

CURRICULUM VITÆ

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

MAÎTRE DE CONFÉRENCES
À L'UNIVERSITÉ DE MONTPELLIER 2

Projet TAL, LIRMM
UMR 5506, CNRS, Université Montpellier 2
Courriel : mroche@lirmm.fr
Web : <http://www.lirmm.fr/~mroche/>



CURSUS

• EXPÉRIENCE PROFESSIONNELLE

- Depuis septembre 2005 : **Maître de Conférences** à l'Univ. Montpellier 2.
- 2004-2005 : **Demi poste d'ATER** (Attaché Temporaire de l'Enseignement et de la Recherche) en Informatique à l'Univ. Paris XI.
- 2001-2004 : **Allocataire de recherche, moniteur en informatique** à l'Univ. Paris XI (Orsay) dans l'équipe Inférence et Apprentissage (IA) du Laboratoire de Recherche en Informatique (LRI).

• FORMATION

- 13 décembre 2004 : **Doctorat en Informatique** à l'Univ. 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'Univ. Paris XI (91), mention Bien. Stage de DEA dans l'équipe IA Encadrement : C. Nédellec et D. Faure.

ANIMATION RECHERCHE & ENSEIGNEMENT (DEPUIS 2005)

• RESPONSABLE DE PROJETS DE RECHERCHE ET TRANSFERT DE TECHNOLOGIE :

- **Porteur du projet ProSigles** – 2007-2008.
Projet sélectionné par le Conseil Scientifique de l'Université Montpellier 2.
Le but de ce projet consiste à extraire et gérer 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.
Ce projet a pour but d'extraire des connaissances à partir de corpus écrits en français médiéval.
- **Membre de l'ANR Forum** (2005-2008) portée par Z. Bellahsene (LIRMM).
L'objectif est d'étendre les techniques d'intégration pour permettre le partage de données dans un environnement à grande échelle.
- **Accompagnement de divers Projets Incubés du Languedoc-Roussillon Incubation**
 - * EvalAcces (porteur M. Faure) – incubé en 2008 (part du LIRMM : 12Keuros)
 - * Namae Concept (porteuse B. Laurent) – incubé en 2007 (part du LIRMM : 4Keuros)

• RESPONSABLE & MEMBRE DE COMITÉS D'ORGANISATION :

- **Co-responsable avec F. Saïs (LRI - Orsay) et N. Béchet (LIRMM - Montpellier) de l'atelier EvalECD'09** dans le cadre de la conférence EGC'09
L'objectif de cet atelier est de confronter les 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
Le but de cet atelier est de faire émerger 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**
Cette conférence a pour but de présenter les 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**
Le but de ce challenge international est de classifier le trafic HTTP en plusieurs classes permettant de distinguer les différents types d'attaques et le trafic normal.
- **Création 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)

- ÉDITION DE REVUES ET D'OUVRAGES :

- "Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration", IGI Publishing, Medical Information Science Reference, 460 pages, March 2009 [V. Prince, M. Roche Ed.]
- Numéro spécial de la revue RNTI (Revue des Nouvelles Technologies de l'Information). "Fouille de Données d'Opinions, à paraître en 2009 [M. Roche, P. Poncelet Ed.]
- Numéro spécial de la revue RNTI (Revue des Nouvelles Technologies de l'Information). "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 BDI/Région de Nicolas Béchet co-encadrée avec J. Chauché (10%) - depuis octobre 2006. Titre de la thèse : *Catégorisation automatique de documents à l'aide de techniques hybrides (syntaxiques et apprentissage)*. Soutenance prévue : octobre/novembre 2009
- * Thèse CIFFRE de Hassan Saneifar co-encadrée avec P. Poncelet (30%) – Thèse en collaboration avec la société Satin-IP – depuis octobre 2008. Titre de la thèse : *Extraction d'Information à partir de Logs*.
- * Thèse UM2/Région co-encadrée avec V. Prince (30%) - à partir d'octobre 2009 (*sélection du candidat en cours*). Titre de la thèse : *Titrage Automatique de Pages Web*.

