

**MATHIEU ROCHE,**  
Né le 30 octobre 1975 à Tours (37)

MAÎTRE DE CONFÉRENCES  
À L'UNIVERSITÉ DE MONTPELLIER 2

Équipe-Projet TAL, LIRMM  
UMR 5506, CNRS, Université Montpellier 2  
Courriel : [mroche@lirmm.fr](mailto:mroche@lirmm.fr)  
Web : <http://www.lirmm.fr/~mroche/>



## CURSUS

---

### • EXPÉRIENCE PROFESSIONNELLE

- Depuis septembre 2005 : **Maître de Conférences**  
à l'Université Montpellier 2, Faculté des Sciences, Département Informatique  
**Titulaire de la PES** (Prime d'Excellence Scientifique) depuis décembre 2009.
- 2004-2005 : **Demi poste d'ATER** (Attaché Temporaire de l'Enseignement et de la Recherche)  
à l'Université Paris XI (Orsay), Faculté des Sciences, Département Informatique
- 2001-2004 : **Allocataire de recherche, Moniteur en informatique**  
à l'Université Paris XI dans l'équipe Inférence et Apprentissage (IA) du Laboratoire de Recherche en Informatique (LRI).

### • FORMATION

- Décembre 2004 : **Doctorat en Informatique** à l'Université Paris XI, mention Très Honorable.  
Titre : "Intégration de la construction de la terminologie de domaines spécialisés dans un processus global de fouille de textes".  
Jury : C. Froidevaux (Présidente), C. Vrain (Rapporteuse), A. Napoli (Rapporteur), Y. Toussaint (Rapporteur avec A. Napoli), P. Sébillot (Examinatrice) et Y. Kodratoff (Directeur de thèse).
- 2001 : **DEA Information Interaction Intelligence** à l'Université Paris XI (Orsay), mention Bien.

## ANIMATION RECHERCHE & ENSEIGNEMENT (DEPUIS 2005)

---

### • RESPONSABLE ET MEMBRE DE PROJETS DE RECHERCHE :

- **Porteur du projet PEPS RESENS** – 2009-2010.  
Dans le cadre des Projets Exploratoires Pluridisciplinaires (PEPS) mis en place par le CNRS, l'Institut INS2I a sélectionné 10 projets sur 69 soumis (15%).  
*Recueil des descriptions spontanées pour accéder automatiquement au sens des mots.*
- **Porteur du projet ProSigles** – 2007-2008.  
Projet sélectionné par le Conseil Scientifique de l'Université Montpellier 2.  
*Extraction et gestion des sigles issus de textes plus ou moins spécialisés.*
- **Porteur du projet TSAL** (Traitement Statistique et Algorithmique du Langage) – 2006.  
Projet "Pluridisciplinaires 2006" - Département STICS de l'Université Montpellier 2.  
*Extraction des connaissances à partir de corpus écrits en français médiéval.*
- **Membre de l'ANR Forum** (2005-2008) portée par Z. Bellahsene (LIRMM).  
*Extension des techniques d'intégration pour permettre le partage de données dans un environnement à grande échelle.*

### • TRANSFERT DE TECHNOLOGIE :

- **Accompagnement de Projets Industriels** (part du LIRMM en K€)
  - \* WebReport – 2010 (16 K€) – responsable scientifique
  - \* Nevantropic – 2009/2010 (8 K€) – co-responsable scientifique
  - \* Expernova – 2009/2010 (30 K€)
  - \* Satin – 2008/2011 (40 K€)
  - \* Itesoft – 2008 (5 K€) – responsable scientifique
- **Accompagnement de Projets Incubés du Languedoc-Roussillon Incubation** (part du LIRMM en K€)
  - \* EvalAcces (porteur M. Faure) – incubé en 2008 (12 K€)
  - \* Namae Concept (porteuse B. Laurent) – incubé en 2007 (4 K€)

● RESPONSABLE & MEMBRE DE COMITÉS D'ORGANISATION :

- **Co-responsable avec F. Saïs (LRI - Orsay) et N. Béchet (LIRMM - Montpellier) des ateliers EvalECD'09 et EvalECD'10** dans le cadre des conférences EGC'09 et EGC'10

*Confrontation des méthodes d'évaluation dans un processus d'Extraction de Connaissances à partir de Données (ECD)*

- **Co-responsable avec P. Poncelet (LGI2P - Nîmes) de l'atelier FODOP'08** (FOuille des Données d'OPinions) dans le cadre de la conférence INFORSID'08

*Développement d'une communauté pluridisciplinaire (fouille de données, aide à la décision, modélisation des connaissances, TAL, linguistique, etc) qui s'intéresse au traitement des données d'opinion.*

- **Membre du Comité d'Organisation de la conférence EDA'09**

*Présentation de travaux de recherche innovants liés aux Entrepôts de Données.*

- **Membre du Comité d'Organisation du challenge "Analyzing Web Traffic" dans le cadre des conférences ECML/PKDD'07**

*Classification du trafic HTTP afin de distinguer les différents types d'attaques et le trafic normal.*

- **Concepteur avec J. Azé (LRI - Orsay) du premier Défi Francophone de Fouille de Textes : DEFT'05.** Responsable avec J. Azé du Comité d'Organisation de DEFT'05 et DEFT'06.

