

PERSONAL INFORMATION

Mathieu Roche



📍 UMR TETIS, 500, rue J.F. Breton, 34093 Montpellier (France)

✉ mathieu.roche@cirad.fr

🌐 <http://textmining.biz/Staff/Roche>

Sex Male | Date of birth 1975 | Nationality French

WORK EXPERIENCE

01/10/2013–Present

Senior Research Scientist

Cirad, Montpellier (France)

- **Head of SISO group** (i.e. Spatial Information, Modelling, Data Mining, and Knowledge Extraction) at TETIS
- **Member of Scientific Committee** at [LIRIMA](#)
- **Leader of several academic and industrial projects in text-mining** (e.g. Songes - FEDER, Animitex - CNRS, Senterritoire - MSH-M)
- **Supervision of 12 PhD students** at TETIS
- **Co-chair of 13 conferences and workshops** since 2013
- **Publication of more than 100 papers** in conferences and journals (since 2013): [Agritrop-Cirad](#)

01/09/2005–30/09/2013

Associate Professor (Maître de Conférences)

Université Montpellier 2, Montpellier (France)

- **Co-head** (2010-2012) and **head** (2012-2013) of **TEXTE group** (LIRMM)
- **Head of the Computer Science Master *Intégration de Compétences***, Univ. Montpellier 2 (2008-2011)
- **Elected member of Scientific Committee** at LIRMM (2010-2012)
- **Leader of several projects in text-mining** (e.g. PEPS-CNRS ECOMESS, PEPS-CNRS RESENS).
- **Supervision of 3 PhD students** at LIRMM
- **Co-chair of 5 conferences and workshops** (2005-2013)
- **Publication of more than 100 papers** in conferences and journals (2005-2013): [HAL-LIRMM](#)

EDUCATION AND TRAINING

01/09/2001–13/12/2004

PhD in Computer Science

Université Paris-Sud, Orsay (France)

Title: Intégration de la construction de la terminologie de domaines spécialisés dans un processus global de fouille de textes [*in French*]

Supervisor: Yves Kodratoff (CNRS)

01/01/2011–09/12/2011

HDR in Computer Science (accreditation to supervise research)

Université Montpellier 2, Montpellier (France)

Title: Fouille de Textes : de l'extraction des descripteurs linguistiques à leur induction [*in French*]

PERSONAL SKILLS

Mother tongue(s) French

Foreign language(s)

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	B2	B2	B1	B2	C1
Spanish	B1	B1	A2	A2	A2

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

Organisational / managerial skills

- **Head of research groups** (TEXTE at LIRMM and SISO at TETIS)
- **Head of a Computer Science Master:** [Description](#)
- **Leader of academic and industrial projects**
- **Supervisor of PhD students**

ADDITIONAL INFORMATION

Research topics:

- Text Mining
- Natural Language Processing (NLP)
- Information Retrieval and Machine learning

Applications

Current applications in agriculture domain ([Description](#))

- Animal disease surveillance [1,6]
- Information Extraction from experimental data of agricultural domain [5]

Current applications in biomedical domain:

- Biomedical term extraction [4]
- Word-sense induction [7]

Projects

- Leader of **SONGES project** (2016-2019) [project supported by Occitanie and European Union (FEDER)]. *Topic: Mining and matching heterogeneous data based on 3 dimensions: spatial, thematic, and temporal.*
- Co-scientific leader of **BigDataPol project** (2017) [project supported by Cirad]. *Topic: Using Big Data for Policy Management.*
- Leader of CNRS-Mastodons **ANIMITEX project** (2013-2014) / Scientific direction with M. Teisseire (UMR TETIS) [selection of 5 projects over 20 proposals (25%) by CNRS]. *Topic: Image analysis based on textual information.*
- Leader of CNRS-PEPS **ECOMESS project** (2013) [selection of 23 projects over 113 proposals (20%) by CNRS] *Topic: Contrastive analysis of emotions contained in short messages.*
- Leader of **Senterritoire project** (2012-2014) [project supported by the Scientific Committee of MSH-M]. *Topic: Sentiment analysis and geospatial information.*
- Leader of CNRS-PEPS **RESENS project** (2010-2011) [selection of 10 projects over 69 proposals (15%) by the INS2I institut of CNRS]. *Topic: Extraction of lexical knowledge from text by exploiting the spontaneous descriptions.*

Publications

List of publications:

- [DBLP](#)
- [ORCID](#)

Selection of publications:

