# *A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applications and Visualizations*

# A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applications, and Visualizations

GILBERT BADARO, American University of Beirut, Lebanon
RAMY BALY, Massachusetts Institute of Technology, USA
HAZEM HAJJ and WASSIM EL-HAJJ, American University of Beirut, Lebanon
KHALED BASHIR SHABAN, Qatar University, Qatar
NIZAR HABASH, New York University of Abu Dhabi, United Arab Emirates
AHMAD AL-SALLAB, Cairo University, Egypt
ALI HAMDI, Qatar University, Qatar

Opinion-mining or sentiment analysis continues to gain interest in industry and academics. While there has been significant progress in developing models for sentiment analysis, the field remains an active area of research for many languages across the world, and in particular for the Arabic language, which is the fifth most-spoken language and has become the fourth most-used language on the Internet. With the flurry of research activity in Arabic opinion mining, several researchers have provided surveys to capture advances in the field. While these surveys capture a wealth of important progress in the field, the fast pace of advances in machine learning and natural language processing (NLP) necessitates a continuous need for a more up-to-date literature survey. The aim of this article is to provide a comprehensive literature survey for state-of-the-art advances in Arabic opinion mining. The survey goes beyond surveying previous works that were primarily focused on classification models. Instead, this article provides a comprehensive system perspective by covering advances in different aspects of an opinion-mining system, including advances in NLP software tools, lexical sentiment and corpora resources, classification models, and applications of opinion mining. It also presents future directions for opinion mining in Arabic. The survey also covers latest advances in the field, including deep learning advances in Arabic Opinion Mining. The article provides state-of-the-art information to help new or established researchers in the field as well as industry developers who aim to deploy an operational complete opinion-mining system. Key insights are captured at the end of each section for particular aspects of the opinion-mining system giving the reader a choice of focusing on particular aspects of interest.

This work was made possible by NPRP 6-716-1-138 grant from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.
Authors' addresses: G. Badaro and H. Hajj, American University of Beirut, P.O.Box 11-0236 / Electrical and Computer Engineering Department, Riad El-Solh / Beirut 1107 2020, Lebanon; emails: {ggb05, hh63}@aub.edu.lb; R. Baly, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America; email: baly@mit.edu; W. El-Hajj, American University of Beirut, P.O.Box 11-0236 / Computer Science Department, Riad El-Solh / Beirut 1107 2020, Lebanon; email: we07@aub.edu.lb; K. B. Shaban and A. Hamdi, Qatar University, P.O. Box: 2713 – Doha / Computer Science and Engineering Department, Qatar; emails: khaled.shaban@qu.edu.qa, alihamdif@gmail.com; N. Habash, Computer Science Department, New York University of Abu Dhabi, Saadiyat Marina District, Abu Dhabi, United Arab Emirates; email: nizar.habash@nyu.edu; A. Al-Sallab, Computer Engineering Department, Cairo University, 1 Gamaa Street, Giza 12613, Egypt; email: ahmad.elsallab@gmail.com.

2375-4699/2019/05-ART27 $15.00
https://doi.org/10.1145/3295662

CCS Concepts: • **Information systems** → **Sentiment analysis**;

Additional Key Words and Phrases: Sentiment analysis, arabic natural language processing, sentiment lexicons, opinion mining, deep learning, sentiment analysis applications

**ACM Reference format:**
Gilbert Badaro, Ramy Baly, Hazem Hajj, Wassim El-Hajj, Khaled Bashir Shaban, Nizar Habash, Ahmad Al-Sallab, and Ali Hamdi. 2019. A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applications, and Visualizations. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 18, 3, Article 27 (May 2019), 52 pages.
https://doi.org/10.1145/3295662

## 1 INTRODUCTION

People are often interested in determining other people's opinions when, among other examples, seeking to buy products, sensing the public opinion on certain issues, or identifying trends. Governments and politicians are often interested in opinion (or sentiment) mining for defining policies and campaign strategies. As a result, opinion mining, which aims at automatically extracting people's opinions, has found significant importance in politics, social media, and business. Furthermore, the application of opinion mining in Arabic text (OMA) is a timely subject, given the importance of the Arabic language, which emerged as the fifth most-spoken language worldwide and recently became a key source of Internet content and now stands as the fourth most-used language on the Internet [343]. Arabic opinion mining has been an active area of research but still has many open challenges.

Arabic is both morphologically rich and highly ambiguous; it has complex morpho-syntactic agreement rules and a lot of irregular forms; and it has a large number of dialectal variants with no writing standards. Without proper processing and handling, learning robust general models over Arabic text can be difficult. Furthermore, compared to English, there are fewer freely available resources for Arabic opinion mining in terms of large clean sentiment lexica and annotated collections of opinions. These challenges have fueled extensive research interest in Arabic opinion mining. To support this wide interest in opinion-mining research for Arabic, several researchers have developed informative surveys of advances in the field. As technology is moving at a fast pace, machine learning (ML) and data-mining-rich fields are moving even faster. There is a need to have the latest progress frequently. There is also a need to cover different perspectives.

In the recent past, there have been efforts to provide informative surveys of trends and challenges in Arabic sentiment analysis [22, 34, 53, 92, 97, 99, 114, 142, 161, 163, 233, 239, 262, 302, 305]. An early survey [277] covered early and limited work in the field up to 2012. The article listed available corpora in Arabic and discussed features originally adopted by researchers and that can be categorized into language-independent features, information retrieval (IR) methods, and Arabic-specific methods. Another survey [77] followed in 2014 to cover updated methods. The coverage divided the literature into supervised, unsupervised, and hybrid approaches. The author concluded that stemming improved accuracies in most studies. N-grams choices varied with their successes, and the most successful models were Support Vector Machines (SVM) and Naïve Bayes (NB). The authors concluded that custom models were needed for different genres and reflected on open challenges that were getting little attention. Recently, other surveys [104, 146, 374] closed these gaps that included coverage of methods that have received more attention since earlier surveys and included: subjectivity classification, sentiment classification, aspect-based sentiment classification, resources, opinion holder, and opinion spam detection. Other surveys have also presented different perspectives on sentiment analysis, such as the impact of preprocessing approaches [114],

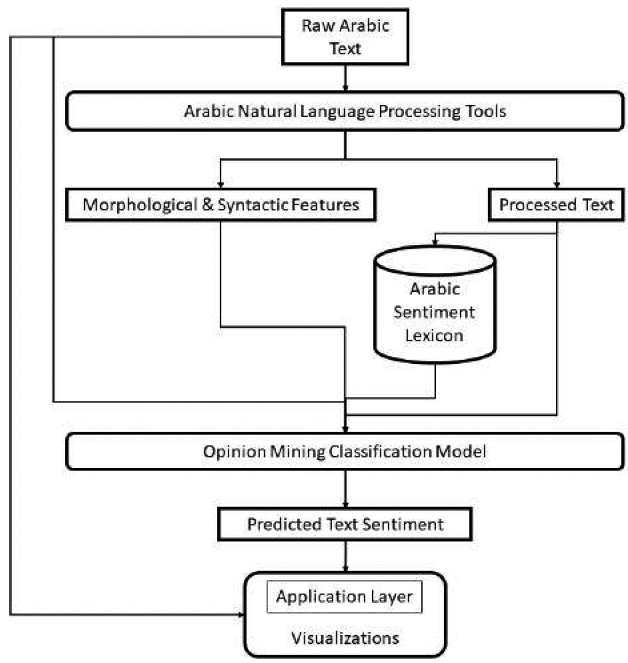


Fig. 1. Overall opinion-mining system.

but typically focused on machine-learning models. Reference [190] also covered the pros and cons of Arabic opinion-mining models covering couple of works on deep learning models. Salloum et al. [357] presented a survey on different applications and methods of Arabic text mining including opinion mining, mining the Holy Quran, and Arabic Web mining. More recently, there have been efforts to further advance progress with contests such as the ones introduced by SemEval [273, 296, 328, 361]. The outcome from these contests may provide some indication of new trends to analyze Arabic sentiment analysis.

Despite the increased attention received by the field with these surveys, several gaps remain. Most of the previous surveys have focused on models and algorithms with feature-engineering-based approach. There is a need for a wider coverage and synthesis of available processing software tools, development of lexical resources, features used, recent advances with deep learning approaches, a reflection of applications in Arabic sentiment analysis, system architecture, and visualization approaches. This survey aims at addressing the gaps in surveying these aspects. Furthermore, it provides a comprehensive system perspective. An overview of this system can be summarized as in Figure 1.

This article will be of interest to researchers and industry developers who may be interested in deploying state-of-the-art system modules for automated sentiment analysis. It helps them be up to date with the latest components needed for an opinion-mining system. The rest of the article is organized as follows: Section 2 covers challenges in Arabic language and presents advances in available tools for Arabic Natural Language Processing. Section 3 covers Arabic sentiment lexicons and sentiment corpora needed for opinion-mining models that are covered in Section 4. Section 5 presents applications of opinion mining in Arabic and Section 6 presents future directions for opinion mining in Arabic covering system architecture, visualizations, and open challenges.

## 2 ARABIC NATURAL LANGUAGE PROCESSING

In this section, we preview the main challenges associated with Arabic language and that affect performing NLP tasks in this language. Then, we describe the different tools and resources that are developed to remedy these gaps.

### 2.1 Challenges

Arabic language is both morphologically rich and highly ambiguous and hence is challenging for most NLP applications. According to Reference [232], a complete part-of-speech (POS) tagset in Modern Standard Arabic (MSA) has over 300K tags, whereas English has around 50. Also, MSA words have 12.2 morphological analyses, on average, whereas English has only 1.25 analyses per word. This high ambiguity is primarily due to Arabic orthography, which almost always omits the diacritics that are used to specify short vowels and consonantal doubling. Furthermore, Arabic has complex morpho-syntactic agreement rules and a lot of irregular forms. Finally, although MSA is the official form of Arabic, it is no one's mother tongue. In fact, Arabic has a large number of dialectal variants that are as different from MSA as Romance languages are different from Latin. These different dialects are used for everyday communication but do not follow MSA language rules. As a result, it becomes challenging to create NLP resources for Arabic dialects [87, 109]. Moreover, dialectal Arabic includes several misspellings and typical Arabic NLP morphological tools do not perform well on dialects [44]. The lack of writing standards is particularly challenging for processing Arabic dialect big data such as blogs and tweets, where we find a wide range of variations in spelling and even script choice, e.g., using Latin characters to write Arabic (a.k.a. Arabizi).

### 2.2 Tools and Resources

Many tools and resources have been developed to perform Arabic NLP functionalities, with some reaching competitive performance levels compared to languages that are extensively studied such as English [232]. Central to these efforts was the creation of corpora that are annotated with linguistic information. Examples include the Penn Arabic Treebank (PATB) [283], which includes phrase structure trees along with POS tags and morphological information, and the Columbia Arabic Treebank (CATiB) [229], which contains dependency trees. Such corpora have been instrumental in the development of different Arabic NLP tools for tokenization [166, 227], diacritization [227, 397], POS tagging and syntactic parsing [223, 289], and morphological analysis [284].

It is worth mentioning that most Arabic NLP tools are designed and trained on MSA datasets. Some of these tools were developed to perform specific functionalities. For instance, segmentation is tackled in Reference [359] by training deep recurrent neural models such as Long Short-term Memory (LSTM). Systems for Named Entity Recognition (NER) [324, 362] and diacritization [127] were developed as well. However, some other tools are comprehensive in the sense that they perform several NLP tasks under one framework. An example of such tools is MADAMIRA [326] that performs morphological disambiguation, which encompasses tokenization, POS tagging, base phrase chunking (BPC), diacritization, lemmatization, gender, number, person, in addition to NER. MADAMIRA is based on the ancestor systems MADA [227, 228] and AMIRA [165]. Also, in Reference [252], a grammar analyzer is developed based on basic morphology analysis using the Buckwalter Arabic morphological analyzer (BAMA-v2.0) [150], stemming, POS, BPC, and, finally, a rule-based system applying 400 Arabic grammar rules. Other software tool examples include TectoMT [330], ASMA [9], and NERA [324].

Although dialectal Arabic (DA) is quite prevalent, especially in social media, it has received little attention with a few early efforts [169, 270]. With the development of accurate NLP tools for MSA, more attention has been given to DA. For instance, a small corpus of Egyptian Arabic was annotated for morphological segmentation to learn segmentation models [295]. Al-Sabbagh and Girgu [62] described a supervised tagger for Egyptian Arabic social networking. Habash et al. [226] presented the first large-coverage morphological analyzer for Egyptian Arabic. A tool for automatic identification of dialectal Arabic (AIDA) was also developed in Reference [196]. Dialects of the Arabian Gulf region were also studied very recently in Reference [265]. Given the inherited similarity between MSA and DA, there have been efforts to map from DA back to MSA to exploit the existing MSA resources [127, 159, 358]. Additional work has also been done for dialect identification [140, 197, 394]. Recently, given the significant performance improvements achieved for sentiment analysis when using word vector representations, Soliman et al. [371] developed AraVec. AraVec is a pre-trained distributed word representation open source project that aims to provide the Arabic NLP research community word-embedding models. The first version of AraVec provides six different word-embedding models built from three different Arabic content domains: tweets, World Wide Web pages, and Wikipedia Arabic articles. The total number of tokens used to build the models amounts to more than 3.3 billion.

In summary, several Arabic NLP tools and resources, specifically those targeting MSA, have been developed achieving performances that are comparable to those of the English language. Further efforts are still needed to achieve better results when it comes to DA given the prevalence of using different dialects across the Arab world.

## 3 SENTIMENT LEXICAL RESOURCES

While Arabic NLP tools help in providing linguistic and grammatical features of the text, this section presents Arabic sentiment resources that were developed to provide a deeper understanding of sentiment and semantic aspects of the text.