*DEFT'05 (Défi Fouille de Textes) est le premier défi francophone du domaine. L'édition 2005 de DEFT consistait à déterminer les phrases issues d'un corpus d'allocutions de F. Mitterrand présentes dans un corpus de discours de J. Chirac. Onze équipes représentant neuf laboratoires ont participé. La campagne DEFT'06 organisée dans le cadre de la Semaine du Document Numérique à Fribourg en Suisse avait pour but de déterminer les segments thématiques à partir de trois corpus de domaines différents (politique, juridique, scientifique). Depuis sa création, le défi DEFT est organisé annuellement (participation en tant que membre du CP de 2007 à 2010).*

● ÉDITION DE REVUES ET D'OUVRAGES :

- **"Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration"**, IGI Publishing, Medical Information Science Reference, 460 pages, 2009 [V. Prince, M. Roche Ed.]
- **Numéro spécial de la revue RNTI** (Revue des Nouvelles Technologies de l'Information). "Fouille de Données d'Opinions, volume E-17, 202 pages, 2009 [M. Roche, P. Poncelet Ed.]
- **Numéro spécial de la revue RNTI.** "Défi Fouille de Textes : reconnaissance automatique des auteurs de discours - Campagne DEFT'05 (TALN'05)", 148 pages, volume E-10, 2007 [V. Prince, Y. Kodratoff, J. Azé, M. Roche Ed.]

● ENCADREMENT DE STAGES ET THÈSES :

– **Thèses de Doctorat :**

- \* Thèse UM2/Région de Cédric Lopez co-encadrée avec V. Prince (30%) – depuis octobre 2009.  
*Titre : Titrage automatique.*
- \* Thèse CIIFFRE de Hassan Saneifar co-encadrée avec P. Poncelet (30-40%)– depuis octobre 2008. Thèse en collaboration avec la société Satin-IP  
*Titre : Extraction d'information à partir de Logs.*
- \* Thèse BDI CNRS/Région de Nicolas Béchet co-encadrée avec J. Chauché (10%) – 2006-2009. Thèse soutenue le 08 décembre 2009.  
*Titre : Extraction et regroupement de descripteurs morpho-syntaxiques pour des processus de Fouille de Textes.*

– **Stages Recherche (Master 2) :**

- \* Stage de Walid Belhaoues – 2008-2009 – co-encadré avec V. Prince (LIRMM), S. Léon (LIRMM), M. Faure (Open-S/EvalAcces)  
*Thème : Titrage automatique de pages Web*
- \* Stage de Miguel Macias – 2007-2008 – co-encadré avec V. Prince (LIRMM), A. Labadié (LIRMM)  
*Thème : Titrage automatique de pages Web*
- \* Stage de Ali Harb – 2007-2008 – co-encadré avec P. Poncelet (LGI2P) , M. Plantié (LGI2P)  
*Thème : Fouille de données d'opinion*
- \* Stage de Sami Laroum – 2007-2008 – co-encadré avec N. Béchet (LIRMM)  
*Thème : Recherche d'Information*
- \* Stage de Nicolas Boffo – 2007-2008 – co-encadré avec A. Laurent (LIRMM), J. Bres (PRAXILING)  
*Thème : Extraction d'information linguistique*
- \* Stage de Larbi Mesbahi – 2005-2006 – co-encadré avec M. Teisseire (LIRMM), J. Chauché (LIRMM)  
*Thème : Recherche d'Information*
- \* Stage de Tang Yi Fei – 2005-2006 – co-encadré avec Z. Bellahsene (LIRMM)  
*Thème : Mise en correspondance de schémas*

– **Stages de fin d'étude d'Ingénieur :**

- \* Stage d'ingénieur CNAM de Pierre Nghiem – 2009-2010 – co-encadré avec P. Poncelet (LIRMM)  
*Thème : Entrepôts de données textuelles*