- **Stages Recherche (Master 2) :**

- * Stage de Walid Belhaoues – 2008-2009 – co-encadré avec V. Prince (LIRMM), S. Léon (LIRMM), M. Faure (Open-S/EvalAcces)
- * Stage de Miguel Macias – 2007-2008 – co-encadré avec V. Prince (LIRMM), A. Labadié (LIRMM)
- * Stage de Ali Harb – 2007-2008 – co-encadré avec P. Poncelet (LGI2P) , M. Plantié (LGI2P)
- * Stage de Samy Laroum – 2007-2008 – co-encadré avec N. Béchet (LIRMM)
- * Stage de Nicolas Boffo – 2007-2008 – co-encadré avec A. Laurent (LIRMM), J. Bres (PRAXILING)
- * Stage de Larbi Mesbahi – 2005-2006 – co-encadré avec M. Teisseire (LIRMM), J. Chauché (LIRMM)
- * Stage de Tang Yi Fei – 2005-2006 – co-encadré avec Z. Bellahsene (LIRMM)

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

- * Stage de fin d'étude d'ingénieur d'Ines Bayouh – INSAT Tunisie / Société Paperblog - 2007-2008 – co-encadré avec N. Béchet (LIRMM)
- * Stage d'ingénieur CNAM d'Emmanuel Cazal – 2006-2007 – co-encadré avec A. Laurent (LIRMM), M. Teisseire (LIRMM)

- **Stage Post Doctoral :**

- * Stage de F. Chakkoura - 2008-2009 (3 mois) – co-encadré avec S. Bringay (LIRMM), M. Teisseire (LIRMM)

- **TER stages et projets tutorés (Masters 1 et 2) :**

- * Stages de quatre étudiants en Master 1 Informatique – 2005-2009 – 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/quatre étudiants en Master 2 Informatique – 2007-2008 – co-encadré avec A. Laurent (LIRMM), B. Laurent (Namae Concept), S. Jaillet (CrysaLEAD)
- * Projets tutorés en Master 1 Informatique (Intégration de Compétences) – 2006-2009 – co-encadrés avec A. Laurent (LIRMM), B. Laurent (Namae Concept), I. Mougenot (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) :

- * Remy Kessler (Univ. Avignon) – soutenue le 10 juillet 2009
- * Dong (Haoyuan) Li (Univ. Montpellier 2) – soutenance prévue en septembre 2009

- **Rapporteur de stages de M2 Informatique** (hors LIRMM/Univ. Montpellier 2) :

- * rapporteur extérieur de stages du Master ECD (Univ. Lyon 2) - depuis 2007

- RESPONSABILITÉS PÉDAGOGIQUES D'ENSEIGNEMENT (UNIVERSITÉ MONTPELLIER 2) :

- Co-responsable du module "Stages d'analyse ou Projets Tutorés" du Master Informatique (spécialité IC) – depuis 2008

- Co-responsable du module "Extraction de Connaissances dans les Données" du Master Informatique (spécialité IFPRU) – depuis 2007
 - Co-responsable du module "Fouille de Données Avancée" du Master Informatique (spécialité IFPRU) – depuis 2007
 - Responsable du module "Fouille de Données" du Master Informatique (spécialité IC) – depuis 2007
 - Co-responsable du module "Administration des Bases de Données" du Master Informatique (spécialité IC) – depuis 2007
 - Co-responsable du module de "Systèmes d'Information et Bases de Données" en licence (L3) – depuis 2007
 - Co-responsable du module "Systèmes d'Information et Bases de Données" du Master Informatique (spécialité IC) – depuis 2006
 - Co-responsable du module de "Systèmes d'Information et Bases de Données" en Licence (L2) – 2006-2007
 - Responsable du module "Fouille de Données" du Master Professionnel Informatique – 2005-2007
- **RESPONSABILITÉS ADMINISTRATIVES D'ENSEIGNEMENT :**
 - **Responsable du Master Informatique spécialité "Intégration de Compétences" (IC)** – depuis 2008.

