



Karim Sayadi

26 ans. Célibataire.

Formations

- 2013–2016 **Thèse de doctorat en informatique**, UPMC, Université Pierre et Marie Curie, Paris, France.
Actuellement en 3e année
- 2011–2013 **Master Cognitions naturelle et artificielle. Étude des systèmes complexes**, Ecole Pratique des Hautes Études, Paris, France.
Spécialité informatique. Mention très bien
- 2008–2011 **Licence fondamentale en informatique.**, Institut d'informatique et de mathématiques, Monastir, Tunisie.
Mention bien.

Sujet de thèse

- Titre *Construction d'algorithmes d'apprentissage automatique pour l'analyse des documents.*
- Description Ma thèse porte sur le développement d'algorithmes d'apprentissage automatique pour la fouille de textes capables d'exploiter le contenu sémantique d'une part et de proposer une analyse structurée des documents d'autre part. Le suivi dynamique de la structure thématique d'un document constitue un défi scientifique en raison du passage à l'échelle lié au développement de l'espace sémantique au cours du temps. Une seconde partie plus applicative de la thèse est orientée vers l'analyse des documents anciens où les techniques d'apprentissages automatique sont également déployés.
- Mots clés *Topic Modeling, Probabilistic Graphical Model, Machine Learning, Deep Learning, Digital Humanities, Natural Language Processing*
- Directeur Professeur Marc Bui, France

Autres formations

- Sept 2015 Ecole d'été du Helmholtz-Zentrum Berlin sur l'analyse des manuscrits anciens, Frauenthemsee, Germany (1 semaine).
- Juil 2014 Ecole d'été du labex Dynamite sur les systèmes complexes et la modélisation des territoires, Florence, Italie (1 semaine).
- Jan-Juil 2013 Semestre Michel Serres, heSam, École des Arts et Metier, École Pratique des Hautes Études. Sujet de travail : proposition d'un design pattern d'une ontologie interdisciplinaire. Paris, France. (6 mois)

Activités d'enseignement

- 2015-2016 TD sur l'utilisation de \LaTeX . École Pratique des Hautes Études (8h). Formation humanités numériques EPHE.
- 2014-2015 TD sur l'utilisation de processing. École Pratique des Hautes Études (12h). Formation humanités numériques EPHE.
- 2014-2015 TD sur la modélisation avec UML. École Pratique des Hautes Études (8h). Master 1 CNA-PC.
- 2014-2015 TD sur le traitement des informations spatiales avec python. Université Paris 8 (16h). Master 1 géomatique.
- 2013-2014 Cours sur les méthodes d'apprentissage automatique pour l'analyse des corpus de textes. École Pratique des Hautes Études (8h). Master 2 CNA-PC.

Activités de recherche

- Fév-Juin 2016 Séjour de recherche au sein du groupe de recherche sur l'analyse de documents, images et voix. Université de Fribourg, Suisse. (Pr. Rolf Ingold, Pr. Marcus Liwicki) (5 mois)
- Nov-Dec 2015 Séjour de recherche au sein du groupe d'analyse des données et calcul haute performance. Institut John Von Neumann. Université Nationale de Ho Chi Minh Ville, Vietnam. (Pr. Vu Duong) (1 mois)
- Sept 2015 Workshop sur l'analyse des documents historiques. Centre de recherche allemand pour l'intelligence artificielle. Kaiserslautern, Allemagne. (Pr. Marcus Liwicki) (1 semaine)
- Dec 2014 Workshop sur la modélisation du socio écosystème de l'île de Moorea. Université de Berkeley, Californie, États-Unis d'Amérique. (Dr. Joachim Claudet) (1 semaine)
- Nov-Dec 2014 Séjour de recherche au sein du groupe de recherche sur les technologies d'information et de communication. Institut polytechnique de Hanoi, Vietnam. (Associate Pr. Ha Quoc Trung) (1 mois)

Travaux de recherche

Mes travaux de recherche s'inscrivent dans le cadre du Knowledge Data Discovery (KDD) où il s'agit de fournir à l'utilisateur l'information pertinente en rendant efficace la phase de recherche d'informations.

Travaux effectués

Dans le cadre de ma thèse, j'ai développé des applications pour la détection et la visualisation sous la forme de réseau complexe, des similarités thématiques entre les différents documents dans un corpus de textes. Les documents sont représentés par des sommets et les liens thématiques entre les différents documents sont représentés par des arcs. Les méthodes employées permettent l'analyse de grands corpus de documents grâce à une distribution du calcul sur une grappe d'ordinateurs. Elles sont en mesure de traiter des tâches de catégorisations thématiques, multilingues, multi domaines, mais l'originalité de l'approche proposée consiste à aborder l'aspect dynamique des contenus à analyser en vue de produire une expertise automatisée de qualité. Il s'agit par exemple de détecter les centres d'intérêt exprimés sur la Blogosphère mondiale ou les corpus Twitter pour un suivre les tendances.

Travaux en cours de réalisation

Dans le cadre d'une collaboration scientifique avec l'université de Fribourg en Suisse, je travaille actuellement sur des solutions pour automatiser complètement ou partiellement la tâche de la reconnaissance des styles d'écritures dans un document ancien. J'approche ce problème avec une structure d'apprentissage profonde implémenté avec les réseaux de neurones.