### 3.1 Sentiment Lexica

The establishment of advanced Arabic NLP tools and resources, as discussed in Section 2.2, is useful for opinion mining as syntactic and morphological features can mitigate the impact of Arabic's complex morphology [11, 14, 15]. Also, it is crucial to equip opinion models with deeper insights into the text semantics. Sentiment lexica are crucial resources for most sentiment analysis algorithms [215]. Several researchers worked on developing Arabic sentiment lexica to improve the accuracy of OMA. In general, there are two approaches for creating such lexicons; manual and automatic. The manual approach consists of manually determining the sentiment of a list of Arabic words that are extracted from a certain domain or dataset. The resulting lexicon is usually highly accurate but limited in size due to the time-consuming task of annotation. Several resources were developed manually, from which we mention the following. ArabSenti [14] contains 3,982 adjectives extracted from 400 documents belonging to the Arabic Tree Bank (ATB) part 1 V3.0 [283]. These adjectives were manually labeled by two Arabic native speakers as positive, negative, or neutral and were reviewed by a linguist expert. Similarly, SIFAAT [8] was manually developed and consisted of 3,325 adjectives labeled as positive, negative, or neutral. Using SIFAAT showed significant improvement in the accuracy results of the proposed subjectivity and sentiment analysis system. Al-Kabi et al. [47] collected 1,080 Arabic reviews that only include Arabic characters. They applied first a set of preprocessing steps: removal of digits, punctuation, symbols, and special characters; normalization; and tokenization. They annotated the reviews for sentiments (positive versus negative) as well as for domain (eight domains in total). Based on annotation, they created two general-purpose lexicons one including positive terms and another one consisting of negative terms. Moreover, eight domain-specific lexicons were manually developed. Al-Rowaily et al. [59] presented BiSAL, a sentiment lexicon specifically related to cyber threats, radicalism and conflicts.

BiSAL is composed of an English version that includes 279 sentiment terms and an Arabic version of 1,019 terms. The resource is publicly available.[1] Abdulla et al. [19] also used a manual approach for creating their lexicon by translating 300 English seed words from SentiStrength [379] to Arabic. This seed set was then expanded using synonyms and antonyms and by including emoticons. Unlike ArabSenti and SIFAAT, this lexicon does not include neutral terms and is not publicly available. NileULex is a recently developed sentiment lexicon [179, 181] that includes single-word terms as well as compound phrases for MSA and dialects. Although the extraction of the terms and compound phrases was done automatically from social media, the annotation was performed manually. The authors chose single-word terms to be as unambiguous as possible, and compound phrases were used to overcome ambiguity in single-word terms. NileULex consists of 5,953 expressions annotated either as positive or negative. NileULex was used by the NileTMRG team [178] that participated at SemEval 2016 Task 7 [273]. El-Beltagy [180] extended NileULex by presenting a method for automatically assigning strength scores or weights to NileULex entries as well as making the resulting lexicon "WeightedNileULex" publicly available. In Reference [248], Ibrahim et al. presented AIPSeLEX, idioms/proverbs sentiment lexicon for MSA and DA. AIPSeLEX was manually collected and annotated at sentence level with semantic orientation (positive or negative). AIPSeLEX consists of 3,632 phrases annotated for sentiment with the help of three native Arabic speakers.

The larger the lexicon, the better the coverage of the language vocabulary and thus the better the sentiment modeling. Therefore, automatic approaches to develop sentiment lexica are required. Although an automatic approach might be prone to errors and noise, it is cheap and less time-consuming. Usually, automatic approaches are supported with some manual efforts to allow for automatic expansion. They usually consist of harvesting existing sentiment lexical resources in English and trying to create equivalent versions in Arabic. For instance, Mourad and Darwish [310] utilized the manually developed ArabSenti to perform automatic expansion through graph reinforcement. They translated ArabSenti into English and then used machine translation tables of English-MSA and English-DA to enrich ArabSenti with new MSA and DA terms. They also translated the MPQA lexicon [390] from English to Arabic using the Bing machine translation tool and combined all lexica together to use them in their opinion-mining classification system. However, the authors did not report the total number of terms they finally obtained. Alhazmi et al. [89] linked the Arabic WordNet (AWN) [143] to the English SentiWordNet (ESWN) [118, 211] through synset offset information. Their approach had limited coverage (around 10K lemmas only) and did not define a process for using the lexicon in practical applications given Arabic's complex morphology. Furthermore, it was not made publicly available and was not evaluated in the context of a sentiment classification application. In Reference [240], starting with an English sentiment lexicon derived from the General Inquirer English sentiment lexicon, Hassan et al. proposed an automatic approach for creating a lexicon in other languages by using semantic relationships available through WordNet. They applied their approach on Arabic and Hindi using English WordNet, Arabic WordNet, and Hindu WordNet. Following a similar concept, Mahyoub et al. [287] used AWN 2.0 synset relations such as "near_antonym" and "near_synonym" to automatically expand a seed list translated from English to Arabic where the list is proposed by Turney and Littman [383]. Since they could not cover all words in the AWN with the initial seed list, they randomly picked up terms and added them to the list. The resulting lexicon had a size of around 7.6K words and improved sentiment classification results when used in addition to other stylistic and syntactic features. Badaro et al. [120] benefited from the success and availability of the ESWN and developed a lemma-based large-scale Arabic sentiment lexicon: ArSenL. This lexicon is the result of combining two sub-lexicons.

[1] http://www.abulaish.com/bisal.

The first consists of mapping AWN 2.0 to ESWN 3.0 by going through sense map files across English WordNet versions similar to the work of Alhazmi et al. [89] but with an important addition: standardizing the lemma format of AWN terms to LDC format. This step is important in making the resource easy to integrate with other Arabic NLP tools. The second sub-lexicon, ArSenL-Eng, was the result of mapping Standard Arabic Morphological Analyzer (SAMA) [284] directly to ESWN 3.0 by matching SAMA gloss terms to ESWN synset terms. ArSenL is publicly available and includes around 29K lemmas along with their corresponding POS tag, EWN synset ID, and ESWN sentiment scores. It also improved accuracy in subjectivity and sentiment classification tasks. Recently, Badaro et al. [219] presented ArSEL, a large-scale Arabic Sentiment and Emotion Lexicon, an extension to ArSenL with the addition of eight emotion scores to most ArSenL Arabic lemmas. The emotion scores are extracted from a WordNet-based English emotion lexicon, EmoWordNet [125]. Eskander and Rambow developed SLSA [210] using almost the same approach as ArSenL, but they used instead AraMorph [149], which is the publicly available version of SAMA. Moreover, when matching gloss terms to synset terms, they applied different heuristics and back-off measures to ensure higher coverage but at the expense of additional noisy mappings. SLSA consists of around 35K Arabic lemmas and is publicly available. Similarly, Sabra et al. [354] developed a sentiment lexicon for Arabic by first developing an English sentiment lexicon and then mapping it to Arabic. Starting with a seed list of 18 words equally split between positive and negative, the authors utilized WordNet semantic relations to expand the list. During the expansion, the depth needed to reach a certain term was recorded and then used to compute scores for positive and negative sentiments. The result was an annotated EWN for sentiment. Next, the authors mapped the EWN terms to SAMA lemmas through synset terms to SAMA gloss matching. The lexicon that consists of around 78K entries was tested on an OCA corpus and achieved comparable performance to ArSenL with a simple sentiment classification model. Abdulla et al. [18] adopted a semi-automatic approach to develop their sentiment lexicon. They manually translated from English to Arabic 300 terms from SentiStrength and then they used synonym tables to expand the initial list. For the automatic part, they have translated the remaining of SentiStrength using Google Translate, and they also investigated using an annotated corpus for sentiment to extract positive and negative words using a Term-Frequency weighting scheme. Abdul-Mageed and Diab [13] extended their manually developed sentiment lexicon (SIFAAT) automatically by using machine translation and statistical formulation based on pointwise mutual information to create SANA. SANA included 224,564 entries that cover Modern Standard Arabic (MSA) as well as Egyptian and Levantine dialects. These entries are not distinct and possess many duplicates. Developing SANA involved gloss matching across Arabic lexical resources such as THARWA [167] and SAMA [284] and English sentiment resources, the Affect Control Theory lexicon [243], and ESWN [118]. It is also composed of different lexicons such as Yahoo Maktoob, a dataset from Twitter, and an automatically translated YouTube comments dataset [16]. Unlike SIFAAT, SANA was not tested in a sentiment classification task and is not publicly available. Chen and Skiena [153] proposed an expansion approach of an English sentiment lexicon to more than 130 languages using knowledge graph construction. Using Wiktionary, Google Translate, transliteration, and WordNet, the authors tried to get semantic links between words in their different translations. Using graph propagation algorithm and the semantic links, they extended sentiment polarities from a set of terms to their neighbors. For Arabic, they were able to construct a sentiment lexicon of 2,794 words starting with a set of 1,422 positive and 2,956 negative English terms. Al-Ayyoub et al. [33] developed a sentiment lexicon of 120K stems that were collected by crawling the web specifically (AlJazzera.net) and by using the Abuaiadah dataset for document classification [1]. ArSeLEX [247, 250] is an automatically generated and publicly available sentiment lexicon that contains 5,244 adjectives. ArSeLEX was developed by getting synonyms and antonyms of 400 adjectives manually annotated for sentiment. El Sahar and El-Beltagy [208] used a set of lexico-syntactic rules to automatically extract Arabic phrases that represent opinion. They applied the rules on DA, specifically the Egyptian Cairene dialect. Starting with a seed set manually translated from English to Arabic, the authors defined a set of patterns that could capture subjective text in Arabic. After extracting the subjective slang, pointwise mutual information (PMI) was used to determine the polarity of the extracted phrase using an annotated set of tweets. Of 7.5M cleaned tweets, they were able to extract 633 expressions with 89% precision. El Sahar and El-Beltagy [209] used a supervised learning model to generate a sentiment lexicon with positive and negative labels. From an annotated corpus of 35K sentences, they extracted unigrams and bigrams and used them for sentiment classification with SVM. The terms with highest positive coefficients were labeled as positive sentiment, and the terms with lowest negative coefficients were considered as holding negative opinion. Arabic Senti-Lexicon [52] was developed using a combination of automatic approaches followed by manual adjustments. Terms from the MPQA lexicon [390] were translated to Arabic using Google Translate, and the translation was manually adjusted. The list was expanded by adding synonyms and antonyms. Two types of annotations were provided for the words: a manual annotation provided by three annotators consisting of a score on a scale varying between −5 to +5 with −5 being very negative and +5 very positive. A second automatic score was computed using PMI such that the sum of the positive and negative scores is equal to 1. Arabic Senti-Lexicon consists of 13,760 terms and covers MSA as well as dialects. Al-Twairesh et al. [80] focused on developing sentiment lexicons for DA to improve the accuracy of opinion-mining systems applied on Twitter data. The authors first collected around 2.2 million tweets that included specific positive and negative seed words. These seed words, in addition to emoticons, helped in performing automatic sentiment annotation for the collected tweets. Using the annotated twitter dataset, Al-Twairesh et al. created two sentiment lexicons: AraSenti-Trans and AraSenti-PMI. For AraSenti-Trans, the authors ran the tweets through MADAMIRA [326], extracted Arabic lemmas, and tried to match the provided English gloss with existing English sentiment lexicons, Liu Lexicon [246] and MPQA [390], to assign sentiment labels for the Arabic lemma based on defined matching rules. AraSenti-Trans included around 132K Arabic terms with around 60K positive and 72K negative. AraSenti-PMI is generated by computing PMI for all the words occurring at least more than 5 times in the positive and negative sets of tweets and then generating a sentiment score for the word using its PMI. AraSenti-PMI included around 57K positive terms and 37K negative terms. They evaluated both lexicons in sentiment classification tasks on different Twitter datasets (AraSenti-Tweet [79], ASTD [314], and RR [337]) and obtained best average F1-scores of 88.92%, 59.8%, and 63.60% on the three datasets, respectively. The authors of Reference [274] applied the annotation technique of Best-Worst Scaling to obtain real-valued sentiment association scores for words and phrases in three different domains: general English, English Twitter, and Arabic Twitter. They showed that on all three domains the ranking of words by sentiment remains remarkably consistent even when the annotation process was repeated with a different set of annotators. The authors also asked the annotators to indicate the minimum perceptible difference in sentiment strength between two terms. The annotated data are publicly available.[2] For annotation, they utilized CrowdFlower (now known as Figure Eight). Three different lexicons were created: (1) SemEval 2015 English Twitter Lexicon consisting of 1,515 high-frequency English single words and simple negated expressions commonly found in tweets; (2) the SemEval 2016 Arabic Twitter Lexicon that includes 1,367 most-frequent terms and expressions in Arabic tweets; and (3) the SemEval 2016 General English Sentiment Modifiers lexicon, also known as the Sentiment Composition Lexicon for Negators, Modals and Degree Adverbs: It consisted of 1,621 positive and negative single words in addition to 1,586 high-frequency phrases. Last but not least, Assiri et al. [113] created a sentiment lexicon targeting

[2] www.saifmohammad.com/WebPages/BestWorst.html.

Saudi dialects by manually annotating terms extracted from Saudi social media and adding it to a modified version of ArSenL. The modified version of ArSenL was obtained by removing punctuations and diacritics from lemmas. Table 1 summarizes the important characteristics of some of the sentiment lexicons presented above. In summary, similarly to Arabic NLP tools, a good amount of work has been invested to develop sentiment lexicons for MSA, but further efforts are needed for DA. Moreover, although some efforts were put to develop large-scale Arabic sentiment lexicons, more energy is needed to achieve a similar scale with good accuracy compared to English.