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
 - **Responsable du Parcours CCI (Compétences Complémentaires en Informatique)** du Master IC – 2006-2008
 - Responsable des stages du Master IC – depuis 2006
 - Responsable du Certificat Professionnel CP09 : "Webmestre : Conception de Sites Administration et Serveurs Web" pour le centre CNAM du Languedoc Roussillon (2008-2009)
 - **AUTRES RESPONSABILITÉS ADMINISTRATIVES :**
 - **Correspondant de l'association SPECIF** (Société des Personnels Enseignants et Chercheurs en Informatique de France) pour le LIRMM – depuis 2007
 - Membre de la Commission bibliothèque du LIRMM – depuis 2007 – Exposé invité dans le cadre de la conférence FREDOC'08 rassemblant les documentalistes du CNRS.
 - **Membre de la Commission de Spécialistes** de l'Université d'Avignon (CSE 7, 27, 61) – 2007-2008

ACTIVITÉS DE RECHERCHE (DEPUIS 2005)

- **DOMAINES DE RECHERCHE :** fouille de textes, traitement automatique du langage naturel, recherche d'information.
- **DESCRIPTION DES DIFFÉRENTS THÈMES DE RECHERCHE.** L'objectif principal de ces différents axes de Recherche consiste à **déterminer les descripteurs pertinents pour différentes tâches de fouille de textes.**
 - **Enrichissement des Descripteurs Textuels en Recherche d'Information :**

Actuellement, la plupart des approches utilisées pour la catégorisation automatique de textes s'appuient sur des méthodes statistiques. Cependant, ces méthodes ne tiennent pas toujours compte de la spécificité du matériau (type de langage utilisé, domaine des textes, etc). L'ajout de connaissances linguistiques pour la Recherche d'Information est un axe de recherche crucial.

Collaborations :

 - * N. Béchet (LIRMM - thèse), J. Chauché (LIRMM) : ajout de connaissances syntaxiques aux méthodes de classification
 - * M. Teisseire (LIRMM) : ajout de connaissances linguistiques aux motifs séquentiels
 - * P. Poncelet (LIRMM), M. Plantié (LGI2P-EMA), G. Dray (LGI2P-EMA), F. Troussel (LGI2P-EMA) : utilisation de connaissances linguistiques pour la classification de textes d'opinion
 - * N. Verdier (société PaperBlog), I. Bayouhd (stage d'ingénieur - INSAT Tunisie) : classification de blogs
 - * V. Poulain d'Andecy (société ITESOFT), Hatem Hamza (société ITESOFT) : classification de données textuelles pauvres et hétérogènes
 - * R. Kessler (LIA - thèse), J.M. Torres-Moreno (LIA), M. El-Bèze (LIA), N. Béchet (LIRMM - thèse) : classification de candidatures d'offres d'emploi
 - * S. Bringay (LIRMM), M. Teisseire (CEMAGREF) : recherche de liens entre les connaissances extraites (motifs séquentiels) via un processus de fouille de données et les données textuelles (articles scientifiques) dans les domaines Santé et Environnement
 - **Extraction Automatique de Descripteurs Linguistiques**

L'extraction d'information à partir de textes spécialisés exige l'application d'un processus complet de fouille de textes. Une des étapes de ce processus consiste à extraire les termes dans les textes qui sont définis comme des groupes de mots représentant des traces linguistiques de concepts.

