

# Chapter 64

## Sentiment Analysis of Arabic Documents: Main Challenges and Recent Advances

**Hichem Rahab**

 <https://orcid.org/0000-0002-4411-0901>


*ICISI Laboratory, University of Khenchela, Algeria*

**Mahieddine Djoudi**

 <https://orcid.org/0000-0002-2998-5574>

*TechNE Laboratory, University of Poitiers, France*

**Abdelhafid Zitouni**

 <https://orcid.org/0000-0003-2498-4967>

*LIRE Laboratory, University of Constantine 2, Algeria*

### ABSTRACT

*Today, it is usual that a consumer seeks for others' feelings about their purchasing experience on the web before a simple decision of buying a product or a service. Sentiment analysis intends to help people in taking profit from the available opinionated texts on the web for their decision making, and business is one of its challenging areas. Considerable work of sentiment analysis has been achieved in English and other Indo-European languages. Despite the important number of Arabic speakers and internet users, studies in Arabic sentiment analysis are still insufficient. The current chapter vocation is to give the main challenges of Arabic sentiment together with their recent proposed solutions in the literature. The chapter flowchart is presented in a novel manner that obtains the main challenges from presented literature works. Then it gives the proposed solutions for each challenge. The chapter reaches the finding that the future tendency will be toward rule-based techniques and deep learning, allowing for more dealings with Arabic language inherent characteristics.*

DOI: 10.4018/978-1-6684-6303-1.ch064

## **INTRODUCTION**

The evolution of Internet use in today's world is coupled with an important advancement in offering services for users. A tremendous amount of information and data is generated, and more needs emerge to take benefit from it (Liu, 2012). The importance of taking into accounts other opinions and advises in decision-making process (sales, voting, etc.) is a result of the neutrality of this information and its independence from any conflict of interest (Liu, 2015). When someone would sell a new product, ask for a service like a hotel booking or go to a new restaurant, or even take the decision in elections, he is no more limited to advice of family members and near friends. On Internet there are several web sites, discussion forums and social networks allowing their visitors to open debates and giving their comments on subjects, products and services of their interest (Guellil et al., 2019).

Sentiment analysis seeks to discover positive and negative sentiments about objects (ex. Cellular phones) and their attributes (image quality, weight, etc.) through natural language processing NLP, text mining and data-mining techniques (Aggarwal, 2018). Sentiment analysis aims to classify discovered people opinions into well-defined categories to facilitate hidden phenomenon understanding. Sentiment analysis can be seen as an automatic summarization of subjective documents, which allow positive or negative polarity extraction from textual documents (Pang & Lee, 2004). Opinion mining use is not limited to product reviews; it can reach user attitudes, political attitudes etc. (Aggarwal, 2018).

The application of sentiment analysis techniques covers a widespread of domains, such as business, politics, security and healthcare, to cite a few. Using sentiment analysis in healthcare domain can profit from available opinionated data in social media and web forums to help in the improvement of healthcare systems by controlling epidemics and guarantee a better care for patients (Ramírez-Tinoco et al., 2019). In the security domain, opinion mining can be used to control exchanged discussions in e-mails (Danowski, 2012), in social networks or even in phone conversations (Iskra et al., 2004). The available data at low cost in social media can be very beneficial to prevent possible perturbations in different events. This information can provide useful information for authorities and organizations, allowing them to take suitable decisions by understanding their people's mood (Subramaniaswamy et al., 2017).

An important number of works in data mining and sentiment analysis is achieved in European languages, especially in English. Resources in these languages are available and enough in term of quantity and quality. However, in low resourced languages such as Arabic, the number of dedicated resources is very limited. Arabic is a Semitic language with more than 400 million speakers in 22 countries, and it is the fourth most used language in the Internet by 226 million users (*Internet World Stats*, n.d.). Arabic letters are used to write other languages, such as Persian and Urdu. The Arabic language is also important as the language of the Holy Quran the book of 1.5 billion muslims around the world.

The aim of the present chapter is to provide sentiment analysis community, newcomers especially, with the important problematic questions and their last proposed solutions in the literature. The research questions that this chapter intends to answer are: what are the recent problems that sentiment analysis has to solve? What are the important methods used in sentiment analysis for the resolution of each problem? Are these methods sufficient to overcome the new challenges in the domain? And what will be the future tendency of the research in opinion mining and sentiment analysis? Unlike traditional surveys, listing sentiment analysis works according to adopted approaches and used methods and tools, the authors in this chapter extract, from literary works, the main challenges and their corresponding solutions. By the adopted methodology, the authors reach the new tendency in the domain, in order to respond to new challenges.

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/sentiment-analysis-of-arabic-documents/308542](http://www.igi-global.com/chapter/sentiment-analysis-of-arabic-documents/308542)

## Related Content

---

### Basic Principles of Data Mining

Karl-Ernst Erich Biebler (2009). *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 266-289).

[www.irma-international.org/chapter/basic-principles-data-mining/29155](http://www.irma-international.org/chapter/basic-principles-data-mining/29155)

### Computing Join Aggregates Over Private Tables

Rong She, Ke Want, Ada Waichee Fuand Xu Yabo (2008). *International Journal of Data Warehousing and Mining* (pp. 22-41).

[www.irma-international.org/article/computing-join-aggregates-over-private/1816](http://www.irma-international.org/article/computing-join-aggregates-over-private/1816)

### Approaches to Large-Scale User Opinion Summarization for the Web

William Darling (2014). *Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding* (pp. 163-184).

[www.irma-international.org/chapter/approaches-to-large-scale-user-opinion-summarization-for-the-web/96744](http://www.irma-international.org/chapter/approaches-to-large-scale-user-opinion-summarization-for-the-web/96744)

### Dynamic Discovery of Fuzzy Functional Dependencies Using Partitions

Shyue-Liang Wang, Ju-Wen Shenand Tuzng-Pei Hong (2010). *Intelligent Soft Computation and Evolving Data Mining: Integrating Advanced Technologies* (pp. 44-60).

[www.irma-international.org/chapter/dynamic-discovery-fuzzy-functional-dependencies/42355](http://www.irma-international.org/chapter/dynamic-discovery-fuzzy-functional-dependencies/42355)

### Optimization of Mean and Standard Deviation of Multiple Responses Using Patient Rule Induction Method

Jin-Kyung Yangand Dong-Hee Lee (2018). *International Journal of Data Warehousing and Mining* (pp. 60-74).

[www.irma-international.org/article/optimization-of-mean-and-standard-deviation-of-multiple-responses-using-patient-rule-induction-method/198974](http://www.irma-international.org/article/optimization-of-mean-and-standard-deviation-of-multiple-responses-using-patient-rule-induction-method/198974)