### 3.2 Sentiment Corpora

Similarly to other machine-learning tasks, sentiment analysis requires the existence of annotated data. Annotated sentiment corpora will help in training opinion-mining models as well as testing the performance of these models. Abdul-Mageed [14] annotated 2,855 sentences from PATB part 1 corresponding to the first 400 documents [283]. Two college-educated Arabic native speakers annotated the data with four possible labels; objective (OBJ), subjective positive (S-POS), subjective negative (S-NEG), and subjective neutral (S-NEUT). The dataset can be used for subjectivity classification as well as sentiment classification. The distribution of the annotations was 1,281 OBJ, a total of 1,574 SUBJ, where 491 were deemed S-POS, 689 S-NEG, and 394 S-NEUT. The dataset is available on request. Abdul-Mageed and Diab [12] expanded the initial corpus by annotating 5,342 sentences from 30 Wikipedia talk pages and 2,532 sentences from threaded web forums. In total, AWATIF consisted of 10,729 sentences annotated for sentiment. Rushdi-Saleh et al. [351] presented a corpus consisting of 500 movie reviews in Arabic equally divided between positive reviews and negative reviews. The data were extracted from different web pages and blogs. Aly and Atia [110, 313] collected a large-scale corpus consisting of 63,257 book reviews in Arabic, each rated with a scale between 1 and 5; 16,486 users submitted their reviews for 2,131 different books. They utilized the dataset to test different sentiment classification techniques. They also created a sentiment lexicon consisting of single and compound words extracted from the corpus. Saad et al. [352, 353] proposed an annotation strategy for subjectivity in multilingual setting. Starting with an annotated corpus in English with 5K subjective sentences and 5K objective sentences, the authors built an English sentiment analysis model using trigrams and Naïve Bayes to automatically label English sentences in a parallel English-Arabic corpus. The predicted labels were assigned to the Arabic text as well. Several parallel corpora were used and a total of 148K sentences were automatically annotated to be either subjective or objective. A sample of 330 sentences was manually annotated to evaluate the automatic annotation, and an average F-measure of 68.83% was obtained with an average accuracy of 68.8%. Elnagar and Einea [202] presented BRAD, a large-scale book reviews dataset in Arabic. The dataset includes 510,598 book reviews provided by 76,530 users on 4,993 books. Each review is annotated with a scale from 1 to 5. A cleaned and balanced version of 156,506 reviews was also released. Several opinion-mining models were tested to create baseline results for future research. As an extension to BRAD corpus [202], Elnagar et al. [204] released a large dataset of hotel reviews written in Arabic and extracted from Booking.com. HARD includes 490,587 hotel reviews collected from the Booking.com website. Each record contains the review text in the Arabic language, the reviewer's rating on a scale of 1 to 10 stars, and other attributes about the hotel and the reviewer. The dataset includes both MSA and dialects. The authors made publicly available the full unbalanced dataset as well as a balanced subset and published some baseline results on the dataset. Farra et al. [214] worked on annotating an Arabic corpus for sentiment as well as for the targets of the sentiment in a two-stage process using crowdsourcing via Amazon Mechanical Turk. Task 1 involved identifying the main named entities in a list of comments. Three annotators worked on the task and the intersection of their annotations was selected. The

Table 1. Summary of Arabic Sentiment Lexicons

| Lexicon | Lemmas | # Entries | Labels | Method | Tested for SA | Source | POS Tags | MSA | Dialect | Emoticons | Availability |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ArabSenti [14] | Not specified | 3,982 | Pos, Neg, Obj | Manual annotation by 2 Arabic native speakers and 1 expert linguist | Yes | Newswire | Adj | Yes | No | No | On request |
| SIFAAT [8] | Yes | 3,325 | Pos, Neg, Obj | Manual annotation by 2 Arabic native speakers and 1 expert linguist | Yes | PATB [283] | Adj | Yes | No | No | On request |
| Abdulla et al. [19] | No | 3,479 | Pos, Neg | Manual translation and expansion with synonyms | Yes | SentiStrength [379] | All | Yes | No | Yes | No |
| Mourad and Darwish [310] | No | Not reported | Pos, Neg, Obj | Automatic approach using graph reinforcement applied on MT tables | Yes | ArabSenti [14], MT tables from Moses [275] | All | Yes | Yes | No | No |
| Alhazmi et al. [89] | Yes | 10K | Scores for Pos, Neg, Obj | Automatic Mapping of AWN 2.0 to SWN 3.0 | No | AWN 2.0 | All | Yes | No | No | No |
| Abdulla et al. [18] | No | 16,800 | Pos, Neg | Semi-Automatic: manual translation of 300 seed words from SentiStrength and adding synonyms; Machine translating SentiStrength terms and extracting words from annotated Corpus | Yes | SentiStrength, Corpus | All | Yes | Yes | Yes | No |
| ArSenL [120] | Yes | 28,760 | 3 scores for Pos, Neg, Obj whose sum = 1 | Automatic mapping of AWN 2.0 to SWN 3.0 union gloss-synset string matching between SAMA and SWN | Yes | AWN 2.0 [143], SAMA [284] | All | Yes | No | No | Yes |
| Mahyoub et al. [287] | Yes | 7,576 | Score for Pos, Neg, Obj | Expansion starting from seed list using AWN 2.0 synset relations | Yes | AWN 2.0 | All | Yes | No | No | No |
| SANA [13] | No | 224,564 | Pos, Neg, Obj | Manual annotation and automatic expansion through machine translation and a statistical method based on PMI | No | SIFAAT, Yahoo Maktoob, SAMA, Affect Control Theory Lexicon, Tharwa, Twitter, SWN, Youtube Lexicon, General Inquirer | All | Yes | Yes | Yes | No |
| Al-Ayyoub et al. [33] | No | 120,000 | Pos, Neg, Obj | Collecting stems and crawling the web for expanding coverage. Translating terms to English and finding corresponding sentiments in English | Yes | Aljazeera.net | All | Yes | No | No | No |

(Continued)

Table 1. Continued

| Lexicon | Lemmas | # Entries | Labels | Method | Tested for SA | Source | POS Tags | MSA | Dialect | Emoticons | Availability |
|---|---|---|---|---|---|---|---|---|---|---|---|
| El Sahar and El-Beltagy [209] | No | 6,708 | Pos, Neg | Selecting unigrams and bigrams of annotated datasets that correspond to highest and lowest positive and negative coefficients when used in SVM classification | Yes | Reviews of movies, products hotels and restaurants (Souq, Qaym, TripAdvisor, Elcinemas.com) | All | Yes | Yes | No | Yes |
| ArSeLex [247, 250] | No | 5,244 | Pos, Neg, Obj | A list of 400 adjectives is manually annotated and then automatically expanded by checking synonyms/antonyms derived from translations and dictionaries | Yes | Tweets & online dictionaries | Nouns, Adj | Yes | Yes | No | Yes |
| SLSA [210] | Yes | 34,821 | 3 scores for Pos, Neg, Obj | Automatic gloss-synset matching between AraMorph English gloss terms and SWN synset terms adjusted with heuristics and manual back-offs | Yes | SWN 3.0 [211] and AraMorph [149] | All | Yes | No | No | Yes |
| NileULex [179] | No | 5,953 | Pos and Neg | Automatically collected terms and compound phrases from social media postings and manually annotated them; Single-word terms were chosen to be as unambiguous as possible and compound words were used to overcome ambiguity | Yes | Social media | All | Yes | Yes | Yes | No |
| Arabic Senti-Lexicon [52] | Yes | 13,760 | Manual scoring (-5,+5). Automatic scoring for Pos & Neg (sum = 1) | Translation of MPQA using Google translate followed by manual adjustment; Term expansion using synonyms. Labels/scores added manually by 3 annotators, and automatically using PMI between terms and seed words in a large set of reviews | Yes | MPQA [390] | All | Yes | Yes | No | No |
| AraSenti-Trans/PMI [80] | No | 132K/94K | Pos;Neg;Obj/ Score | Automatic generation using gloss matching against English sentiment lexicons / Computing PMI using automatically annotated Twitter dataset | Yes | Twitter, MADAMIRA [326], MPQA [390], Liu Lexcion [246] | All | Yes | Yes | No | Yes |
| Assiri et al. [113] | No | 14,000 | Pos;Neg;Obj | Manual annotation of Saudi Dialect terms in addition to entries from ArSenL obtained after removing diacritics and punctuations from its lemmas | Yes | Saudi dialect and ArSenl [120] | All | Yes | Yes | No | No |

second task was to find the sentiment (positive, negative, neutral) toward the entities derived in Task 1. The dataset was selected from QALB [303, 393] and consisted of 1,177 comments of three different domains: politics, culture, and sports. They annotated a set of 4,345 Targets: Forty-three percent of them had a positive sentiment, and 57% had negative opinions about them. The corpus is publicly available.[3] The Human Annotated Arabic Dataset (HAAD) [71] consists of 1,513 book reviews extracted from LABR and manually annotated for aspects and their corresponding sentiments with four class labels (positive, negative, conflicting, and neutral) and aspect categories and their corresponding aspect category sentiments. HAAD included 1,296 distinct aspect terms. Al-Sarhan et al. [65] worked on developing models for aspect extraction and sentiment classification toward those aspects similarly to the subtasks presented at SemEval-2014 Task 4 [329]. For this purpose, they extracted data from Facebook and Twitter related to the 2014 attacks on Gaza and manually annotated them for aspects, aspects' categories, and the corresponding sentiments toward the aspects and toward their categories. In total, they collected a set of 2,265 news posts divided among three sentiment classes: positive, negative, and neutral. In total, 9,655 aspects were extracted belonging to four categories: Plans, Results, Peace, and Parties. They ran simple sentiment classification techniques on their annotated corpus to set baseline results for future research. In Reference [337], the authors collected a set of 8,868 Arabic tweets and manually annotated them for subjectivity and sentiment with five labels: polar, positive, negative, neutral, and mixed. They also annotated the corpus automatically with a variety of linguistically motivated features: morphological, syntactic, semantic, stylistic, and social signals. ElSahar and El-Beltagy [209] extracted around 33,116 Arabic reviews about movies, hotels, restaurants, and products. They collected the reviews from several websites, such as Elcinemas.com, Souq.com, and tripadvisor.com. They made the dataset publicly available. While the previous datasets were mainly focused on MSA, Al-Kabi et al. [45, 49] developed an Arabic sentiment corpus covering MSA as well as several Arabic dialects. The corpus consists of 250 topics and 1,442 reviews extracted from five domains: Economy, Food-Life style, Religion, Sport, and Technology. The corpus was built manually to ensure high accuracy in terms of annotation. Siddiqui et al. [366] developed a multifaceted, multilingual corpus for hierarchical sentiment analysis. The different facets included hierarchical nominal sentiment labels, a numerical sentiment score, language used in the review, and the dialect. The annotated corpus consisted of 191K reviews of hotels in Saudi Arabia. The reviews were divided into 11 different sentiment categories ranging from exceptional to very poor. The corpus contains 1.8 million tokens. Reviews were mostly written in Arabic and English, but there were instances of other languages as well. A Twitter corpus written mostly in Saudi dialect and consisting of 4,700 tweets was also annotated for sentiment by Assiri et al. [112]. Duwairi et al. [174] annotated a corpus of 3,206 Arabizi tweets using crowdsourcing. Emerging work handling Arabizi is also presented by Guellil et al. [225]. Nabil et al. [314] created the Arabic Sentiment Tweets Dataset (ASTD) consisting of 10,006 tweets classified as positive (799 tweets), negative (1,684), mixed (832), and objective (6,691). They investigated different statistics on the dataset and presented a set of benchmark experiments for baseline comparisons. Medhaffar et al. [292] manually annotated an Arabic corpus written in the Tunisian dialect consisting of 17K comments extracted from Facebook. The Tunisian Sentiment Analysis Corpus (TSAC) consists of an almost equal number of positive and negative comments. The authors tested several sentiment analysis models on the developed corpus. Recently, AraSenti-Tweet [79] was developed and made publicly available at the end of 2017. AraSenti-Tweet includes 17,573 tweets mainly covering Saudi dialect and annotated with four sentiment labels: positive, negative, neutral, and mixed. Benchmark and baseline results are also provided for easier research

[3] www.cs.columbia.edu/ noura/Resources.html.

use and comparison. MIKA [249] is a multi-genre tagged corpus for MSA and DA. MIKA was manually collected and annotated at sentence level with semantic orientation (positive or negative or neutral). A number of rich set of linguistically motivated features such as contextual intensifiers, contextual shifters, negation handling, and syntactic features for conflicting phrases were used for the annotation process. The data consist of MSA and Egyptian DA from different sources such as Twitter, hotel reviews, product reviews, and TV programs reviews. MIKA includes 4,000 reviews annotated for sentiment. ArSAS [200] is an Arabic corpus of tweets annotated for the tasks of speech-act recognition and sentiment analysis. A set of 21K Arabic tweets covering multiple topics were collected, prepared, and annotated for six different classes of speech-act labels, such as expression, assertion, and question. In addition, the same set of tweets were also annotated with four classes of sentiment: positive, negative, mixed, and neutral. ArSentD [335] is a dataset of 4,000 tweets with the following annotations: the overall sentiment of the tweet, the target to which the sentiment was expressed, how the sentiment was expressed, and the topic of the tweet. The authors claimed that the results confirm the importance of these annotations at improving the performance of a baseline sentiment classifier. They also claimed that ArSentD helps in closing the gap of training in a certain domain and testing in another domain. Hamdi et al. [234] annotated a set of 15,274 reviews extracted from online reviews, Facebook comments, and Twitter posts related to governmental services. The annotation was multifaceted and was not only limited toward identifying the sentiment polarity but also domains, dialects, and linguistic issues. Annotators were from different Arab countries and they were trained for the annotation task. In References [255, 256], Itani et al. extracted a corpus from Facebook comments consisting of DA. The corpus was manually annotated into five sentiment labels: positive, negative, neutral, dual, and spam. The corpus included 2K comments in total distributed among the five classes. In Reference [54], Al Mukhaiti et al. extracted a dataset from multiple social network websites, namely Youtube, Twitter, Facebook, Instagram, and Keek. After filtering the data, a total of 2,009 reviews were annotated to be either positive (1,004) or negative (1,005). While previous efforts were mainly about annotating Arabic texts, and similarly to References [352, 353], Elnagar et al. [203] studied the impact of translating resources from English to Arabic to perform sentiment analysis on Arabic text. They claimed that although the translation might not be accurate grammatically, sentiment was preserved in most of the cases due to main keywords. In 2018, Ahmed [30] developed a large-scale Arabic sentiment corpus (GLASC) consisting of online news articles and metadata. GLASC was annotated for both sentiment labels (positive, negative, neutral) and sentiment scores between −1 and +1. GLASC consists of around 620K news articles. Ahmed also worked on translating the English SenticNet [151] to Arabic and creating ArSenticNet, which includes 48K Arabic concepts. The author experimentally tested multiple sentiment classification approaches at document level and sentence level with the best models using fusion between SVM and HMM for documents and SVM and Linear regression for sentences.

Development of Arabic sentiment lexicons and corpora has gained interest among researchers. In fact, Arabic sentiment corpora are currently of comparable sizes compared to English such as, for example, the balanced version of BRAD [202] with more than 156K reviews. Most of the datasets created were developed in-house, and several were not made publicly available. It is crucial to share these resources publicly to help accelerate research work in the domain. It would also be beneficial to have a unified platform where these resources can be found to increase visibility to researchers interested in Arabic opinion mining.

After discussing different aspects of Arabic resources covering tools, lexicons, and annotated sentiment corpora, we describe in the next section state-of-the-art opinion-mining approaches that utilize extensively these resources.

## 4 OPINION MINING APPROACHES

Over the past two decades, several approaches have been adopted to perform sentiment analysis (SA) in different genres of English text (reviews, messages, and micro-blogs) and have reached satisfactory performances [164, 271, 278]. Sentiment models have benefited from available NLP resources in English such as POS taggers and grammatical parsers, as well as lexical resources including sentiment lexica such as SentiWordNet [211] and MPQA [390] and sentiment-labeled corpora such as the Stanford Sentiment Treebank [370].