- [1] E. Arsevska, S. Valentin, J. Rabatel, J. De Goër de Hervé, S. Falala, R. Lancelot, M. Roche. Web monitoring of emerging animal infectious diseases integrated in the French Animal Health Epidemic Intelligence System. *PLoS One* 13 (8), 25 p., 2018
- [2] S. Zenasni, E. Kergosien, M. Roche, M. Teisseire. Spatial information extraction from short messages. *Expert Systems with Applications, Elsevier*, 95:351-367, 2018
- [3] J. C. Valverde-Rebaza, M. Roche, P. Poncelet, A. de Andrade Lopes. The role of location and social strength for friendship prediction in location-based social networks. *Information Processing and Management, Elsevier*, 54 (4):475-489, 2018
- [4] J.A. Lossio-Ventura, J. Bian, C. Jonquet, M. Roche, M. Teisseire. A novel framework for biomedical entity sense induction. *Journal of Biomedical Informatics, Elsevier*, 84:31-41, 2018
- [5] S.L. Berrahou, P. Buche, J. Dibie, M. Roche. Xart: Discovery of correlated arguments of n-ary relations in text. *Expert Systems with Applications, Elsevier*, Volume 73, p.115-124, 2017
- [6] E. Arsevska, M. Roche, P. Hendriks, D. Chavernac, S. Falala, R. Lancelot, B. Dufour. Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web. *Computers and Electronics in Agriculture, Elsevier*, 123: 104-115, 2016
- [7] J.A. Lossio Ventura, C. Jonquet, M. Roche, M. Teisseire. Biomedical term extraction: overview and a new methodology. *Information Retrieval Journal, Springer*, 19(1-2): 59-99, 2016
- [8] H. Saneifar, S. Bonniol, P. Poncelet, M. Roche. Enhancing passage retrieval in log files by query expansion based on explicit and pseudo relevance feedback. *Computers in Industry, Elsevier*, 65(6): 937-951, 2014
- [9] E. Kergosien E., B. Laval., M. Roche, M. Teisseire. Are opinions expressed in land-use planning documents? *International Journal of Geographical Information Science, Taylor & Francis*, p.739-762, Volume 28, Issue 4, 2014
- [10] C. Lopez, V. Prince, M. Roche. How can catchy titles be generated without loss of informativeness? *Expert Systems with Applications, Elsevier*, p.1051-1062, Volume 41, Issue 4, 2014

Conferences

- Co-chair (Program Committee) of Spatial Analysis and GEOMatics Conference (**SAGEO'2018**), Montpellier
- Co-chair of the national workshop **Humans'2018** - workshop of SAGEO conference, Montpellier
- Co-chair of the Special Session "Environmental and geo-spatial data analytics (**EnGeoData**)" IEEE International Conference on Data Science and Advanced Analytics - DSAA'2014 (Shanghai), DSAA'2015 (Paris), DSAA'2016 (Montreal), DSAA'2017 (Tokyo), and DSAA'2018 (Turin)
- Co-chair of the Special Session "Learning from heterogeneous data analytics (**EAST'2017**)" - FLAIRS'2017 - AAAI
- Co-chair of the Session "Knowledge Organization and Mining" - **CARI'2014/16/18** (African Conference on Research in Computer Science and Applied Mathematics)
- Co-chair (Program Committee) of the 19th International Conference on Application of Natural Language to Information Systems (**NLDB'2014**), Montpellier.
- Co-chair of the national workshop **CerGEO'2014** - workshop of EGC conference
- Co-chair of the international workshop **GeoDoc'2012** - workshop of PAKDD'2012 conference
- Co-chair of the national workshop **FODOP'2008** (Opinion Mining / Fouille de Données d'OPinions) - workshop of INFORSID'2008 conference
- Co-chair of the national workshop **EvalECD'2009, EvalECD'2010, EvalECD'2011** - workshop of EGC conferences
- **Creation in 2005 with J. Azé (LRI - Orsay) of the French Text Mining Challenge: DEFT (DEFT Fouille de Textes)**. Since its creation in 2005, the DEFT challenge is organized each year.

Editions **Edition of book**

- Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration. Prince V., Roche M. (Eds). *Medical Information Science Reference, IGI Global*, 460 pages, March 2009.

Edition of proceedings of international conference

- Natural Language Processing and Information Systems - 19th International Conference on Applications of Natural Language to Information Systems, NLDB 2014, Montpellier, France, June 18-20, 2014. Proceedings. Métais E., Roche M., Teisseire M. (Eds). *Lecture Notes in Computer Science (LNCS), Springer Verlag*, Vol. 8455, 268 pages, 2014.

Edition of proceedings of international workshop

- Proceedings of the First Workshop on Modeling, Learning and Mining for Cross/Multilinguality (MultiLingMine 2016) co-located with the 38th European Conference on Information Retrieval (ECIR 2016) Ienco D., Roche M., Romeo P., Rosso P., Tagarelli A. *CEUR Workshop Proceedings (CEUR-WS.org)*, Vol. 1589, 65 pages, 2016.

Edition of special issues of journals

- Special issue on Environmental and Geospatial Data Analytics. Inkpen D., Roche M., Teisseire M. (Eds). *International Journal of Data Science and Analytics*. Volume 5, Issue 2–3. Springer, 2018
- Special Issue of 19th International Conference on Applications of Natural Language Processing to Information Systems. Roche M., Teisseire M., Métais E. (Editors) *Special issue of DKE (Data & Knowledge Engineering)*, Elsevier, Volume 100, Part B, Pages 189-254, 2015
- TAL et réseaux sociaux. Farzindar A., Roche M. (Rédacteurs invités). *Special issue of TAL (Traitement Automatique des Langues)*, volume 54, Numéro 3, 2013
- Mesurer et évaluer la qualité des données et des connaissances. Azé J., Béchet N., Berti-Equille L. Guillaume S., Roche M., Sais F. (Rédacteurs invités). *Special issue of RNTI (Revue des Nouvelles Technologies de l'Information)*, volume E-22, 2012
- Fouille de Données d'Opinions. Roche M., Poncelet P. (Rédacteurs invités). *Special issue of RNTI (Revue des Nouvelles Technologies de l'Information)*, 202 pages, volume E-17, 2009