Publications

- Avr 2016 Karim Sayadi, Marcus Liwicki, Rolf Ingold, Marc Bui,. *Tunisian Dialect and Modern Standard Arabic Dataset for Sentiment Analysis : Tunisian Election Context*, IEEE-CICLing (Computational Linguistics and Intelligent Text Processing) Intl. conference, Konya, Turkey, 7-8 Avril 2016.
- Dec 2015 Quang Vu Bui, Karim Sayadi, Marc Bui. *A multi-criteria document clustering method based on topic modeling and pseudoclosure function*, ACM-SOICT (Symposium on Information and Communication Technology) Intl. conference, Hue, Vietnam, 3-4 December 2015. (Extended version submitted to informatica journal)
- Juil 2015 Karim Sayadi, Quang Vu Bui, Marc Bui. *Multilayer classification of web pages using Random Forest and semi-supervised version of the Latent Dirichlet Allocation*, IEEE-I4CS (International Conference on Innovations for Community Services) Intl. conference, Nuremberg, Germany, 8-10 July 2015.
- Juil 2013 Karim Sayadi, Marc Bui, Michel Lamure. *Predictive topic modeling : Complex Networks approach using dynamics of author's communities*, EURO INFORMS (Operational Research), Rome, Italy, 1-4 July 2013.
- Mai 2012 Karim Sayadi, Marc Bui, Vigile Hoareau, Sofian Ben Amor, *Une approche prétopologique pour la catégorisation des données de microblogging*, Conférence nationale VSST'12 (Veille Scientifique et Technologique), Ajaccio, Corse, France, 24-25 Mai 2012.

Rapport de recherche

- Oct 2013 Marc Bui, Karim Sayadi. *Modèle de recommandation de contacts basé sur l'analyse thématique des échanges*. Rapport du projet n 45 Nexboo. Convention Techno Pole de la Réunion. 20 Octobre 2013

Présentation et séminaires

- Nov 2015 Towards an automatic annotation in the computer science ontologies. University El Manar et Institute Pasteur. Tunis, Tunisie. 19-20 November 2015
- Sept 2015 Poster : Multilayer classification of the text in ancient documents. Frauenchiemsee, Germany. 10 September 2015
- Dec 2014 Data science, and modeling tools in investigating social-ecological systems. IDEA project, University of California Berkley, California, USA. 9 December 2014
- Juin 2014 Modeling the Seine Estuary with an ontology, Global Estuaries Forum, Deauville, France. 1 July 2014
- Oct 2013 Identification des thématiques dans les corpus de textes non structurés : application à un corpus santé (QALY), Conférence intl, VSST 2013, Nancy, France, 24 Octobre 2013.

Projets de recherche

- Dec 2014 Moorea IDEA. <http://mooreaidea.org>.
- Fév 2016 HisDoc DIVA Group <https://diuf.unifr.ch/main/diva/research/research-projects/hisdoc-2-towards-computer-assisted-palaeography>

Expériences professionnelles

- Juil-Oct 2013 Système de recommandation d'amis basé sur l'analyse des interactions sur les réseaux sociaux. Nexboo. Paris, France (4 mois)
- Fév-Juil 2013 Construction d'un référentiel conceptuel pour la recherche d'informations autour de l'estuaire de la Seine. École des Arts et Métiers. Paris, France (6 mois)
- Sept-Dec 2012 Approche prétopologique pour la catégorisation de textes. École Pratique des Hautes Études. Stage effectué au sein du laboratoire Laisc. Paris, France (4 mois)
- Fév-Juil 2012 Analyse des réseaux sociaux et visualisation des interactions entre utilisateurs. Semdee. Stage effectué au sein du laboratoire Laisc. Paris, France (6 mois)
- Juil-Aout 2010 Construction de base de données virtuelles. Metro Group IT. Düsseldorf, Allemagne (2 mois)

Compétences informatiques

- Langage PYTHON, JAVA, R, HTML, C
- Utilitaires LaTeX, BibTex, IPe scientifique
- Logiciels Matlab, Anylogic, Scilab, Scikit learn, Hadoop, Spark, Lucene
- Systèmes UNIX, Linux, Microsoft Windows
- Production Pretopological Semantic Analyzer (PSA for health). Déposé à l'agence de protection des programmes le 18/10/2012.

Langues

- Langue maternelle **Arabe, Français**
- Maitrisée **Anglais**
- Basique **Allemand**

Vie associative & centres d'intérêt

- 2012-2014 Responsable logistique et événementiel pour les activités culturelles à la cité international universitaire de Paris.
- 2010 Fondateurs et président d'honneur du club des logiciels libres à L'ISIMM Monastir

Centres d'intérêt

- Cuisine
- Water Polo, cyclisme
- Philosophie

Madame, Monsieur,

C'est avec intérêt que je candidate au poste d'attaché de recherche et d'enseignement à l'École Pratique des Hautes Études. Je suis en troisième année de thèse, inscrit à l'université de Pierre et Marie Curie. Je réalise mes travaux de recherche au sein de l'équipe CHArt EA4004 à l'EPHE sous la direction du Professeur Marc Bui. Mon sujet de recherche porte sur la conception d'algorithmes d'apprentissage automatique pour la fouille de texte. Ces algorithmes sont appliqués à l'analyse sémantique et à l'analyse des documents anciens.

J'ai eu l'opportunité au sein de diverses formations (Master BSE spécialité CNA, M1 & M2) d'exposer mes travaux de recherche sur l'analyse de corpus de textes. J'ai eu l'occasion d'assurer quelques cours et TD, dans le cadre de diverses formations à l'EPHE et à l'université de Paris 8, concernant divers langages de programmation et de modélisation (python, processing, java, L^AT_EX, UML).

Au cours de ma thèse, j'ai pu constater que l'environnement de l'enseignement et de la recherche au sein de l'EPHE est un environnement riche en collaborations scientifiques, ceci me motive à adresser ma candidature en tant qu'ATER.

En espérant que mon dossier retiendra votre attention et en restant à votre disposition pour vous apporter, si vous le souhaitez, des précisions supplémentaires, je vous prie d'agréer, Madame, Monsieur, l'expression de mes salutations les plus distinguées.

Karim Sayadi