systems ICARIS 2012. Taormina, Italy 2012. Presentation orale

Alaa Abi-Haidar [2011]. Data mining and information visualization: a new approach to capitalize from on line public opinion trends (Invited Speaker). Club of Venice Plenary 2011. Venice, Italy.

Alaa Abi-Haidar and Luis Rocha [2011]. Collective Classification of Textual Documents by Guided Self-Organization in T-Cell Cross-Regulation Dynamic (Contributed talk). NetSci-2011: International Conference on Network Science and Its Applications. Budapest, Hungary.

Alaa Abi-Haidar and Luis Rocha [2010]. Collective Classification of Textual Documents Using Self-Organized Cross-Regulatory T-cells in the Adaptive Immune System (Contributed talk). GSO2010 : Guided Self-Organization 2010. Bloomington, IN.

Alaa Abi-Haidar and Luis Rocha [2008]. Adaptive Spam Detection Inspired by a Cross-Regulation Model of Immune Dynamics (Poster). 1st Portuguese Computational Biology Forum (FPBC 2008). Winner of Best Poster Award.

## Presse et Media

IGC News (17/08/2010) [Best paper awarded to IGC scientist at the 9th ICARIS](#)

Telegraph (05/08/2008)

[Can we make software that comes to life?](#)

Science Daily (06/08/2008) [ALife Conference To Reveal Bio-inspired Spam Detection](#)

Silicon Republic (06/08/2008) [Fighting spam and thinking robots – bio-inspired AI](#)

Presse Portugaise (07/08/2008) Correio Manhã (p20 ext136180), Diário Notícias (p15 ext79040),  
Jornal Notícias (p5,28 ext150515), Público (p8 ext75000)

CiênciaHoje (11/08/2008) [Cientistas portugueses criaram aspirador de lixo electrónico](#)

Tecnomania (12/08/2008) [Podemos hacer software que nazca a la vida?](#)

L'atelier (22/08/2008) [Spams et bactéries : même combat !](#)

Telegraph (09/08/2008) [Spore: The science of Spore](#)