These resources are often scarce or nonexistent in other low-resource languages, which poses serious obstacles to performing accurate sentiment analysis in these languages. Many of these resources have only recently been developed for Arabic, as discussed in Sections 2 and 3, which opened doors for extensive efforts to explore the application of sentiment analysis in Arabic text. Early efforts included manual attempts by the use of conceptual model to represent opinion as described in Reference [279]. This section discusses approaches and highlights the progress made at developing sentiment analysis solutions in Arabic.

Sentiment analysis approaches can be categorized, at the macro-level, into *lexicon-based* and *supervised learning* approaches.

Lexicon-based approaches mainly depend on algorithms that use sentiment lexica to predict the sentiment, whereas supervised learning approaches are modeling algorithms trained on labeled examples to learn complex relations between features extracted from the texts and the associated sentiment labels. Supervised models can be further categorized into models based on *feature engineering* and others based on *deep learning*. Feature engineering-based approaches predict sentiment by learning from a variety of features that are selected to capture different aspects of the text. However, deep learning is considered the state of the art of machine learning and has achieved significant success in many domains, mainly computer vision and NLP [391].

It generally uses embedded representations of text units (characters or words) as input features to train different neural networks architectures.

Recently, combination of multiple approaches have been used for Arabic sentiment classification for Twitter data [78, 81, 158, 364] and online reviews [184, 264]. For instance, Refaee and Rieser [342] participated in SemEval 2016 Task 7 [273] detecting sentiment intensity of English and Arabic phrases. The authors proposed a hybrid approach that consists of a supervised model that uses ensemble of trained linear regression models followed by a rule-based approach with sentiment lexica to adjust the predicted intensities. Refaee and Rieser were able to achieve the first rank in this task with a 0.536 Kendall score. Such models can also be observed across the systems participating in SemEval 2017 Task 4: Sentiment Analysis for Twitter [347]. Five subtasks were dedicated for Arabic Sentiment Analysis in Twitter data. The tasks were as follows:

- subtask A: Given a tweet, classify it as either positive, negative or neutral;
- subtask B: Given a tweet and a topic, decide whether the tweet is positive or negative;
- subtask C is similar to subtask B, but there are five sentiment classes: strongly and weakly positive/negative and neutral;
- subtask D: Given a tweet and a topic, estimate the distribution of tweets across the positive and negative labels, and, last,
- subtask E is similar to subtask D but with five different sentiment labels as in subtask C.

Below, we provide for each category a detailed description of the main sentiment approaches that were developed and were applied to Arabic.

### 4.1 Lexicon-based Approaches

Sentiment lexicon-based approaches are to some extent simple and usually unsupervised. They do not need large-scale and expensive sentiment-labeled datasets that are used to train supervised machine-learning models. However, these approaches require external resources, mainly sentiment lexica, to predict sentiment. One of the most widely used lexicon-based unsupervised algorithms predicts the sentiment of a text by accumulating the scores of its words, based on some lexicon, and then checking the sign of the resulting score. Usually, scores of +1, −1, and 0 were assigned to positive, negative, and neutral words, respectively. Alternatively, numerical scores indicating sentiment intensity were also used when available.

Awwad and Alpkocak [116] evaluated the performance of four different sentiment lexicons derived from English sentiment lexicons Harvard IV-4 Dictionary and MPQA. Awwad and Alpkocak obtained similar performance on different datasets and concluded that lexical-based approaches for sentiment analysis have similar performances. Hamdi et al. [234] proposed CLASENTI a class-specific sentiment analysis framework. CLASENTI includes a lexicon-based model to calculate polarity strengths. The annotated lexicon is filtered based on the specific classes of the domain and dialect. Ahmad et al. [28] developed a general approach for sentiment analysis from Financial data streams by applying their English approach to Chinese and Arabic financial texts. By identifying keywords specific to financial news, the authors looked at the neighboring terms of those keywords to identify the polarity by using a sentiment lexical resource. Mohammad et al. [297] experimented with different available sentiment lexicons to perform sentiment classification on Arabic text extracted from social media. They used existing Arabic sentiment lexicons and translated ones from English to Arabic. A combination of an Arabic Dialectal hashtag lexicon [396] and a translation of NRC Emotion lexicon achieved best results. Elhawary and Elfeky [199] utilized an unsupervised approach to perform sentiment classification on Arabic Business reviews. Using a sentiment lexicon and a set of rules that detect intensification and negations, each sentence was labeled with a total score that defined as a result the polarity of the sentence. In References [74, 75], Al-Subaihin et al. proposed an unsupervised technique to perform sentiment analysis on Arabic dialectal text at the sentence level. They utilized an unsupervised lexicon-based approach to try to overcome the challenge of limited resource availability for dialects. They used human computing, which is a technique that integrates human effort to solve certain steps in a system. To keep the sentiment lexicon enriched, the authors built an online game where players help in annotation of the lexicon. The lexicon was used to assess the polarity of a given sentence. Siddiqui et al. [367, 368] presented a sentiment analysis system that is based on a set of handcrafted rules extracted with the help of sentiment lexicons such as text begins with, ends with, includes, or is equal to. They tested their technique on two different corpora achieving accuracies of 93.9% and 85.6%.

In Reference [18], the lexicon-based algorithm was applied to MSA comments using a sentiment lexicon of 16,800 words that was constructed automatically and manually. In Reference [20], the same approach was applied to tweets and comments, written in both MSA and DA, using a sentiment lexicon of 4,800 words that was created by expanding a seed of 300 words using synonym and antonym relations. Experimental results indicate that light-stemming degraded the performance. Furthermore, in Reference [33], a lexicon-based algorithm was applied to tweets, written in both MSA and DA, using a sentiment lexicon of 120,000 words that was constructed by translating existing lexica in English. Other works that adopted this algorithm include References [19] and [170]. Improvements to this algorithm were explored in Reference [175] to account for negations by adding hard-coded rules derived after extensive analysis of negation forms in Arabic. The algorithm was used along with ArSenL lexicon [120]. Also, the work in Reference [323] proposed a rule-based parser for document segmentation and then used an Arabic translated version of the

MPQA lexicon [390] to aggregate the sentiment scores over the document taking into consideration negations, intensifiers, and conjunctions. Similarly, Bayoudhi et al. [139] proposed to develop models for Arabic sentiment classification at the sub-sentential level that could provide very useful trends for information retrieval and extraction applications, Question Answering systems, and summarization tasks. They started by (1) building a high-coverage sentiment lexicon with a semi-automatic approach; (2) creating a large multi-domain annotated sentiment corpus segmented into discourse segments to evaluate the sentiment approach; and (3) applying a lexicon-based approach with an aggregation model taking into account challenging linguistic phenomena such as negation and intensification. Obaidata et al. [69, 320] proposed a lexicon-based approach for aspect extraction and classification of sentiments toward those aspects. They experimented on a dataset HAAD [71]. The authors utilized HAAD to automatically generate an Arabic sentiment lexicon using the frequency of occurrence of the term with a specific polarity. When using the lexicon in a sentiment analysis task, words that were not in the lexicon were translated to English and sentiment scores were obtained from SentiWordNet. Matoui et al. [291] proposed a lexicon-based approach for sentiment analysis from dialectal Arabic text, mainly the Algerian dialect. The Algerian dialect includes a lot of code switching, specifically switching between Arabic and French. They showcase the challenges present in the Algerian dialect and develop techniques to overcome them. They also develop three sentiment lexicons manually. The first one is based on existing sentiment lexicon for Egyptian dialect where the authors only kept terms that are common with Algerian dialect. The second lexicon is a list of negation words that are frequently used in Algerian dialect, and the third one is a list of intensifiers also frequently used in Algerian dialect. They tested different configurations for their model. The first one works at the phrase level and computes similarity of a given comment with existing labeled phrases. The second configuration involves going to the word level after the application of their own developed parser for Algerian dialect to perform tokenization, stop words removal, and normalization. The tokens are then processed by a language detection and stemming module that identifies the language of the tokens. For Arabic tokens, stemming was applied while tokens in other languages were first translated to Arabic and then stemmed. Stems were matched against the developed sentiment lexicons to compute a text semantic orientation score. They collect and manually annotate for polarity a set of 7,698 Facebook comments that cover multiple topics and include MSA and Algerian dialect. They achieved best accuracy of 79.13% when combining the two system's configurations. Recently, in SemEval 2017 Task 4 [347], Mulki et al. [311] applied an unsupervised lexicon-based approach for Arabic message polarity classification in subtask A. As sentiment lexicons, they have utilized NileULex [179], Arabic Emotion lexicon for Emojis, and Arabic Hashtag Lexicon for MSA and DA [301, 356]. They have also built two manually annotated sentiment lexicons to cover Levant and Gulf dialects. Specific efforts toward the Emirati dialect were also seen in Reference [76]. In Reference [245], Htait et al. also implemented an unsupervised sentiment classification model using word embeddings and sentiment lexicon extracted from annotated Arabic tweet corpora [314]. The sum of cosine similarity measures between the vector representation of the tweet and the words in the sentiment lexicon is used for message polarity classification. This technique placed the Htait et al. team, LSIS, in fifth rank among eight participants in subtask A.

### 4.2 Feature Engineering "Supervised" Approaches

With the availability of NLP tools and several lexical resources, a wide range of features were made available through feature engineering to train supervised machine-learning models for opinion mining, achieving better performances compared to the lexicon-based models. The aim of extracting these features is to reflect the different aspects of the given text. Thus, these features vary in complexity depending on the information they convey. According to Reference [266], an optimal set of features for opinion mining is dataset dependent. They typically range from shallow features that capture the surface form of the text to syntactic features that capture the grammatical rules that govern the language construction to deeper semantic features that capture the underlying meanings of the different components of the text.

*Surface Features.* These mainly include word *n*-grams; sequences of *n* consecutive words. These features were extensively evaluated under different settings including different context lengths *n* and different feature representations: binary presence, term frequency (TF), and term frequency inverse document frequency (TFIDF) [176]. These features were used to train several machine-learning algorithms for classification, mainly SVM, Multinomial Naïve Bayes (MNB), Conditional Random Fields (CRF), Decision Trees, and *k*-Nearest Neighbors (*k*-NN). Overall, in some cases, SVM achieved better results [21, 48, 57, 58, 66, 90, 91, 100, 101, 110, 144, 145, 147, 148, 175, 194, 195, 205, 206, 218, 237, 307, 351, 363, 365], and in other cases, NB performed better [51, 115, 117, 191, 257, 286], especially in the case of unbalanced datasets such as in References [306, 308, 309]. Mostafa [304] claimed that the best classifier is dataset dependent. Ensemble classifiers achieved further improvements [39, 138, 141, 269, 321]. To handle the impact of out-of-vocabulary (OOV) words due to dialectal variations, a lexicon was manually created in Reference [173] to map dialectal words into their corresponding MSA forms. The use of this lexicon introduced a slight improvement. Azmi and Alzanin presented Aara' [117], a system for mining opinion polarity through the pool of comments that readers write anonymously at the Online edition of Saudi newspapers. N-grams were used as features with a Naïve Bayes classifier and the output consisted of four class labels: positive, negative, neutral, and mixed. For training, they manually marked the comments as belonging to one of the four categories. All the words in the documents of the training set were removed except those with explicit sentiment connotations. The system carried out polarity classification over informal colloquial Arabic that is unstructured and with a reasonable proportion of spelling errors. The result of testing the system showed a macro-averaged precision of 86.5%, while the macro-averaged F-score was 84.5%. The accuracy of the system was 82%. Al-Obaidi and Samawi [55] worked on developing sentiment classification model for DA specifically Jordanian and Saudi dialects. They contributed to the preprocessing part of the system by developing a tailored stop words list for the two dialects and a light stemmer specific for the dialects. They experimented with different classification techniques along with bag-of-words (BoW) and *n*-gram features. Maximum Entropy performed best with trigrams. Results in these articles indicate that there is no consensus regarding the best experimental setup (length of *n*-gram or representation) and that results are corpus dependent. This is expected given the simplicity of these features, which does not match the complexity of the task. According to Halees [187], word-embeddings representation performed better in terms of accuracy on four different datasets compared to when using *n*-grams with the same classification model. Badaro et al. [121] also found word embeddings extracted from AraVec [371] to perform better than other features with ensemble classification. The same conclusions were drawn in References [107, 108]. Similarly, Al-Azani et al. [40] observed that simpler models consisting of word embeddings and emojis performed better than when using typical *n*-grams. El Razzaz et al. [207] evaluated the use of Arabic word embeddings. Additional surface features, also referred to as stylistic features, were extracted to indicate lexical and structural style markers. Character *n*-grams improved the classification performance significantly when combined with word *n*-grams to classify political articles using SVM and *k*-NN [171] and to classify newspaper articles [32]. Other stylistic features, including digit *n*-grams, word length distributions, vocabulary richness measure, function words, and punctuation, were also combined with word *n*-grams to train an SVM classifier and performed well after applying the Entropy-Weighted Genetic Algorithm (EWGA) selection method [2]. Recently, TwStAR [311] tried using unigrams, bigrams, and trigrams with SVM to perform sentiment classification on Arabic Twitter data in SemEval 2017 Task 4. They achieved rank 7 among eight participating teams in subtask A.