- \* Stage de fin d'étude d'ingénieur d'Ines Bayouh – INSAT Tunisie / Société Paperblog - 2007-2008 – co-encadré avec N. Béchet (LIRMM)  
*Thème : Recherche d'Information*
- \* Stage d'ingénieur CNAM d'Emmanuel Cazal – 2006-2007 – co-encadré avec A. Laurent (LIRMM), M. Teisseire (LIRMM)  
*Thème : Extraction de la terminologie*
- **Stage Post Doctoral :**
  - \* Stage de Feirouz Chakkoura - 2008-2009 (3 mois) – co-encadré avec S. Bringay (LIRMM), M. Teisseire (LIRMM)  
*Thème : Recherche d'Information*
- **TER stages et projets tutorés (Masters 1 et 2) :**
  - \* Stages de quatre étudiants en Master 1 Informatique – 2005-2010 – co-encadrés avec A. Mela (Univ. Montpellier 3), A. Laurent (LIRMM), B. Laurent (Namae Concept), S. Jaillet (CrysaLEAD), M. Hascoët (LIRMM), V. Prince (LIRMM), M. Faure (Open-S/EvalAcces)
  - \* Stages de trois ou quatre étudiants en Master 2 Informatique – 2007-2008 – co-encadrés avec A. Laurent (LIRMM), B. Laurent (Namae Concept), S. Jaillet (CrysaLEAD)
  - \* Projets tutorés en Master 1 Informatique (Intégration de Compétences) – 2006-2010 – co-encadrés avec A. Laurent (LIRMM), B. Laurent (Namae Concept), I. Mougnot (LIRMM)
  - \* Stages Polytech – 2005-2009 – co-encadrés avec M. Teisseire (LIRMM), A. Laurent (LIRMM), B. Laurent (Namae Concept)
- **JURY DE DOCTORAT ET MASTER :**
  - **Jury de Doctorat** (hors doctorants co-encadrés) :
    - \* Dong (Haoyuan) Li (Univ. Montpellier 2) – septembre 2009
    - \* Rémy Kessler (Univ. Avignon) – juillet 2009
  - **Rapporteur de stages de M2 Informatique** (hors LIRMM/Univ. Montpellier 2) :
    - \* rapporteur extérieur de stages du Master Recherche (Univ. Rennes) - depuis 2009
    - \* rapporteur extérieur de stages du Master ECD (Univ. Lyon 2) - depuis 2007
- **RESPONSABILITÉS PÉDAGOGIQUES D'ENSEIGNEMENT (UNIVERSITÉ MONTPELLIER 2) :**
  - Depuis 2008 : Co-responsable du module "Stages d'analyse ou Projets Tutorés" – Master Informatique IC
  - Depuis 2007 : Co-responsable du module "Extraction de Connaissances dans les Données" – Master Informatique IFPRU
  - Depuis 2007 : Co-responsable du module "Fouille de Données Avancée" – Master Informatique IFPRU
  - Depuis 2007 : Responsable du module "Fouille de Données" – Master Informatique IC
  - depuis 2007 : Co-responsable du module "Administration des Bases de Données" – Master Informatique IC
  - Depuis 2007 : Co-responsable du module de "Systèmes d'Information et Bases de Données" – Licence Informatique L3
  - Depuis 2006 : Co-responsable du module "Systèmes d'Information et Bases de Données" – Master Informatique IC
  - 2006-2007 : Co-responsable du module de "Systèmes d'Information et Bases de Données" – Licence Informatique L2
  - 2005-2007 : Responsable du module "Fouille de Données" – Master Professionnel Informatique
- **RESPONSABILITÉS ADMINISTRATIVES D'ENSEIGNEMENT :**
  - Depuis 2008 : **Responsable de la Spécialité "Intégration de Compétences" (IC) – Master Informatique**  
*La mention Informatique de l'Université Montpellier 2 comporte deux spécialités. Une des spécialités, Intégration de Compétences, a pour but d'apporter, au cours d'un cursus de deux ans, les connaissances en informatique essentielles pour acquérir une double compétence. Cette spécialité comporte trois parcours : "Compétences Complémentaires en Informatique", "Bioinformatique", "Système d'Information Géographique" et mutualise des modules avec le parcours "Physique-Informatique" de la mention Physique. En moyenne, il comporte une soixantaine d'étudiants en M1 et une cinquantaine d'étudiants en M2. Site web du Master IC : [www.lirmm.fr/IC](http://www.lirmm.fr/IC)*
  - Depuis 2006 : **Responsable du Parcours CCI** (Compétences Complémentaires en Informatique) – Master Informatique Spécialité IC
  - Depuis 2006 : Responsable des stages de la Spécialité IC – Master Informatique
  - 2008-2009 : Responsable du Certificat Professionnel CP09 : "Webmestre : Conception de Sites Administration et Serveurs Web" – CNAM Languedoc Roussillon
- **AUTRES RESPONSABILITÉS ADMINISTRATIVES :**
  - Depuis 2010 : **Membre élu du Conseil Scientifique du LIRMM**
  - Depuis 2007 : **Correspondant de l'association SPECIF** (Société des Personnels Enseignants et Chercheurs en Informatique de France) pour le LIRMM
  - Depuis 2006 : Membre de la Commission bibliothèque du LIRMM – Exposé invité dans le cadre de la conférence FRE-DOC'08 rassemblant les documentalistes du CNRS.