Collaborations :

 - * V. Prince (LIRMM) : étude des sigles / extraction de la terminologie orientée tâches
 - * H. Saneifar (LIRMM/Satin-IP - thèse), S. Bonniol (société Satin-IP), P. Poncelet (LIRMM), A. Laurent (LIRMM) : extraction de la terminologie à partir de logs

- * I. Mougenot (LIRMM) : étude des sigles en Sciences du Vivant
- * C. Serp (Université Montpellier 3 - thèse), E. Cazal (Ingénieur CNAM), M. Teisseire (LIRMM), A. Laurent (LIRMM) : étude de la terminologie du français médiéval
- * Y. Kodratoff (LRI) : étude de la terminologie dans un processus global de fouille de textes
- * A. Mela (Université Montpellier 3) : repérage automatique des gloses

– **Étude Lexicale des Descripteurs Linguistiques :**

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

Collaborations :

- * F. Duchateau (LIRMM - thèse), Z. Bellahsene (LIRMM), F. Pinet (CEMAGREF) : découverte de correspondances entre schémas – ANR Forum
- * B. Laurent (Namea Concept), S. Jaillet (CrysaLEAD), A. Laurent (LIRMM) : étude de connotations, détection de tendances – Projet Incubé Namea Concept

– **Titration Automatique :**

Afin de respecter un des critères de la norme W3C et d'apporter de manière automatique une connaissance synthétique des documents, des méthodes de titrage de pages web à partir de descripteurs linguistiques sont proposées.

Collaborations : V. Prince (LIRMM), M. Faure (Open-S/EvalAcces)

RELECTURES D'ARTICLES

- **Revue internationale :** Int. J. of Human-Computer Studies, International Journal of Software and Informatics, Journal of Digital Information
- **Conférences et Workshops internationaux :** CIRSE'09 (membre du CP), DBKDA'09 (membre du CP), DEXA'09, ICADIWT'09, APWeb-WAIM'09, SAC'08, MMD'08, CAISE'07, ISWC'06, LREC'06, ISWC'06, ECML/PKDD'04/03/02, ISMIS'02
- **Revue nationale :** TSI, TAL, ISI
- **Conférences et Workshops nationaux :** RJCRI'09 (Vice-Président du CP), RECITAL'09/08 (membre du CP), TALN'09/08, BDA'09, SIDE'09/08 (membre du CP), DEFT'09/08/07 (membre du CP), RFIA'08, LFO'07, EDA'07, INFORSID'07/06 (membre adjoint du CP en 2006), CORIA'07 (membre du CP), EGC'09/08/07/05/03, MAJECSTIC'07/06/05/04, CAp'04, JFT'03

PUBLICATIONS SCIENTIFIQUES DEPUIS 2005 (HORS ÉDITIONS)

• CHAPITRE DE LIVRE

- [1] **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

- [2] H. Li, A. Laurent, P. Poncelet, **M. Roche.** Extraction of Unexpected Sentences: A sentiment Classification Assessed Approach. In *Intelligent Data Analysis Journal (IDA)*, 14(1), 2010, to appear.
- [3] 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)*, 2009, to appear.
- [4] **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%], num 2, vol 2, p163-179, 2008

• REVUES NATIONALES AVEC COMITÉ DE LECTURE

- [5] 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*, 2009, à paraître.
- [6] 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*, numéro 1-2, volume 11, p.37-61, 2008.
- [7] F. Duchateau, Z. Bellahsene, **M. Roche.** Improving Quality and Performance of Schema Matching in Large Scale. *Revue ISI (Ingénierie des Systèmes d'Information)*, numéro 5, volume 13, p.59-82, 2008.
- [8] C. Serp, A. Laurent, **M. Roche**, M. Teisseire. La quête du Graal et la réalité numérique. *Revue Corpus*, numéro 7, 2008.
- [9] 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)*, p15-38, volume E-10, 2007.

• ACTES DE CONFÉRENCES INTERNATIONALES AVEC COMITÉ DE LECTURE

- [10] 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, to appear, 2009.
- [11] 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)*, to appear, 2009.
- [12] 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, to appear, 2009.
- [13] 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)*, Bradford, United Kingdom, May 2009.
- [14] 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, Short paper (poster), p786-790, Toulouse, France, April 2009.
- [15] 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)*, Sarajevo, Bosnia and Herzegovina, September 2009.

