

## Fiche de Lecture

### Numéro spécial de RNTI « Fouille de Données d'Opinions »

Titre de l'article : Extracting Opinions and Facts for Business Intelligence

Noms des auteurs : H. Saggion and A. Funk

Affiliation : U. Sheffield

---

### Evaluation générale :

(cocher la case correspondante)

Originalité, nouveauté, résultats significatifs	Excellent	Bon	Passable	Insuffisant X
Qualité technique	Excellent	Bon	Passable	Insuffisant X
Présentation de l'article, style et organisation	Excellent	Bon	Passable X	Insuffisant
Impression générale	Excellent	Bon	Passable	Insuffisant X

### Recommandation :

Accepter tel quel	
Accepter avec des modifications mineures	
Accepter avec des modifications majeures	
Rejeter	X

### Compétences que vous estimez avoir sur le sujet de l'article :

Spécialiste	
Bonne connaissance	X
Connaissance des concepts de base	
Domaine mal connu	

## **Remarques générales :**

### **Résumé de l'article :**

This article seeks to retrieve the semantic orientation of a text using machine learning algorithms. Two corpora are used: a consumer forum where each comment comes along a thumbs-up or -down rating, and another consumer forum where each comment comes with a one to five stars. The thumbs and the stars reflects the opinion expressed in the text. Several linguistic features are computed to feed the machine learning algorithm: basic features and semantic information obtained for the SentiwordNet. The evaluation is done on the same data than the parameter estimation. The best results are of 80 % of F-measure: 85.8% for positive and 51.7 for negative for the binary classification; for the star 79.9 for 1-star and 74.9 for 5 stars, the scores for the intermediate stars are low.

### **Donner les points forts de l'article :**

This article falls within the scope of the special issue of the RNTI journal.

### **Donner les points faibles de l'article :**

This article is far from the objectives it claims: p.2 “identify statements which reflect the concepts of reputation, reliability and quality of business entities and track them over time in order to create an accurate picture of a business entity”. Only the semantic orientation is computed. There is no link to the business entities, no idea how to link company information extraction module and the opinion mining module. Not a word about “the over time” and “accuracy” assumptions.

The work presented is not original; many similar works have been done using the same technics and the same kind of data (Pang and al. 2004; Airolti et al. 2006; Dave et al. 2003; Gamon, 2004; etc.).

The state of art is incomplete (see above).

The work lacks of scientific content: the choice to use machine learning tools is not motivated; the features are basic, some are redundant; there is no discussion about the specificity of the business corpus and opinion expression in this particular domain; what are the constraints of the use of GATE platform, etc.

The paper is unclear: the screen dumps could not be read; the Gate processing resources presented are not used in the experiments; the organization of section 5 is on contradiction with what is announced in the introduction (4 types of features); the summary does not reflect the content of the paper, etc.

There is a methodological flaw as the evaluation is done on the same data than the parameter estimation.