## ACTIVITÉS DE RECHERCHE (DEPUIS 2005)

---

- **DOMAINES DE RECHERCHE** : fouille de textes, traitement automatique du langage naturel, recherche d'information.
- **THÈME PRINCIPAL** : **les descripteurs textuels en Fouille de Textes.**

### – Leur extraction :

*L'extraction de connaissances à partir de textes de spécialité exige l'application d'un processus complet de fouille de textes. Une des étapes de ce processus complexe consiste à extraire des descripteurs linguistiques pertinents dans les textes. Cette extraction dépend des tâches pointues de fouille de textes que nous menons actuellement.*

#### Sous-thèmes :

- \* Extraction de la terminologie à partir de logs / Extraction d'information dans les logs.  
H. Saneifar (LIRMM/Satin-IP - thèse), S. Bonniol (société Satin-IP), P. Poncelet (LIRMM), A. Laurent (LIRMM)
- \* Extraction de descripteurs linguistiques pour le titrage automatique.  
C. Lopez (LIRMM - thèse), V. Prince (LIRMM), M. Faure (Open-S/EvalAcces)
- \* Extraction et désambiguïsation des sigles/expansions.  
V. Prince (LIRMM), I. Mougnot (LIRMM)
- \* Extraction de la terminologie dans un processus global de fouille de textes.  
Y. Kodratoff (LRI), T. Heitz
- \* Extraction de la terminologie du français médiéval.  
C. Serp (Université Montpellier 3 - thèse), E. Cazal (Ingénieur CNAM), M. Teisseire (CEMAGREF), A. Laurent (LIRMM).
- \* Extraction automatique des gloses.  
A. Mela (Université Montpellier 3), A. Steuckardt (LPL, Université de Provence)
- \* Extraction et agrégation de descripteurs dans les entrepôts de données textuelles.  
P. Poncelet (LIRMM), A. Laurent (LIRMM), M. Teisseire (CEMAGREF), S. Bringay (LIRMM), P. Nghiem (Ingénieur CNAM)

### – Leur enrichissement en Recherche d'Information :

*La plupart des approches utilisées en Recherche d'Information (RI) s'appuie sur des méthodes statistiques. L'enrichissement des descripteurs textuels de base se révèle crucial en RI. Ces approches d'enrichissement combinent des méthodes statistiques aux connaissances linguistiques (lexicales, syntaxiques, sémantiques). Notons que des étapes d'action et de rétro-action sont souvent nécessaires entre les processus d'extraction et d'enrichissement.*

#### Sous-thèmes :

- \* Utilisation de connaissances syntaxiques pour la classification.  
N. Béchet (LIRMM - thèse), J. Chauché (LIRMM)
- \* Utilisation de connaissances morpho-syntaxiques pour la classification de textes d'opinion.  
P. Poncelet (LIRMM), M. Plantié (LGI2P-EMA), G. Dray (LGI2P-EMA), F. Troussat (LGI2P-EMA)
- \* Utilisation de connaissances morpho-syntaxiques pour la classification de blogs.  
N. Verdier (société PaperBlog), I. Bayouh (stage d'ingénieur - INSAT Tunisie)
- \* Classification de données textuelles pauvres et hétérogènes.  
V. Poulain d'Andecy (société ITESOFT), H. Hamza (société ITESOFT)
- \* Classification de candidatures d'offres d'emploi.  
R. Kessler (LIA - thèse), J.M. Torres-Moreno (LIA), M. El-Bèze (LIA), N. Béchet (LIRMM - thèse)
- \* Visualisation de connaissances et de données textuelles de biomédecine.  
S. Bringay (LIRMM), M. Teisseire (CEMAGREF), A. Sallaberry (LABRI & PIKKO - thèse)

### – Leur caractéristique lexicale :

*Les travaux menés consistent à étudier la proximité lexicale entre termes (tâche de mise en correspondance de schémas) et d'apporter une analyse lexicale de noms permettant d'identifier une tendance.*

#### Sous-thèmes :

- \* Découverte de correspondances entre schémas  
F. Duchateau (LIRMM - thèse), Z. Bellahsene (LIRMM), F. Pinet (CEMAGREF)
- \* Etude de connotations, détection de tendances  
B. Laurent (Namel Concept), S. Jaillet (CrysaLEAD), A. Laurent (LIRMM)

## RELECTURES D'ARTICLES

---

- **Revue internationale** : IEEE Intelligent Systems, International Journal of Human-Computer Studies, International Journal of Software and Informatics, Journal of Digital Information, International Journal of Web Applications
- **Revue nationale** : TSI, TAL (membre du Comité Editorial d'un numéro spécial), ISI
- **Conférences et Workshops internationaux** : CIRSE'10/09 (membre du CP), DBKDA'10/09 (membre du CP), ICDM'09, DEXA'09, ICADIWT'09, APWeb-WAIM'09, SAC'08, MMD'08, CAISE'07, ISWC'06, LREC'06, ISWC'06, ECML/PKDD'10/04/03/02
- **Conférences et Workshops nationaux** : RJCRI'09 (Vice-Président du CP), RECITAL'09/08 (membre du CP), TALN'10/09/08, Web Social'10 (membre du CP), JFO'09, BDA'09, SIDE'10/09/08 (membre du CP), DEFT'10/09/08/07 (membre du CP), RFIA'08, LFO'07, EDA'07, INFORSID'10/07/06 (membre adjoint du CP en 2006), CORIA'07 (membre du CP), EGC'10/09/08/07/05/03, MA-JECSTIC'07/06/05/04