*Syntactic Features.* These are used to reflect the structure of the text and understand how the words function and combine together to express meaning. Knowing that Arabic is a morphologically rich and complex language, it is of high importance to incorporate syntactic and morphological information of the language into the sentiment models. One of the earliest grammatical approaches proposed to generalize verbal and nominal phrases into one form based on "actors" and "actions" and then to train SVM using the following features: actors, actions, adjectives, nouns, syntactic type of sentence, conjunction with previous sentence, and word sentiment polarity [212]. Most of this information was manually labeled due to the lack of Arabic NLP tools at that time. Results are considered a "proof of concept" of the importance of incorporating syntactic information into the model. The recent establishment of advanced Arabic NLP tools and resources allowed the automatic extraction of syntactic and morphological features, which are used to mitigate the impact of complex morphology on sentiment. Examples of such resources include the ATB [283], the SAMA [284], MADAMIRA [326], and many other tools mentioned in Section 2.2. For instance, adding word-level inflectional morphological features (gender, number, voice, ...) to basic surface features improved the performance of sentiment classification in MSA data [14], whereas they resulted in performance degradation when applied to Twitter data [340]. This is mainly because most Arabic NLP tools are trained on MSA data, whereas tweets generally contain significant amounts of dialects and misspellings, thus resulting in error-prone features. One way to alleviate the complexity of concatenative morphology in Arabic words was to represent words by stems or lemmas. Stemming was widely used as a preprocessing step, where all clitics and suffixes were chopped off from the base words, whereas lemmatization refers to selecting one word form to represent a set of words that are related by inflectional morphology [232]. It has become conventional to extract stem, root, or lemma *n*-grams, instead of raw word *n*-grams, and use them to train sentiment classifiers [6, 7, 10, 14, 19, 26, 42, 50, 66, 155–157, 172, 322]. Both forms provided the capability of generalizing to new unseen words that are morphologically related to words in the training data. However, it was not clear which form performed better. Furthermore, the impact of augmenting lemma *n*-grams with POS tags was measured for both tasks of subjectivity analysis and sentiment analysis. It was found that POS tags do not provide further improvements for sentiment analysis, whereas they do slightly improve subjectivity analysis [10]. Abd-Elhamid et al. [3] used rules and syntactic features for sentiment classification on Arabic text using decision trees. Al-Smadi et al. [67, 72] presented an aspect-based sentiment analysis approach using a feature engineering approach. As features, they considered POS tags, named entity recognition and *n*-grams with *n* varying between 1 and 3. They experimented with different classifiers: CRF, *k*-NN, Decision Trees, and Naïve Bayes. They extracted 2,265 news posts from Al Arabiya and Al Jazeera. The dataset was annotated by three Arabic native speakers. The annotators had to extract aspect terms and had to assess their polarity: positive, negative, or neutral. Preprocessing included tokenization, segmentation, stemming, POS tagging, punctuation and stop words removal, and normalization. For the task of extracting aspects, decision trees performed best with F1-measure of 81.7% while for the task of sentiment classification CRF performed best with an accuracy of 86.5%. CRF were also utilized successfully in References [381, 382]. *k*-NN performed best when used to perform multi-way classification in a set of cascaded classifiers [36, 319]. In References [355], the authors presented an approach to extract and classify opinion in micro-blogs. Their approach is based mainly on linguistic features extracted from Kuwaiti dialect and employed with a SVM classifier. They tested their approach on a corpus of 340,000 tweets about "interrogation of ministers by the National Assembly of Kuwait" during the past two years. Tweets were collected automatically by a module developed in java. The corpus had been manually annotated by three Kuwaiti dialect native speakers. An average value of 76% and 61% were obtained for precision and recall, respectively, in terms of sentiment classification.

*Semantic Features.* Many lexical resources were developed to reflect the semantic aspects of words and phrases in Arabic. Examples include the Arabic WordNet [143], an equivalent of English WordNet [216], which groups words into sets of synonyms (synsets) that are also enriched with semantic relations such as hyponymy, synonymy, and antonymy. Additionally, several Arabic sentiment lexica have been developed using automatic, semi-automatic, and manual approaches, as described in Subsection 3.1. The development of these resources enabled the extraction of more complex sets of engineered features that reflect surface, syntactic, and semantic aspects of the texts. Scores from sentiment lexica were frequently used to represent *n*-gram features, replacing presence and TFIDF scores [46, 88, 102, 119, 126, 217]. Sentence-level sentiment features were also derived by aggregating word scores through averaging or summation or by generating binary features indicating presence of positive and negative terms in the text. These features were combined with surface and syntactic features to train sentiment classifiers that performed well [4, 10, 12, 14, 120, 201, 242, 310, 372]. In References [83, 186], ontology was used to derive features for sentiment analysis from Arabic text reviews. To extract the ontologies present in a review, the author utilized ConceptNet [373] and WordNet [216]. The ontology allows the system to detect the sentiment for each element in the ontology. The technique was applied on two different datasets, one for hotels and one for books, each consisting of 2,000 reviews equally split between positive and negative. An average accuracy of 79% was claimed to be achieved. Ontologies were also used in References [93, 103, 135, 280, 377]. Entries of the ArSeLEX lexicon were used along with a set of linguistically and syntactically motivated features, including contextual intensifiers, contextual shifters, and negation particles, to train a SVM sentiment classifier and achieved high performances on small Twitter, comments, and reviews datasets, written in both MSA and Egyptian dialect [250]. Refaee and Rieser [338, 339] tested the efficiency of using emoticons for automatically classifying tweets into subjective or objective and also detecting whether the tweet is positive or negative. They found that emoticons perform well for detecting subjectivity; however, they perform poorly in distinguishing polarity. The same analysis of emoticons was suggested in Reference [56]. Rizkallah et al. [345] addressed the problem of sentiment analysis on Twitter data by comparing two different approaches: The first consisted of applying feature extraction directly on Twitter data in its dialectal form, and the second consisted of first translating the dialectal Arabic to MSA and then applying sentiment analysis model. The authors annotated manually a set of 2,010 tweets written in Saudi dialects into four labels: positive, negative, neutral, and mixed. Based on their results, translating to MSA seemed to improve the overall performance of sentiment classification with an accuracy of 76.2% when using logistic regression. In Reference [43], Al-Harbi et al. also worked on developing a sentiment analysis model for Saudi dialect. They studied the effect of different preprocessing techniques on the performance of sentiment classification. They used a dataset of 5,484 Saudi tweets annotated for sentiment. The preprocessing included no stemming, stemming, light stemming, or replacing Saudi dialect terms with their corresponding MSA term. Using the latter resulted in best performance along with the *k*-NN as classification technique. Mustafa et al. [312] proposed an improvement over the typical BoW model for sentiment analysis by incorporating different feature sets and performing cascaded analysis that contains lexical analysis, morphological analysis, and semantic analysis. AWN was used to extract semantic relations between terms in the dataset. Moreover, specific feature extraction components were integrated to account for the linguistic characteristics of Arabic. Emoticons and smileys were as well extracted to reflect the nature of the social media content. An average F-measure of 89% was achieved with a claimed significant improvement compared to BoW. El-Naggar et al. [192] presented a hybrid approach for sentiment analysis of MSA and Egyptian DA using verbal and non-verbal cues in the form of text and emojis. Using Plutchik's Wheel of Emotions, an emoji lexicon was created as a resource for non-verbal emotion classifications. Their feature set consisted of unigrams and bigrams with a minimum frequency of five, sentiment score derived based on different syntactic rules, emotion labels, number of sentiment and emotion tokens, presence of negations, and total number of tokens in the tweet. SVM performed best when used with bagging. The authors achieved an accuracy of 90%. CLASENTI [234] includes a two-stage machine-learning sentiment analysis. First, full-corpus (i.e., trained on all the annotated dataset) models classify the incoming text polarity, domains, dialects, and linguistic issues. Second, class-specific models are trained on filtered subsets of the corpus according to the performances of the full-corpus models. Moreover, a set of hand-crafted features that proved successful in English [128] was adopted in Arabic [129]. This feature set consisted of character and stem/lemma *n*-grams, counts of punctuation marks, elongated words, negated contexts, positive and negative emoticons, POS tags, and positive and negative words found in MSA and dialectal sentiment lexica. These features were used to train the model submitted by the OMAM team [130] to SemEval-2017 Task 4 on sentiment analysis on an Arabic Twitter dataset. OMAM ranked first in subtasks C (*3-way topic-based sentiment classification*) and E (*3-way tweet quantification*). The NileTMRG team [182] trained a Naïve Bayes classifier using the different features described above with additional Twitter-specific features such as whether the tweet starts with a hashtag [182]. This system ranked first in subtask A with an average recall of 58.3%. The same team implemented an ensemble system for subtasks B and D, consisting of a convolutional neural network (CNN), a multi-layer perceptron (MLP), and a logistic regression (LR) predictor. The input to the CNN classifier was an aggregation of the embeddings of each tweet's words, where these embeddings are obtained by training Word2Vec [293] using a Arabic Twitter dataset of four million tweets. The other classifiers were trained using a set of hand-crafted features including: a BoW representation of the tweet, the sentiment of the tweet as predicted by the team's system for subtask A, the number of positive/negative terms, the position of the target topic in the tweet, the presence of positive/negative emoticons, the presence of positive/negative terms around the target topic, and the number of positive/negative terms in the first and second halves of the tweet. This system also ranked first for both subtasks.

Other participants tried combining features from different approaches to improve the accuracy of their systems. Jabreel and Moreno [258] trained SVM with a rich set of surface, syntactic, and semantic features, in addition to tweets embeddings generated by summing the word embeddings of their constituent words. SiTAKA was ranked second in subtask A (*sentiment classification*). The INGEOTEC team used an ensemble classification system and ranked fourth in subtask A [294]. In this system, the output of a generic sentiment classification (B4MSA) system [378] was combined using the EvoDAG Genetic programming system [221, 222].

*Aspect-based Sentiment Analysis.* Ismail et al. [254] proposed an approach for performing aspect-based sentiment analysis on 500 Arabic movie reviews, 1,000 restaurant reviews, and 500 reviews of mixed domains. They started by manually annotating a set of Arabic syntactic patterns and roots with sentiment binary scores. They used the manually annotated lexicon to tag the words in the corpus. They also identified intensifiers and negations. They formulated a set of patterns to detect automatically aspects. They assigned for each aspect the sum of sentiment scores of the words that describe it. They claimed to achieve an average accuracy of 85.9% in terms of detecting the correct orientation of a given aspect. They also achieved an average F1 measure of 79.1% in terms of accurately extracting aspects. To enable aspect-based sentiment analysis from Arabic texts, Mataoui et al. [290] proposed a syntax-based approach for aspect detection from Arabic reviews extracted from Trip Advisor and Souq.com. Ibrahim and Salim [251] developed aspect-based sentiment analysis system that takes as inputs an Arabic tweet and generates as outputs the sentiment of the tweet and the different aspects or entities described in the tweet. Their work targeted DA. Extraction of entities was based on a frequent pattern-mining approach. For predicting sentiment, a set of features were generated including POS tags, *n*-grams, and polarity scores from a sentiment lexicon. In References [68, 70], Al-Smadi et al. compared the performance of RNN to SVM when using a set of surface, syntactic, and semantic features. The authors claimed that SVM performed better in terms of accuracy, but RNN had a faster training and testing time in both prediction of aspects and classification of sentiment toward those aspects. Farra and Mckeown [213] presented a system that is applied to complex posts written in response to Arabic newspaper articles. Their goal was to identify important entity targets within the post along with the polarity expressed about each target. They claimed to achieve significant improvements over multiple baselines, demonstrating that the use of specific morphological representations improves the performance of identifying both important targets and their sentiment and that the use of distributional semantic clusters further boosts performances for these representations, especially when richer linguistic resources were not available. Zarra et al. [395] focused on Maghrebi DA. On a corpus extracted from different Facebook pages, the authors implemented a supervised approach to extract the sentiments and an unsupervised approach to extract topic. Then, they proposed a semi-supervised approach that combines the topic and the sentiment in a single model to assign each topic to a specific sentiment.

### 4.3 Deep "Supervised" Approaches

Currently, state-of-the-art artificial intelligence (AI) systems rely on deep learning techniques and have achieved immense successes in many domains, especially computer vision and NLP. In NLP applications, deep learning benefited from the invent of word embeddings; distributional vector representations that encode syntactic and semantic properties of the words into low-dimensional and dense vectors. Examples of word-embedding models include CBOW and Skip-gram models in word2vec [293] and GloVe [327]. Word embeddings proved successful in several NLP applications such as POS tagging, chunking and parsing, word-sense disambiguation, named entity extraction, and sentiment target entities [35, 160, 188, 375]. One of the main advantages of deep learning models lies in their ability to perform semantic composition—generating a vector representation for text units by combining their finer-grained constituents or entities—efficiently and in a low-dimensional space. Consequently, advanced deep neural network architectures have been successfully applied for English sentiment analysis, such as recursive neural networks [370], deep convolutional networks [260], Gated Recurrent Neural Networks (GRNN) [376], dynamic memory networks [278], and the human reading for sentiment (HRS) framework [133].

Following their success in English, deep learning models were first used for sentiment analysis in Arabic in Reference [64], where several deep learning models, including Deep Belief Networks (DBN), Deep Neural Networks (DNN), and Deep Auto Encoders (DAE), were trained using word *n*-grams. Although these models proved better than several SVM sentiment classifiers such as in Reference [5], they still suffer from the sparse input features, indicating the importance of word embeddings. Al-Sallab et al. [63] analyzed and identified several limitations that might prevent deep learning models to achieve high performances in Arabic similarly to what they did in English. These limitations are mainly related to lexical sparsity, which is due to Arabic-rich morphology and a complex concatenative system, as well as the usage of non-standardized dialects, leading to large numbers of OOV tokens. Also, using traditional word embeddings is sub-optimal for sentiment analysis, since these vectors do not capture sentiment properties of the words. These limitations were addressed in the AROMA model [63], where the performance of the Recursive Auto Encoder (RAE), as used in English [369], was significantly improved by using newly proposed word sentiment embeddings as input features and by applying the recursion to morphologically tokenized text (separating clitics from words according to Reference [230]) following the path of phrase structure parse trees. The first Arabic Sentiment Treebank (ArSenTB) was developed in Reference [132] and consists of parse trees, where each node of the tree (corresponding to a word, phrase, or full text) is associated with a sentiment label. ArSenTB allowed us to train Recursive Neural Tensor Networks (RNTN) that can predict the overall sentiment by using the intermediate sentiment labels at the internal nodes, following the idea proposed in English [370]. Furthermore, this model was enriched with morphological features including stems, lemmas, and roots to overcome the lexical sparsity and ambiguity issue. These features, especially stems, achieved significant performance improvements on data containing a mixture of MSA and DA. CNNs have also been utilized for Arabic sentiment classification using word embeddings [86, 162, 224]. Alayba et al. [84, 85] applied sentiment analysis on a health dataset using deep learning models specifically deep neural networks, CNNs, and LSTM networks. The features they used were characters, character *n*-gram, and words. Each feature was represented as an embedding vector based on different sentiment analysis levels. Each feature was tested on its own. The input data wre fed to CNN followed by a max pooling layer. The output vectors of the max pooling layer were then used as input to LSTM networks to measure the long-term dependencies of feature sequences. The output vectors of the LSTMs were concatenated, and an activation function was applied to generate the final output: either positive or negative. The deep neural network accuracy reached 85%, while the CNN accuracy was better, reaching 90%. Ruder et al. [348] presented a hierarchical bidirectional LSTM network for performing aspect-based sentiment analysis on a multilingual dataset. They utilized word embeddings to represent terms and each aspect was represented by an entity and an attribute. Word embeddings were fed into a sentence-level bidirectional LSTM. Final states of forward and backward LSTM were concatenated together with the aspect embedding and fed into a bidirectional review-level LSTM. At every time step, the output of the forward and backward LSTM were concatenated and fed into a final layer, which outputs a probability distribution over sentiments. They evaluated their model on SemEval 2016 dataset [328] that consists of 11 domain-language datasets containing 300–400 reviews with 1,250 to 6,000 sentences. The LSTMs had one layer and an output size of 200 dimensions. The authors used 300-dimensional word embeddings. For Arabic, they learned embeddings using the Leipzig Corpora Collection.[4] They were able to achieve an accuracy of 82.9% for Arabic. Ruder et al. [349] applied a similar approach but with CNNs for both aspect extraction and aspect-based sentiment analysis on the SemEval 2016 dataset and achieved top ranks in multiple languages. In Reference [392], Yu and Al Baadani developed a multilayer perceptron (MLP) sentiment classification model for DA. The data preprocessing consisted of two phases: (1) a segmentation phase, where a separation among MSA, DA, and non-Arabic was performed; and (2) a refinement phase that included removing elongation effects, normalization and removing diacritics. Features included negation and intensification detection, polarity score assignment for terms, and polarity tag assignment for emoticons based on their developed lexicons. As a dataset, the authors collected comments from several social media platforms such as Google Plus, AreebaAreeba, Facebook, Youtube, Twitter, Yahoo News, and WeChat Moments. They annotated the comments into five sentiment labels varying from highly positive to highly negative. The annotated dataset consists of 14,000 comments. They used an ensemble of three-layer MLP networks consisting of one hidden layer in addition to input and output layers. The input for each MLP is a random subspace of their proposed features. Using 10-fold cross validation, they were able to achieve an average accuracy of 89.75%. Al-Azani and El-Alfy [37] tested LSTM models unidirectionally and bidirectionally and GRNN. Bidirectional LSTM outperformed typical classification