- [16] 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, November 2008.
- [17] A. Harb, M. Plantié, G. Dray, **M. Roche**, F. Troussel, 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, October 2008.
- [18] I. Bayoudh, 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, October 2008.
- [19] 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, September 2008.
- [20] H. Li, Laurent A., Roche M., Poncelet P. 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, Turin, Italy, p710-717, September 2008.
- [21] 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 (poster proceedings) - CICLing'08*, pages 213–224, University of Haifa, Israel, 2008.
- [22] N. Béchet, **M. Roche**, J. Chauché. ExpLSA et classification de textes. In *Proceedings of JADT'08 (Journées internationales d'Analyse statistique des Données Textuelles)*, pages 167–177, Volume 1, Lyon, France, 2008
- [23] 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)*, pages 1069–1080, Volume 2, Lyon, France, 2008
- [24] **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, pages 1009–1020, Volume 2, Lyon, France, 2008
- [25] **M. Roche**, V. Prince. AcroDef: A Quality Measure for Discriminating Expansions of Ambiguous Acronyms. In *Proceedings of CONTEXT'07*, Springer-Verlag, LNCS, pages 411–424, Roskilde University, Denmark, 2007.
- [26] 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*, pages 9–20, Ouarzazate, Maroc, 2007.
- [27] 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*, pages 21–26 (short paper), Ouarzazate, Maroc, 2007.
- [28] 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)*, pages 209–219, Brest, France, 2005.
- ACTES DE WORKSHOPS INTERNATIONAUX AVEC COMITÉ DE LECTURE
- [29] 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, Monterrey, Mexico, November 2008.
- [30] 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*, pages 47–52, Poland, 2007.
- [31] N. Béchet, **M. Roche**, J. Chauché. Improving LSA by expanding the contexts. In *Proceedings of Context-Based Information Retrieval (CIR) workshop - CONTEXT'07* (short paper), pages 105-108, Roskilde University, Denmark, 2007
- [32] F. Duchateau, Z. Bellahsene, M. Roantree, **M. Roche**. An Indexing Structure for Automatic Schema Matching. In *Proceedings of the 2nd International Workshop on Self-Managing Database Systems (SMDDB) - ICDE'07*, Istanbul, Turkey, 2007.
- [33] **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, Montpellier, France, pages 1107–1116, 2006.
- ACTES DE CONFÉRENCES NATIONALES AVEC COMITÉ DE LECTURE
- [34] 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.
- [35] 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.
- [36] 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, 2009. Article nominé parmi les 9 meilleurs articles académiques d'EGC'09
- [37] O. Djang, H. Hamzioui, M. Hachi, I. Mougenot, **M. Roche**. Regroupement des Définitions de Sigles Biomédicaux. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information)*, numéro spécial EGC'2009, session démonstrations (résumé), 2009.
- [38] **M. Roche**, V. Prince. Processus global d'acquisition et de gestion des sigles. In *Actes de la conférence INFORSID'08*, Fontainebleau, 2008
- [39] 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
- [40] 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, pages 589-600, Sophia-Antipolis, France, 2008.
- [41] 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, pages 583-588, Sophia-Antipolis, France, 2008.
- [42] 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, session démonstrations, pages 231-232, Sophia-Antipolis, France, 2008.
- [43] 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, session démonstrations, pages 237-238, Sophia-Antipolis, France, 2008.
- [44] 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, France, 2007.
- [45] **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)*, pages 255–270, Saint-Etienne, France, 2007.
- [46] 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.
- [47] **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, pages 205-216, Lille, France, 2006.
- [48] **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)*, pages 151-161, Megève, France, 2006.
- [49] 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, pages 366-376, Victoria, British-Columbia, Canada, 2005
- [50] 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 (session démonstrations), E3:685-690, Paris, France, 2005.