## • LIVRE

- [1] **M. Roche**. Terminologie et Fouille de Textes (*publication de la thèse*), Éditions universitaires européennes, ISBN : 978-613-1-50420-4, 2010

## • CHAPITRE DE LIVRE

- [2] H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, **M. Roche**. Exterlog : Extraction and Evaluation of Terminology of Log Files. In *Communications in Computer and Information Science (CCIS)*, Springer-Verlag, 2010, to appear
- [3] **M. Roche**. Utilisation des ressources du Web pour la classification conceptuelle. Dans les *Techniques de l'Ingénieur (T.I.)*. H 7 420, 2007

## • REVUES INTERNATIONALES AVEC COMITÉ DE LECTURE

- [4] H. Li, A. Laurent, P. Poncelet, **M. Roche**. Extraction of Unexpected Sentences: A sentiment Classification Assessed Approach. In *Intelligent Data Analysis Journal (IDA)*, Vol 14, No1, p31-46, 2010
- [5] **M. Roche**, V. Prince. A Web-Mining Approach to Disambiguate Biomedical Acronym Expansions. In *Informatica* [acceptance rate  $\sim$  30%], Vol 34, 2010, to appear.
- [6] S. Laroum, N. Béchet, H. Hamza, **M. Roche**. Hybred: An OCR document representation for classification tasks. In *International Journal of Data Engineering and Management (IJDEM)*, Vol 1, 2010, to appear.
- [7] **M. Roche**, Y. Kodratoff. Text and Web Mining Approaches in Order to Build Specialized Ontologies. In *Journal of Digital Information (JoDI)*, Vol 10, No 4, 2009.
- [8] G. Dray, M. Plantié, A. Harb, P. Poncelet, **M. Roche**, F. Troussset. Opinion Mining from Blogs. In *International Journal of Computer Information Systems*, Vol 1, p205-213, 2009.
- [9] **M. Roche**, V. Prince. Managing the Acronym/Expansion Identification Process for Text-Mining Applications. In *International Journal of Software and Informatics, Special issue on Data Mining* [acceptance rate: 28%], Vol 2, No 2, p163-179, 2008

## • REVUES NATIONALES AVEC COMITÉ DE LECTURE

- [10] S. Laroum, N. Béchet, H. Hamza, **M. Roche**. Classification automatique de documents bruités à faible contenu textuel. *Numéro spécial de la revue RNTI, Fouille de Données Complexes*, Vol E-18, 2010.
- [11] A. Harb, M. Plantié, **M. Roche**, G. Dray, F. Troussset, P. Poncelet. Détection d'opinion. Comment déterminer les adjectifs d'opinion d'un domaine donné. *Revue du Documents Numériques*, Vol 11, No 1-2, p37-61, 2008.
- [12] F. Duchateau, Z. Bellahsene, **M. Roche**. Improving Quality and Performance of Schema Matching in Large Scale. *Revue ISI (Ingénierie des Systèmes d'Information)*, Vol 13, No 5, p59-82, 2008.
- [13] C. Serp, A. Laurent, **M. Roche**, M. Teisseire. La quête du Graal et la réalité numérique. *Revue Corpus*, Vol 7, p173-189, 2008.
- [14] J. Azé, **M. Roche**, E. Alphonse, A. Amrani, T. Heitz, A.-D. Mezaour. Bilan du premier défi francophone de Fouille de Textes. *Numéro spécial de la revue RNTI, Défi Fouille de Textes : reconnaissance automatique des auteurs de discours - Campagne DEFT'05 (TALN'05)*, Vol E-10, p15-38, 2007.

## • ACTES DE CONFÉRENCES INTERNATIONALES AVEC COMITÉ DE LECTURE

2010

- [15] H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, **M. Roche**. Passage Retrieval in Log Files: An Approach Based on Query Enrichment. In *Proceedings of IceTAL (International Conference on Natural Language Processing)*, Springer-Verlag, LNCS, Reykjavik, Iceland, 2010, to appear
- [16] S. Bringay, **M. Roche**, M. Teisseire, P. Poncelet, R. Abdel Rassoul, J.M. Verdier, G. Devau. Sequential Patterns: application to micro-arrays data analysis for Alzheimer disease. In *Proceedings of the 13th International Congress on Medical Informatics (MedInfo)*, Cape Town, South Africa, 2010, to appear.
- [17] A. Sallaberry, N. Pecheur, S. Bringay, **M. Roche**, M. Teisseire. SequenceViewer: visualization of genes sequences. In *Proceedings of the 13th International Congress on Medical Informatics (MedInfo)*, Demonstration (long abstract), Cape Town, South Africa, 2010, to appear.
- [18] **M. Roche**. FIN<sup>3</sup>E Approach: Identification of Named Entities from Extracted Terms. In *Proceedings of the 7th International Conference on Cognitive Science (ICCS)*, Poster (long abstract), Beijing, China, 2010, to appear.
- [19] A. Laurent, B. Laurent, D. Brouillet, S. Martin, **M. Roche**. Embedding Emotions within Automatically Generated Brand Names. In *Proceedings of the International Conference on Kansei Engineering and Emotion Research (KEER)*, Paris, 2010.