[4]http://corpora2.informatik.uni-leipzig.de/download.html.

techniques when emojis were used as input to the network. In Reference [38], Al-Azani and El-Alfy evaluated different deep neural network architectures, including LSTM, GRNN, and CNN, for sentiment classification using as input word embeddings. LSTM combined with CNN performed best on two different datasets. LSTM was also combined with CRF successfully in Reference [73] for aspect-based sentiment classification. In Reference [136], Barhoumi et al. utilized Doc2Vec to generate paragraph embeddings for Arabic texts such that the vector representation would be used as input feature for a classifier. To generate the embeddings, they utilized the sentiment annotated corpus LABR [110]. MLP and logistic regression were used as classification techniques. They claimed that using light stemming as preprocessing before computing the vector representations improved the performance with MLP. Doc2vec was also employed by Abduallah and Shaikh [23]. The authors participated in SemEval 2018 Task 1 Affect in tweets [296]. They applied the same model across all subtasks for both English and Arabic datasets. They used doc2vec and word2vec to generate word embeddings that were appended to a set of psycholinguistic features. They used a fully connected neural network architecture to train their model. They ranked fourth in both sentiment-related tasks.

We can notice that the popularity of deep learning models for Arabic SA is increasing, where several participants at SemEval 2017 Task 4 [347] employed deep learning models for sentiment classification. For instance, the ELiRF-UPV team [220] proposed a neural network architecture that consists of CNN, MLP, and bidirectional long short term memory (BLSTM) recurrent networks. The first layer of the CNN consists of a unidimensional convolutional layer that allows extracting spatial relations among the words in a tweet. In some subtasks and after the convolutional layer, a down-sampling process is applied using a max pooling layer. The output of the convolutional layer (32–256 neurons) is fed to a BLSTM, which performs semantic composition and generates an output representation that is fed to a fully connected MLP consisting of one to three hidden layers (depending on the subtask). A softmax function is then used to estimate the probability of each class. The input to this system was an aggregation of out-domain word embeddings learned from Wikipedia Arabic articles; in-domain word embeddings learned from the corpus provided by SemEval organizers; and a one-hot vector representing word sentiment polarity derived from NRC lexicon [299, 300]. This system ranked third in subtasks A and D and second in subtasks B, C, and E. One possible interpretation for NileTMRG achieving better results than EliRF-UPV is that the former trained word embeddings on a much larger corpus retrieved from Twitter, which will more likely resemble SemEval dataset, while EliRF-UPV trained word embeddings on Wikipedia, which is somehow different than Twitter data. Also, the results achieved by NileTMRG indicates that while deep learning represents the current state of the art for Arabic sentiment analysis, the performance can further improve when integrating surface, syntactic, and semantic features.

While most of the models described above focused on developing Arabic-specific opinion-mining models, several researchers explored translating Arabic text to English and then using state-of-the-art English sentiment classification systems, as in References [111, 137, 341, 350], or considered the outputs of different sentiment classification models for the same sentence in different languages for improved performance, such as in Reference [134]. Reference [332] evaluated a state-of-the-art English sentiment system, SentiStrength [379], by applying it directly on 11 different Arabic texts and claimed to achieve an average F1 of 68%. Similarly, Reference [268] compared two available online tools for sentiment analysis using a collected dataset of 1,000 Facebook comments and tweets annotated for sentiment. They compared the performance of SentiStrength and SocialMention and claimed that SentiStrength performed better.

In summary, we present in Tables 3 and 4 a summary of the opinion models discussed sampled from the different approaches described. The different opinion models categories presented in the above sections vary in their complexity in terms of how much information they use to infer sentiment from given text, resulting in different degrees of success. This also affects the complexity of training and using model. For instance, while lexicon-based approaches are simple and not highly accurate, they are very fast and light in terms of software requirements, which enables faster responses and model updates and easier integration into mobile applications [119]. However, while feature engineering approaches proved more successful, they require excessive effort and time to create immense and sparse feature matrices, with many features possibly not necessarily existing in new texts. This imposes time limitations, especially when building new models or updating existing ones. These challenges related to language modeling and sparsity caused by Arabic complex morphology are addressed with deep learning models. While no consensus is reached in terms of the order of *n*-grams that should be used for feature-based approaches, deep learning models overcome this challenge by using word embeddings and having a complete embedded sentence representation. Moreover, deep learning models allow encapsulating semantic knowledge and sentence parsing structures in a condensed vector representation of the sentence. However, deep learning models also require large-scale training data, and excessive time and computing resources, with thousands or millions of parameters being learned for the task. Nevertheless, they are preferred over feature engineering approaches, due to the dense and compact input features (embeddings), which is one of the main reasons behind the success of these models. We summarize, in Table 2, the challenges addressed (in bold) by the discussed categories of opinion-mining models along with their respective drawbacks. We can observe that deep learning models address many Arabic opinion mining challenges but at the expense of creating large-scale annotated corpora and of having enough computing power.

In general, developing opinion-mining algorithms and models that are highly accurate has been the goal of several researchers, but it is also very important to be able to integrate these algorithms and models into real-world applications. Opinion-mining applications are discussed next in Section 5 highlighting the usage of opinion mining in several sectors.

## 5 OPINION-MINING APPLICATIONS

Sentiment analysis applications evolved from being isolated applications that analyze sentences for subjectivity to becoming vital entities in key sectors such as politics, healthcare, marketing, finance, services, and education. Applications that have sentiment analysis at their core are continually emerging and are targeting all the above-mentioned sectors. However, only few ones rely on Arabic sentiment analysis. In what follows, we present an overview of the most relevant work on Arabic sentiment analysis in each sector.

*Politics.* Opinions that are shared on social media and blogging sites present valuable information that can be used for political purposes: to alert political leaders about potential problems or threats, to get a sense of how much a certain policy is being perceived by the public, to calculate a popularity index that can be used for elections [316], to know the public emotional status (angry, disgusted, or happy), and many others. Determining the opinion holder [193] helps in developing such systems. Most recently, the authors of Reference [317] implemented a system that, given a political figure, tracks the corresponding opinions presented on the web and presents a summarized report of that figure to monitor their political standing. Other relevant applications include the following: Reference [198] focused on the effect of sentiment during the 2012 presidential elections in Egypt, References [24, 276] proposed an automated tool to determine the political orientation of an Arabic article or comment, and Reference [238] analyzed the users' statuses on "Facebook" posts during the "Arabic Spring" era in Tunisia to get insights on users' behavior. Alsmearat et al. [106] were not able to find a correlation between the gender of a writer and the presence of opinion in Arabic text. In Reference [25], Abu-Jbara et al. utilized Arabic sentiment analysis to detect

Table 2. Arabic Opinion Mining Models: Addressed Challenges and Drawbacks

| OMA Technique | Addressed Challenges | Drawbacks |
|---|---|---|
| **Lexicon Based** Approaches | **Limited Availability of large-scale annotated Arabic corpora for sentiment:** No training needed. **Language sparsity:** Sentence represented by sentiment scores. | Requires external large-scale lexical resources. Context is not taken into consideration. Accuracy depends on quality and size of lexicon. |
| Supervised Approaches Using **Surface Features** | **Sentence representation:** *n*-gram representation of a sentence. **Context modeling:** using a high order of *n*-grams to include context. | Increased sparsity with the increase of order of *n*-grams. No unique solution for the order of *n*-grams, since it is corpus dependent. |
| Supervised Approaches Using **Syntactic Features** | **Ambiguity of language:** through lemmatization for example. **Sentence parsing:** understanding word relations. | Not sufficient on their own. Limited availability of NLP tools for the different Arabic dialects. |
| Supervised Approaches Using **Semantic Features** | **Sentence Sentiment Extraction:** using sentiment lexicons. **Semantic relations across words:** using AWN for example. | Limited availability of large-scale semantic/sentiment lexicons or dictionaries such as AWN. |
| **Deep Learning Models** | **Language Modeling. Language Sparsity. Sentence Representation. Semantic Relations. Sentence Parsing.** | Limited availability of large-scale Arabic sentiment annotated corpora for learning accurate models. Limited size of Arabic sentiment treebank compared to English. Computationally expensive. |

subgroups in an online political debate. Given the opinions of discussants in a debate, discussants would belong to the same subgroup if they shared the same opinion about the same targets or topics.

*Healthcare*. Many people share their health-related data and experience on well-known blogs and on social media. People discuss their health issues, symptoms, diagnosis results, medication given, and their experiences when visiting the healthcare centers. It is many times very crucial to patients to know the experiences of other patients and, consequently, to take decisions of which healthcare center to visit or which medication to choose. This subject is discussed to great detail in a book by Khan et al. [267]. In Reference [84], the authors analyze the sentiments on health services from data collected on Arabic twitter.

*Marketing*. Since social media and the web in general are being extensively used as a platform of customer interaction, sentiment analysis has taken marketing to a whole new level. Companies recognized the importance of sentiment analysis in branding their products and, consequently, invested heavily in recommender systems and social/sentiment analysis tools. Opinion mining can improve the quality of recommendation of recommender systems that are only based on user-item ranking matrix such as those described in References [122–124]. In References [17, 21, 48],

Table 3. Summary of Arabic Opinion Mining Models

| Reference | Preprocessing | Features Used | Classifier | Sentiment Lexicon | Corpus Used for Testing | MSA? | DA? | Evaluation Metrics | Class Labels | Results |
|---|---|---|---|---|---|---|---|---|---|---|
| Azmi and Alzanin [117] | Removal of all words except those with sentiment connotation | N-grams | None | Naïve Bayes | Comments extracted from an Online edition of Saudi Newspapers | Yes | Yes | Precision, F1, Accuracy | Pos, Neg, Neu, Mixed | 86.5%, 84.5%, 82% |
| Abdulla et al. [18] | Dialect elimination, Stemming, Intensification detection, Negation handling | Sentiment lexicon scores | Rule-based, Unsupervised | Corpus-based sentiment lexicon | Maktoob Corpus [48], Twitter Corpus [19] | Yes | No | Accuracy | Pos, Neg, Neu | 74.6% 70.2% |
| Abdulla et al. [20] | Tokenization, Elongation correction, Normalization of Arabic characters, Light Stemming, Stop words removal, Intensification detection, Negation handling | Sentiment lexicon scores | Rule-based, Unsupervised | Corpus-based sentiment lexicon | Maktoob Corpus [48], Twitter Corpus [19] | Yes | Yes | Accuracy | Pos, Neg, Neu | 63.8% 70.1% |
| Al-Ayyoub et al. [33] | Removal of non-Arabic characters, Elongation correction, Spelling correction, Stop words removal, Stemming, Negation handling | Sentiment lexicon scores | Rule-based, Unsupervised | Expand seed list and translation from English sentiment lexicons | Own manually annotated Arabic tweets for sentiment | Yes | Yes | Accuracy | Pos, Neg, Neu | 86.9% |
| Duwairi et al. [170] | Tokenization, Stemming | Sentiment lexicon scores | Rule-based, Unsupervised | Translation of English Sentiment lexicon and Expansion using Arabic Thesauri | Own manually annotated Arabic tweets for sentiment | Yes | Yes | Precision, Recall, Accuracy | Pos, Neg | 70.0% 46.0% 46.0% |
| Oraby et al. [323] | Tokenization, Parse Tree, Negation Handling, Intensifiers handling, Conjunctions detection | Sentiment lexicon scores | Rule-based calculation of sentiment score per review | Translated version of MPQA | Opinion Corpus for Arabic (OCA) [351] | Yes | No | Absolute Error (AE) | Rating score between 1 and 10 | 2.3 |

(Continued)

