

# Atelier "Fouille des Données d'Opinions" (FODOP) INFORSID'08

Mathieu Roche et Pascal Poncelet

LIRMM, Université Montpellier 2  
LGI2P, École des Mines d'Alès

*27 mai 2008*

Un domaine en plein essor : la fouille de données d'opinion (thématique de DEFT'07) :

- Évaluation d'un produit.
- Image que les clients peuvent se faire d'une entreprise.
- Détection de rumeurs (buzz) sur le web.
- etc.

### Verrous scientifiques... pluridisciplinaires

- Construction de **dictionnaire d'opinion spécifiques** : l'adjectif "commercial" est-il positif ?
- Outre la polarité, attribuer des **scores d'intensité** (utilisation des descripteurs linguistiques tels que les adverbes).
- Détection de **sentiments contradictoires**.
- Interpréter des **constructions syntaxiques complexes** ("je ne peux pas dire que je n'aime ce film !").

## Sept papiers présentés à FODOP'08 :

- Un papier invité
- Trois papiers longs
- Trois papiers courts

**Projet** : Édition d'une revue sur le thème de FODOP'08

### Deux thématiques :

- Classification des textes d'opinion
- Opinion et Web

**Remarque** : frontière entre ces deux thèmes pas nécessairement bien marquée.

## Principales étapes pour la détection d'opinions (bilan des articles de FODOP'08) :

- **Représentation des données :**  
Choix des descripteurs pertinents par des méthodes **statistiques**, **linguistiques** (par exemple, n-grammes de mots, de caractères) et/ou **mixtes**,
- **Identification des classes d'opinion :**  
Méthodes de classification "**classiques**" (par exemple, SVM, Naive Bayes) et/ou **hybrides** (adaptées aux données d'opinions par des systèmes de score),
- **Évaluation :**  
Utilisation des ressources des **campagnes d'évaluation** (en français, DEFT'07), données issues de **forums**, etc.