2009

- [20] R. Kessler, N. Béchet, J.M. Torres-Moreno, **M. Roche**, M. El-Bèze. Job Offer Management: How Improve the Ranking of Candidates. In *Proceedings of ISMIS'09*, Springer-Verlag, LNCS, p431-441, Prague, Czech Republic, 2009.
- [21] H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, **M. Roche**. Mining for Relevant Terms from Log Files. In *Proceedings of KDIR'09 (International Conference on Knowledge Discovery and Information Retrieval)*, p77-84, Funchal - Madeira, Portugal, 2009.
- [22] H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, **M. Roche**. Terminology Extraction from Log Files. In *Proceedings of DEXA'09 (International Conference on Database and Expert Systems Applications)*, Springer-Verlag, LNCS, p769-776, Linz, Austria, 2009.
- [23] N. Béchet, **M. Roche**, J. Chauché. A Hybrid Approach to Validate Induced Syntactic Relations. In *Proceedings of IEEE MAW'09 (International Symposium on Mining and Web)*, p727-732, Bradford, United Kingdom, 2009.
- [24] N. Béchet, **M. Roche**, J. Chauché. Towards the Selection of Induced Syntactic Relations. In *Proceedings of ECIR'09 (European Conference on Information Retrieval)*, Springer-Verlag, LNCS, Poster (Short paper), p786-790, Toulouse, France, 2009.
- [25] P. Salle, S. Bringay, M. Teisseire, F. Chakkoura, **M. Roche**, G. Devauc, C. Lautier, J.M. Verdier. GeneMining: Identification, Visualization and Interpretation of Brain Ageing Signatures. In *Proceedings of Medical Informatics Europe (MIE)*, p767-771, Sarajevo, Bosnia and Herzegovina, 2009.

2008

- [26] N. Béchet, **M. Roche**, J. Chauché. How the ExpLSA approach impacts the document classification tasks. In *Proceedings of IEEE ICDIM'08 (International Conference on Digital Information Management)*, p241-246, University of East London, London, United Kingdom, 2008.
- [27] A. Harb, M. Plantié, G. Dray, **M. Roche**, F. Troussset, P. Poncelet. Web opinion mining: How to extract opinions from blogs? In *Proceedings of IEEE/ACM CSTST'08 (International Conference on Soft Computing as Transdisciplinary Science and Technology)*, p211-217, Cergy, France, 2008.

- [28] I. Bayouhd, N. Béchet, **M. Roche**. Blog classification: Adding Linguistic Knowledge to Improve the K-NN Algorithm. In *Proceedings of IIP'08 (International Conference on Intelligent Information Processing)*, Springer IFIP, p68-77, Beijing, China, 2008.
- [29] M. Plantié, **M. Roche**, G. Dray, P. Poncelet. Is a Voting Approach Accurate for Opinion Mining? In *Proceedings of DaWaK 08 (International Conference on Data Warehousing and Knowledge Discovery)*, Springer-Verlag, LNCS, p413-422, Turin, Italy, 2008.
- [30] H. Li, A. Laurent, **M. Roche**, P. Poncelet. Extraction of Opposite Sentiments in Classified Free Format Text Reviews. In *Proceedings of DEXA 08 (International Conference on Database and Expert Systems Applications)*, Springer-Verlag, LNCS, p710-717, Turin, Italy, 2008.
- [31] N. Béchet, **M. Roche**, J. Chauché. ExpLSA: An Approach Based on Syntactic Knowledge in Order to Improve LSA for a Conceptual Classification Task. In *Proceedings of RCS volume - CICLing'08*, Poster, p213-224, University of Haifa, Israel, 2008.
- [32] N. Béchet, **M. Roche**, J. Chauché. ExpLSA et classification de textes. In *Proceedings of JADT'08 (Journées internationales d'Analyse statistique des Données Textuelles)*, Vol 1, p167-177, Lyon, France, 2008
- [33] C. Serp, E. Cazal, A. Laurent, **M. Roche**. TERVOTIQ : un système de vote pour l'extraction de la terminologie d'un corpus en français médiéval. In *Proceedings of JADT'08 (Journées internationales d'Analyse statistique des Données Textuelles)*, Vol 2, p1069-1080, Lyon, France, 2008
- [34] **M. Roche**, V. Prince. Évaluation et détermination de la pertinence pour des syntagmes candidats à la collocation. In *Proceedings of JADT'08 (Journées internationales d'Analyse statistique des Données Textuelles)*, Poster, Vol 2, p1009-1020, Lyon, France, 2008