Table 3. Continued

| Reference | Preprocessing | Features Used | Classifier | Sentiment Lexicon | Corpus Used for Testing | MSA? | DA? | Evaluation Metrics | Class Labels | Results |
|---|---|---|---|---|---|---|---|---|---|---|
| Aly and Atiya [110] | Tokenization | Surface Features: 1, 2 and 3 grams, TFIDF | SVM | None | LABR Unbalanced | Yes | No | Accuracy and F1 | Polarity and Rating Classification | Polarity: 91.0%, 90.1% Rating: 50.3%, 49.1% |
| Al Smadi et al. [67, 72] | Tokenization, segmentation, stemming, POS tagging, punctuation and stop words removal, normalization | POS tags, Named Entities, *n*-grams (*n* between 1 and 3) | CRF | None | Articles from Al Arabiya and Al Jazeera | Yes | No | Accuracy | Pos, Neg, Neu | 86.5% |
| Rushdi-Saleh et al. [351] | Tokenization, Stop words removal, Stemming, Filtering by token length | Surface Features: trigrams, TFIDF | SVM | None | OCA | Yes | No | Precision, Recall, Accuracy | Pos, Neg | 87.4% 95.2% 90.6% |
| Omar et al. [321] | Tokenization, Stop words removal, Stemming, Dialectal and slang words conversion to MSA | Surface Features: Unigrams, bigrams, TFIDF | Ensemble: stacking and logistic regression | None | Reviews from jeeran.com manually annotated | Yes | Yes | Macro F1 | Subjectivity classification and Pos, Neg | 97.5% 91.0% |
| Duwairi [173] | Tokenization, Stop words removal, Negation handling, Emoticons conversion to words, Replacing dialectal terms by MSA, Stemming | Surface Features: bag of words with binary representation | NB | None | Arabic tweets annotated using Crowdsourcing | Yes | Yes | Macro F1 | Pos, Neg, Neu | 87.6% |
| Abdul-Mageed et al. [10] | Tokenization, Lemmatization, Stemming, POS Tagging, Replacing low frequency words by UNIQUE | Sentiment lexicon binary features, Detect if MSA or dialect, Genre, User ID | SVM | Manually annotated lexicon of adjectives | Manually labeled data extracted from 4 different sources on the web | Yes | Yes | Accuracy | Subjectivity Classification, Polarity Classification | 72.5% to 95.8% 65.9% to 81.4% |
| Al Shboul et al. [66] | Tokenization, Stemming, Stop words removal | Bag of words | MNB | None | LABR | Yes | No | Average F1 | Rating classification from 1 to 5 | 42.8% |

Table 4. Summary of Arabic Opinion Mining Models

| Reference | Preprocessing | Features Used | Classifier | Sentiment Lexicon | Corpus Used for Testing | MSA? | DA? | Evaluation Metrics | Class Labels | Results |
|---|---|---|---|---|---|---|---|---|---|---|
| El Naggar et al. [192] | Tokenization, Normalization | unigrams, bigrams, rule-based sentiment score, Emotion labels, number of sentiment and emotion tokens, presence of negations, count of total number of tokens in a tweet | SVM with bagging | Emoji lexicon | Own annotated tweets | Yes | Yes | Accuracy | Pos, Neg, Neu | 90% |
| Badaro et al. [120] | Tokenization, Lemmatization, Stop words removal | Three sentiment scores per sentence | SVM | ArSenL | ATB | Yes | No | Average F1 | Subjectivity Classification, Polarity Classification, Pos, Neg | 72.3% 64.5% |
| Mourad and Darwish [310] | Tokenization, Stemming, POS Tagging | Stem, Stem-POS, Bigram stems, count of POS tags, Counts of stems belonging to strong or weak subjectivity classes, MSA or DA, tweet-specific features (presence of hashtags, user mentions, url, retweet), Presence of elongation, Presence of emoticons, Usage of decorating characters, Polarity from sentiment lexicon | NB | Combination of different lexicons | Manually annotated tweets | Yes | Yes | F1 | Subjectivity Classification (subj/obj), Polarity Classification (pos, neg) | 52.8%/71.0% 76.5%/66.4% |
| Ibrahim et al. [250] | Tokenization | Binary sentiment feature, TF, Polar word position, Negation, Intensifiers, question marks, supplication and wishful expressions, POS tags, Bigrams, extraction of conflicting bigrams polarity | SVM | ArSeLEX and idioms for Egyptian dialect | Manually annotated dataset of tweets and reviews | Yes | Yes | F1 | Pos, Neg | 95.8% |
| Hamdi et al. [234] | Tokenization, Stemming, stop words removal, and oversampling | *n*-grams (n from 1 to 3), TFIDF | MNB; Bernoulli Naïve Bayes (BNB); Logistic Regression (LR); SVM | Manually annotated lexicon | Manually annotated corpus | Yes | Yes | Accuracy and F1 | Pos, Neg, Neu | Up to 95% |

(Continued)

Table 4. Continued

| Reference | Preprocessing | Features Used | Classifier | Sentiment Lexicon | Corpus Used for Testing | MSA? | DA? | Evaluation Metrics | Class Labels | Results |
|---|---|---|---|---|---|---|---|---|---|---|
| Al Sallab et al. [63] | Tokenization, Lemmatization, Elongation adjustment, Replacement of Emoticons, Removal of URLs, hashtags and user mentions, Normalization | Word Embeddings, Sentiment Embeddings, Syntactic Parse Trees | Recursive Auto Encoder, Softmax layer | ArSenL | ATB, QALB, RR tweets | Yes | Yes | Accuracy, F1 | Pos, Neg | Per corpus: 86.5%/84.9% 79.2%/75.5% 6.9%/68.9% |
| Baly et al. [132] | Tokenization, Stemming, Lemmatization, Parse Trees | Morphological features: stems, lemmas, roots; Intermediate sentiment labels for phrases and words | RNTN | ArSenL, ArSenTB | QALB | Yes | Yes | Accuracy, F1 | Pos, Neg, Neu | 80.0% 79.0% |
| Gonzalez et al. [220] | Tokenization | Out-domain Word Embeddings (Wikipedia), In-domain word embeddings (SemEval Data), One hot encoding sentiment sequence vector | Stacking of CRNN, MLP, BLSTM. Softmax layer | NRC | SemEval 2017 Task 4-A | Yes | Yes | Recall, F1, Accuracy | Pos, Neg, Neu | 47.8% 46.7% 50.8% |
| El Beltagy et al. [182, 183] | Removal of URLs and Diacritics, Elongation adjustment, Replacement of positive and negative emoticons by love and anger | Word embeddings obtained by running Word2Vec on 4 million tweets dataset, Bag of words, Overall sentiment of tweet predicted by a proprietary system, Features related to count and position of positive and negative terms | Ensemble Classification using CNN, MLP and LR | Proprietary lexicon | SemEval 2017 Task 4-B | Yes | Yes | Recall, F1, Accuracy | Pos, Neg | 76.8% 76.7% 77.0% |

the authors analyzed reviews and comments collected from Yahoo!-Maktoob, an Arabic social networking website. They extracted different aspects important for marketers such as the reviews' length, the numbers of likes/dislikes, the polarity distribution, and the languages used. Wang et al. [387] developed a social data analytics (SDA) tool that run on top of the IBM BigInsights platform. SDA allows us to identify user characteristics like gender, locations, names, and hobbies; develop comprehensive user profiles across messages and sources; and associate profiles with expressions of sentiment, buzz, intent, and ownership around brands, products, and companies. It enables data analysts with little knowledge in information extraction and sentiment analysis to get results from social data quickly. The authors added support for Arabic by developing a sentiment analysis model using Arabic tweets extracted about three different topics, Egyptian Teleco, the Egyptian government, and Saudi employment. Furthermore, Hathlian and Hafezs [241] employed sentiment analysis on Arabic tweets to predict whether people are interested in a certain product or defined subject.

The interest in providing sentiment analysis for marketing purposes leads to the establishment of many companies that provide tools for Arabic sentiment analysis, such as Repustate[5] and LexisNexis.[6]

*Finance.* Sentiment Analysis is being used as a major factor in making financial decisions given the insights it gives on the subject matter under analysis [96]. Take, for instance, the insights sentiment analysis gives on the stock price movement [281]. Sentiment analysis is also used in measuring the mood of a given investor or the overall investing public, either bullish or bearish.[7] In Reference [41], the authors studied the impact of the Islamic holy month of Ramadan on Islamic Middle Eastern markets, where it was shown that there is always a positive increase in the stock market that can be attributed to the positive investor mood and sentiment during this month. Also in Reference [61], the authors suggested a trading strategy with Mubasher products, a leading stock analysis software provider in the Gulf region, using sentiment analysis from tweets. Al-Rubaiee et al. [235, 236] utilized Saudi Twitter posts to predict Saudi stock market. The authors studied the relationship between opinion from social media and the Saudi market index to help foreign investors gain an insight into the opinions of Saudi investors. Alkubaisi et al. [95] proposed a sentiment analysis model on Arabic tweets for stock market price prediction for Al-Marai dairy company. Alshahrani et al. [105] investigated the impact of sentiment in Arabic tweets on Saudi stock market indicators. They manually annotated a set of 114K tweets for sentiment, and they evaluated different correlation metrics with the price change in the Saudi stock market. In Reference [94], a Granger causality is found between the amount of sentiment tweets and the stock market price changes in the Arab world.

*Services Sector.* Ahmed et al. [29] present an application of sentiment analysis using natural language toolkit (NLTK) for measuring customer service representative (CSR) productivity in real estate call centers. The study describes in detail the decisions made, step by step, in building an Arabic system for evaluating and measuring productivity. The system includes transcription method, feature extraction, training process, and analysis. In Reference [333], Rahamatallah et al. presented a sentiment analysis system prototype to specifically analyze customer reviews of Sudanese telecommunication products. In References [98, 331], Arabic sentiment analysis was used to measure customer satisfaction toward telecommunication companies in Saudi Arabia. Similarly,

[5]https://www.repustate.com/sentiment-analysis/.
[6]https://www.lexisnexis.com/en-us/news-company-research/default.page.
[7]www.investorwords.com.

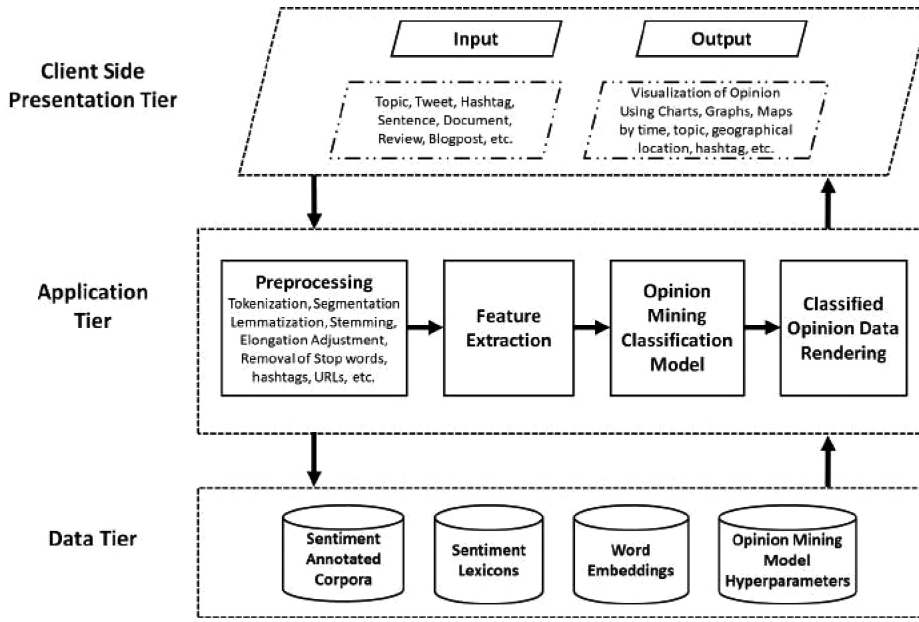


Fig. 2. Three-tier system architecture for arabic opinion mining.

Najadat et al. [315] also employed Arabic opinion mining to measure customer satisfaction toward Jordanian telecommunication companies.

*Education.* Arabic sentiment analysis has been used to evaluate students' satisfaction and experience at university by analyzing students' tweets [60]. Arabic opinion mining was also used to understand the opinion of students toward colleges [177]. El-Halees [185] presented a system to track changes of opinions expressed by Arab students about their courses to improve course evaluation in the future. Additionally, many studies are currently being conducted to evaluate the experience of students who enroll in or take online MOOCs.

After briefly describing typical existing applications for opinion mining, we present in the next section some future directions for opinion mining in Arabic, specifically a possible system architecture for deployment of an opinion-mining system along with pointers for currently used visualizations for English opinion mining as guide for Arabic. We also provide insights on remaining open challenges for OMA, and we outline some potential research directions as solution.

# 6 FUTURE DIRECTIONS FOR OPINION MINING IN ARABIC

## 6.1 System Architecture and Visualizations

After surveying the recent advancements of the different essential components for Arabic sentiment analysis, we present in this section the most common system architecture that allows these components to operate together. We also present, as an example, the architecture and technologies we used in implementing our system. Moreover, we briefly describe the latest efforts in visualizations for English SA. This description will serve as a guidance for Arabic opinion mining given that there were no specific efforts for visualizations in Arabic.

Most related work suggests a three-tier architecture for Arabic opinion-mining systems as presented in Figure 2. The client side, presentation tier, is a lightweight graphical user interface that is accessed by clients of the system. The access could be through a web page, a mobile application, or a desktop application. The graphical interface will include different input options that the client can select based on her needs. For example, the user can input a specific topic, a sentence, a hashtag, a document, and so on. Tweet sentiment viz[8] is an example of a Twitter sentiment analysis system where the user can provide a topic or a keyword and the system will retrieve the latest tweets relevant to the input and will display as output different visualizations for the analyzed tweet sentiments, such as heatmap, tag cloud, and graphs. The middle layer, i.e., application tier, constitutes the logic layer where all the heavy computations are performed: preprocessing, features extraction and computation, opinion classification, and output data rendering based on the classification. The application layer communicates with the client side layer to process the inputs and to display the outputs, and, with the third back-end layer, the data tier, to get the required data for each of the logic functions. The data tier includes all the data storage and the databases that the application tier utilizes to process the input and update the learning parameters, as, for example, the annotated sentiment corpora, word embeddings, sentiment lexicons, and the classification model parameters.

Although the architecture presented in Figure 2 is the norm when implementing opinion-mining systems, the technologies used greatly affect the performance of the system and its effectiveness. We give here an example of the technologies we used in implementing our system and that can be adopted as implementaton guidlines.

**Back-End.** The main API is built using an ASP.NET MVC 5 back-end by an SQL database. It relies on the open source library Hangfire to manage processing jobs. The back-end also relies on a few other external APIs including the Twitter API. The Twitter API is used in two ways:

- To search Twitter for tweets.
- To track selected news source twitter accounts to acquire article URLs for our Twitter-based crawler.

We also implemented on the back-end several sentiment analysis algorithms that provide tradeoff between speed and accuracy. For instance, we implemented a fast real-time algorithm that uses a stemmed version of ArSenL to predict sentiment. This implementation is written fully in C# and runs as part of the OMA Azure App Service. Consequently, it runs fast and does not require any additional resources. Another algorithm that we implemented is very accurate, but slower, and relies on compute intensive deep learning algorithms and a resource-intensive java library, namely MADAMIRA [326] (used for morphological analysis). To address the issue of compute-intensive resources and the issue of supporting multiple languages, we implemented plugin microservices. The plugin microservices fork the workflow to execute externally, while the regular workflow is running. The microservices were implemented using a NodeJS service and a queue. The queue holds all jobs that are to be processed. The NodeJS service deletes jobs off the queue, grabs the corresponding records from table storage, and then executes the corresponding sentiment analysis algorithm. After the execution is finished, it writes the sentiment labels to an Azure Table. Communication between OMA API's is done over HTTP requests.