2007

- [35] **M. Roche**, V. Prince. AcroDef: A Quality Measure for Discriminating Expansions of Ambiguous Acronyms. In *Proceedings of CONTEXT'07*, Springer-Verlag, LNCS, Poster, p411-424, Roskilde University, Denmark, 2007.
- [36] F. Duchateau, Z. Bellahsene, **M. Roche**. A Context-based Measure for Discovering Approximate Semantic Matching between Schema. In *Proceedings of Research Challenges in Information Science (RCIS) - IEEE*, p9-20, Ouarzazate, Maroc, 2007.
- [37] W. Guedria, Z. Bellahsene, **M. Roche**. A Flexible Approach Based on the user Preferences for Schema Matching. In *Proceedings of Research Challenges in Information Science (RCIS) - IEEE*, p21-26, Ouarzazate, Maroc, 2007.

2005

- [38] J. Azé, **M. Roche**, Y. Kodratoff, M. Sebag. Preference Learning in Terminology Extraction: A ROC-based approach. In *Proceedings of ASMDA'05 (Applied Stochastic Models and Data Analysis)*, p209-219, Brest, France, 2005.

• ACTES DE WORKSHOPS INTERNATIONAUX AVEC COMITÉ DE LECTURE

- [39] R. Kessler, N. Béchet, **M. Roche**, M. El-Bèze, J.M. Torres-Moreno. E-Gen: automatic profiling system for ranking candidates answers in Human Resources. In *Proceedings of QSI workshop (Quantitative Semantic methods for the Internet) - OTM'08*, Springer-Verlag, LNCS, p625-634, Monterrey, Mexico, 2008.
- [40] C. Raissi, J. Brissaud, G. Dray, P. Poncelet, **M. Roche**, M. Teisseire M. Web Analyzing Traffic Challenge: Description and Results. In *Proceedings of ECML/PKDD'2007 Discovery Challenge*, p47-52, Poland, 2007.
- [41] N. Béchet, **M. Roche**, J. Chauché. Improving LSA by expanding the contexts. In *Proceedings of Context-Based Information Retrieval (CIR) workshop - CONTEXT'07*, Poster (Short paper), p105-108, Roskilde University, Denmark, 2007
- [42] F. Duchateau, Z. Bellahsene, M. Roantree, **M. Roche**. An Indexing Structure for Automatic Schema Matching. In *Proceedings of the International Workshop on Self-Managing Database Systems (SMDB) - ICDE'07*, p485-491, Istanbul, Turkey, 2007.
- [43] **M. Roche**, Y. Kodratoff. Pruning Terminology Extracted from a Specialized Corpus for CV Ontology Acquisition. In *Proceedings of onToContent'06 workshop (Ontology content and evaluation in Enterprise) - OTM'06*, Springer Verlag, LNCS, p1107-1116, Montpellier, France, 2006.

• ACTES DE CONFÉRENCES NATIONALES AVEC COMITÉ DE LECTURE

2010

- [44] **M. Roche**. Filtrage des Entités Nommées par des méthodes de Fouille de Textes. In *Actes de la conférence Toth'10 (Terminologie & Ontologie : Théories et Applications)*, Annecy, 2010, à paraître.
- [45] C. Lopez, V. Prince, **M. Roche**. Titrage automatique de documents électroniques par extraction de syntagmes nominaux. In *Actes des Journées francophones d'Ingénierie des Connaissances*, Nîmes, 2010, à paraître.
- [46] D. Breton, **M. Roche**, P. Poncelet, F. Marques. Cartographie de dépêches pour analyser le développement d'épidémies. In *Actes des Journées francophones d'Ingénierie des Connaissances*, Poster (Article court), Nîmes, 2010, à paraître.
- [47] D. Breton, **M. Roche**, P. Poncelet, F. Marques. Analyse de dépêches pour l'épidémiologie. In *Actes des Journées francophones d'Ingénierie des Connaissances*, Démonstration (Résumé étendu), Nîmes, 2010, à paraître.
- [48] H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, **M. Roche**. Recherche de passages pertinents dans les fichiers logs par enrichissement de requêtes. In *Actes de la conférence francophone en Recherche d'Information et Applications (CORIA)*, p239-254, Sousse, Tunisie, 2010.
- [49] S. Bringay, A. Laurent, P. Poncelet, **M. Roche**, M. Teisseire. Bien cube, les données textuelles peuvent s'agrèger ! In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2010, p585-596, Hammamet, Tunisie, 2010. Article nominé parmi les 8 meilleurs articles applicatifs d'EGC'10.
- [50] A. Sallaberry, N. Pecheur, S. Bringay, **M. Roche**, M. Teisseire. SequencesViewer : Visualisation de séquences ordonnées de gènes ou comment rendre accessible des motifs séquentiels trop nombreux ? In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2010, Article court, p387-392, Hammamet, Tunisie, 2010.

2009

- [51] H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, **M. Roche**. Processus d'extraction et de validation de la terminologie issue de logs. In *Actes des Journées Francophones sur les Ontologies (JFO'2009)*, Poitiers, 2009.
- [52] N. Béchet, **M. Roche**, J. Chauché. Corpus et Web : deux alliés pour la construction de l'enrichissement automatique de classes conceptuelles. In *Actes de la conférence Toth'09 (Terminologie & Ontologie : Théories et Applications)*, Annecy, 2009.
- [53] R. Kessler, N. Béchet, J.M. Torres-Moreno, **M. Roche**, M. El-Bèze. Profilage de candidatures assisté par Relevance Feedback. In *Actes de la conférence TALN'09*, Poster, Senlis, 2009.
- [54] N. Béchet, **M. Roche**, J. Chauché. Comment valider automatiquement des relations syntaxiques induites. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2009, p169-180, Strasbourg, 2009. Article nominé parmi les 9 meilleurs articles académiques d'EGC'09

- [55] O. Djanga, H. Hamzioui, M. Hatchi, I. Mougenot, **M. Roche**. Regroupement des Définitions de Sigles Biomédicaux. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2009, Démonstration (Résumé étendu), p487, Strasbourg, 2009.

2008

- [56] **M. Roche**, V. Prince. Processus global d'acquisition et de gestion des sigles. In *Actes de la conférence INFORSID'08*, Fontainebleau, 2008
- [57] H. Li, A. Laurent, **M. Roche**, P. Poncelet. Recherche de sentiments opposés par une approche floue à partir de textes libres. In *Actes de la conférence LFA'08 (Logique Floue et ses Applications)*, Lens, 2008
- [58] N. Béchet, **M. Roche**, J. Chauché. ExpLSA : utilisation d'informations syntaxico-sémantiques associées à LSA pour améliorer les méthodes de classification conceptuelle. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2008, p589-600, Sophia-Antipolis, 2008.
- [59] M. Plantié, **M. Roche**, G. Dray. Un système de vote pour la classification de textes d'opinion. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2008, Article court, p583-588, Sophia-Antipolis, 2008.
- [60] V. Matviico, N. Muret, **M. Roche**. Processus d'acquisition d'un dictionnaire de sigles et de leurs définitions à partir d'un corpus. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2008, Démonstration (Résumé étendu), p231-232, Sophia-Antipolis, 2008.
- [61] J. Rabatel, Y. Lin, Y. Pitarch, H. Saneif, C. Serp, **M. Roche**, A. Laurent. Visualisation des motifs séquentiels extraits à partir d'un corpus en Ancien Français. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2008, Démonstration (Résumé étendu), p237-238, Sophia-Antipolis, France, 2008.

2007

- [62] F. Duchateau, Z. Bellahsene, **M. Roche**. BMatch: a Semantically Context-based Tool Enhanced by an Indexing Structure to Accelerate Schema Matching. In *Actes des Journées Bases de Données Avancées (BDA)*, Marseille, 2007.
- [63] **M. Roche**, V. Prince. DefAcro : mesure de qualité pour le choix de la définition des acronymes ambigus. In *Actes de la quatrième conférence francophone en Recherche d'Information et Applications (CORIA)*, p255-270, Saint-Etienne, 2007.

2006

- [64] A. Mela, **M. Roche**. Des gloses de mot aux types de textes : un bilan différencié. In *Actes du colloque "Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation", Texto! [en ligne] Vol XI, n°2*, 2006.
- [65] **M. Roche**, Y. Kodratoff. Choix du taux d'élagage pour l'extraction de la terminologie. Une approche fondée sur les courbes ROC. *Revue RNTI (Revue des Nouvelles Technologies de l'Information)* numéro spécial conférence EGC'06, p205-216, Lille, 2006.
- [66] **M. Roche**. Acquisition de la terminologie et définition des tâches à effectuer, deux principes indissociables. Dans *Actes des Journées de Rochebrune (Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels)*, p151-161, Megève, 2006.

2005

- [67] A. Amrani, **M. Roche**, Y. Kodratoff, O. Matte-Tailliez. Inductive Improvement of Part-of-Speech Tagging and its Effect on a Terminology of Molecular Biology. In *Proceedings of the 18th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2005, Lecture Notes in Artificial Intelligence, Springer-Verlag*, p366-376, Victoria, British-Columbia, Canada, 2005
- [68] T. Heitz, **M. Roche**, Y. Kodratoff. Extraction de termes centrée autour de l'expert. *Revue RNTI (Revue des Nouvelles Technologies de l'Information)* numéro spécial conférence EGC'05, Démonstration (Article court), p685-690, Paris, 2005.