**Front-End.** The front-end is built as a web application implemented as a single page application using HTML5, CSS3, and AngularJS. For the charts, we use a combination of D3 and Highcharts. We also use DataTable to display the sentiment processed records as a list, when more details are requested by the user. Finally, we use Bootstrap and Angular Material for the UI components and grid system of the OMA interface.

An important aspect of the presentation tier is the representation of the sentiment through informative visualizations. Users are interested in having informative and multidimensional

[8] https://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/.

visualizations of the sentiment around the topic they are querying. Thus, we present next a brief survey of visualizations used for English SA as guidance for application to Arabic.

The output of an opinion-mining system is the sentiment of a data feed. Given the sheer amount of such feeds, e.g., social media posts, that are multidimensional, the output is often provided as aggregate sentiment distributions over the entire data and its multiple dimensions. Such results are quite challenging for end users to perceive [325]. Visualizations can be devised to be explanatory or exploratory. Explanatory visualizations tend to be static and communicate inferences by visually highlighting and presenting the inferences directly [253]. However, exploratory visualizations, which involve interaction, aid the analysis process and let the users visually discover and make inferences, i.e., performing "Visual Analytics" [263].

Data feeds to opinion-mining systems often have multiple attributes, including geographical, demographical, and temporal variables. Thus, sentiment have been represented with respect to multiple attributes: location [259], time [168, 295, 318, 346], gender of writer [295], or multiple dimensions simultaneously [154, 288, 298, 318, 344, 385] such as location, time, gender, age, activity, and attitude.

Charts such as line, bar, scatter plots, connected scatter plots, heat map, and tree map are used to visualize sentiment of data feeds over attributes [272, 344]. Generally, connected scatter plots with custom shapes are limited and lack the ability to give context and details. These charts can plot a few number of attributes at a time, and they are to be perceived individually by the user. Other shapes of charts used include network flow chart, donut chart, multiple line chart, cluster map, and word cloud [360].

Besides the typical types of charts and plots, visualizations inspired from the real world were also adapted to represent sentiment and dashboards were introduced to link multiple visualizations: a river stream graph [282, 388], rectangular glyph [389], DNA visual shape [152], timeline arrangement with bubble charts [27], and, last, ring shapes [385].

When large amounts of data are visualized as static visualization, it tends to overload users. Thus visualizations are made dynamic with use of interactions to allow data exploration at the user's pace. Florence Ying Wang et al. in their developed system, SentiCompass [386] adopt an interactive visualization for exploring and comparing the sentiments of time-varying Twitter data. The interactive visualization involves zooming, dragging, and hovering to provide details on demand. Data exploration is obtained with use of time zoom buttons, window size slider, and time ring navigation controls. Torkildson et al. [380] visualize the sentiment of tweets during disasters, e.g., oil spills, using multiple stacked area charts with interactions such as zoom and pan. The goal of this visualization is to facilitate easy comparison of values and to provide details to the user with the help of interaction. Collaborative features such as example tweets are also adopted in the visualization to support analysis. More collaborative features such as shared views and annotations are planned for supporting communication between researchers. Visual Twitter Analytics (Vista) by Hoeber et al. [244] uses visual analytics approach to explore the temporally changing sentiment of tweets. The visual interface of Vista consists of a dashboard view with charts and other HTML form elements. Interactions such as zoom and filter are supported to enable data exploration. More details on demand are provided using the click interaction on the visual interface. This fosters analysis and helps the analysts drill down on data at lower levels.

### 6.2 Remaining Open Challenges

While good progress and results have been achieved for Arabic sentiment analysis, specifically for MSA, a lot remains to be performed for dialectal Arabic. In fact, each Arab country has its own dialect, but in general we refer to dialects per region such as, for example, Levantine, Gulf, Egyptian, and North African. Each region possesses different characteristics making sentiment analysis task even harder [31, 129]. For instance, the expression "يعطيك العافية" (yaEtiyk Al EaAfiyah) [231] is a positive complement in the Levantine dialect meaning "May God grant you health," while it has a negative connotation in the Moroccan and Tunisian dialects, "burn in fire" [189]. Moreover, Twitter data include a lot of neologisms, abbreviations, emojis, elongations, negations, intensifiers, polarity shifters, and sarcasm [261, 285, 334, 384]; thus, more sophisticated natural language processing tools are needed to detect these critical features for sentiment and to handle polarity shifts due to negations or sarcasm, for example. Reference [129] also suggests that each dialect requires a customized opinion-mining model. However, having an accurate opinion-mining model for each dialect is only possible if we have large-scale annotated sentiment corpora for each of the dialects. While some efforts have been invested to achieve this goal [131], much larger-scale annotations are required. In general, more efforts are needed to develop large-scale Arabic annotated data to enable training of robust sentiment classification models. We noticed that few of the annotated datasets and corpora are made publicly available. Releasing the data publicly tremendously helps in getting more researchers and scholars interested in working on opinion mining for Arabic. A unified platform where all resources are shared can further increase visibility. Moreover, on the semantic side and compared to English, Arabic is still considered a low-resource language. For instance, AWN [143] contains less than 30% synsets of those of EWN [216]. Thus work on extending AWN is of great importance for Arabic sentiment analysis and Arabic natural language processing in general.

In terms of opinion-mining models, deep learning approaches are currently the state of the art. Specifically for Arabic, these models can be further explored to enhance the accuracy of Arabic sentiment classification. Some of the open challenges for deep learning in Arabic are as follows: What is the best way to handle special syntactic and morphological complexity of Arabic? How should we handle unseen terms and slang language: MSA is a rich sparse language, and DA, with its own specificities, makes the Arabic language even more sparse and complex. How do we transfer learning from one dialect to another? In fact, one can find more data for the Egyptian dialect compared to other Arabic dialects. Hence, it is interesting to study how to successfully transfer the learning corresponding to the Egyptian dialect to other Arabic dialects that do not possess enough annotated data. Successful implementation of transfer learning would be a much cheaper solution and less time-consuming compared to creating annotated corpora. In terms of evaluating opinion-mining models, it is important to try to select a test set that is completely independent from the training set so that we can obtain results that are not biased or misleading as indicated in Reference [336].

In addition to developing resources and models, it is important as well to harvest these efforts in real-world applications using customized visualizations that target the Arabic audience. We observed that limited work describes applications for Arabic sentiment analysis compared to English. It is also important to note that assessing credibility of the text before taking into account its sentiment will help in accurately representing the sentiment toward a certain topic [82]. In terms of visualizations, existing work on English sentiment visualization can serve as a guide for developing customized Arabic centered visualizations.

## 7 CONCLUSION

In conclusion, we presented in this article an extensive literature review about opinion mining in Arabic. Unlike existing survey papers on Arabic sentiment analysis, we covered literature about all elements involved in a comprehensive opinion-mining system: tools, resources, approaches, applications. We discussed the latest in the field, including state-of-the-art deep learning models for sentiment analysis. We also presented a description of a system architecture for an Arabic opinion-mining system to guide people interested in implementing a real opinion-mining application. As future directions for OMA, a brief survey of visualizations used in implementations of English opinion-mining systems was discussed as a basis for developing Arabic-specific visualizations. We also provided some insights on open research challenges in OMA to help newly interested scholars, researchers, and students progress in their future studies in this field.

## REFERENCES

[1] [n. d.]. Dataset for Arabic Document Classification. Retrieved August 22, 2017 from http://diab.edublogs.org/dataset-for-arabic-document-classification/.

[2] Ahmed Abbasi, Hsinchun Chen, and Arab Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.* 26, 3 (2008), 12.

[3] Laila Abd-Elhamid, Doaa Elzanfaly, and Ahmed Sharaf Eldin. 2016. Feature-based sentiment analysis in online Arabic reviews. In *Proceedings of the 2016 11th International Conference on Computer Engineering & Systems (ICCES'16).* IEEE, 260–265.

[4] Laila Abd-Elhamid, Doaa Elzanfaly, and Ahmed Sharaf Eldin. 2018. Arabic feature-based level sentiment analysis using lexicon-based approach. *J. Fundam. Appl. Sci.* 10, 4S (2018), 143–148.

[5] Naglaa Abdelhade, Taysir Hassan A. Soliman, and Hosny M. Ibrahim. 2017. Detecting Twitter users' opinions of Arabic comments during various time episodes via deep neural network. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*. Springer, 232–246.

[6] Muhammad Abdul-Mageed. 2017. Modeling Arabic subjectivity and sentiment in lexical space. *Inf. Process. Manage.* 56, 2 (2017), 291–307.

[7] Muhammad Abdul-Mageed. 2017. Not all segments are created equal: Syntactically motivated sentiment analysis in lexical space. In *Proceedings of the 3rd Arabic Natural Language Processing Workshop*. 147–156.

[8] Muhammad Abdul-Mageed and Mona Diab. 2012. Toward building a large-scale Arabic sentiment lexicon. In *Proceedings of the 6th International Global WordNet Conference*. 18–22.

[9] Muhammad Abdul-Mageed, Mona Diab, and Sandra Kübler. 2013. Asma: A system for automatic segmentation and morpho-syntactic disambiguation of modern standard arabic. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'13).* 1–8.

[10] Muhammad Abdul-Mageed, Mona Diab, and Sandra Kübler. 2014. SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Comput. Speech Lang.* 28, 1 (2014), 20–37.

[11] Muhammad Abdul-Mageed and Mona T. Diab. 2011. Subjectivity and sentiment annotation of modern standard arabic newswire. In *Proceedings of the 5th Linguistic Annotation Workshop*. Association for Computational Linguistics, 110–118.

[12] Muhammad Abdul-Mageed and Mona T. Diab. 2012. AWATIF: A multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'12)*. 3907–3914.

[13] Muhammad Abdul-Mageed and Mona T. Diab. 2014. SANA: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14)*. 1162–1169.

[14] Muhammad Abdul-Mageed, Mona T. Diab, and Mohammed Korayem. 2011. Subjectivity and sentiment analysis of modern standard Arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2.* Association for Computational Linguistics, 587–591.

[15] Muhammad Abdul-Mageed and Mohammed Korayem. 2010. Automatic identification of subjectivity in morphologically rich languages: The case of Arabic. *Computational Approaches to Subjectivity and Sentiment Analysis* 2 (2010), 2–6.

[16] Muhammad Abdul-Mageed, Mohammed Korayem, and Ahmed YoussefAgha. 2011. "Yes we can?": Subjectivity Annotation and Tagging for the Health Domain. (2011), 666–671.

[17] N. A. Abdulla. 2014. *Towards Building a Sentiment Analysis Tool for Colloquial and Modern Standard Arabic Reviews.* Ph.D. Dissertation. Master's thesis. Computer Science Department, Jordan University of Science and Technology, Irbid, Jordan.

[18] Nawaf Abdulla, Salwa Mohammed, Mahmoud Al-Ayyoub, Mohammed Al-Kabi, et al. 2014. Automatic lexicon construction for Arabic sentiment analysis. In *Proceedings of the 2014 International Conference on Future Internet of Things and Cloud (FiCloud'14).* IEEE, 547–552.

[19] Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab, and Mahmoud Al-Ayyoub. 2013. Arabic sentiment analysis: Lexicon-based and corpus-based. In *Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT'13).* IEEE, 1–6.

[20] Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab, Mahmoud Al-Ayyoub, Mohammed N. Al-Kabi, and Saleh Al-rifai. 2014. Towards improving the lexicon-based approach for Arabic sentiment analysis. *Int. J. Inf. Technol. Web Eng.* 9, 3 (2014), 55–71.

[21] Nawaf A. Abdulla, Mahmoud Al-Ayyoub, and Mohammed Naji Al-Kabi. 2014. An extended analytical study of arabic sentiments. *Int. J. Big Data Intell.* 1, 1-2 (2014), 103–113.

[22] Malak Abdullah and Mirsad Hadzikadic. 2017. Sentiment analysis on arabic tweets: Challenges to dissecting the language. In *Proceedings of the International Conference on Social Computing and Social Media.* Springer, 191–202.

[23] Malak Abdullah and Samira Shaikh. 2018. TeamUNCC at SemEval-2018 task 1: Emotion detection in English and Arabic tweets using deep learning. In *Proceedings of the 12th International Workshop on Semantic Evaluation.* 350–357.

[24] Raddad Abooraig, Ahmed Alwajeeh, Mahmoud Al-Ayyoub, and Ismail Hmeidi. 2014. On the automatic categorization of arabic articles based on their political orientation. In *Proceedings of the 3rd International Conference on Informatics Engineering and Information Science (ICIEIS'14).*

[25] Amjad Abu-Jbara, Ben King, Mona Diab, and Dragomir Radev. 2013. Identifying opinion subgroups in arabic online discussions. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 829–835.

[26] Diab Abuaiadah, Dileep Rajendran, and Mustafa Jarrar. 2017. Clustering Arabic tweets for sentiment analysis. In *Proceedings of the 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA'17).* IEEE, 449–456.

[27] Brett Adams, Dinh Phung, and Svetha Venkatesh. 2011. Eventscapes: Visualizing events over time with emotive facets. In *Proceedings of the 19th ACM International Conference on Multimedia.* ACM, 1477–1480.

[28] Khurshid Ahmad, David Cheng, and Yousif Almas. 2007. Multi-lingual sentiment analysis of financial news streams. In *Proceedings of the 1st International Workshop on Grid Technology for Financial Modeling and Simulation*, Vol. 26. SISSA Medialab.

[29] Abdelrahman Ahmed, Sergio Toral, and Khaled Shaalan. 2016. Agent productivity measurement in call center using machine learning. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics.* Springer, 160–169.

[30] Nasser Ahmed. 2018. Large-scale Arabic sentiment corpus and lexicon building for concept-based sentiment analysis systems. Doctoral Dissertation, Hacettepe University. Retrieved from http://www.openaccess.hacettepe.edu.tr:8080/xmlui/handle/11655/4170.

[31] Soha Ahmed, Michel Pasquier, and Ghassan Qadah. 2013. Key issues in conducting sentiment analysis on Arabic social media text. In *Proceedings of the 2013 9th International Conference on Innovations in Information Technology (IIT'13).* IEEE, 72–77.

[32] Fawaz S. Al-Anzi and Dia AbuZeina. 2017. A micro-word based approach for Arabic sentiment analysis. In *Proceedings of the 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA'17).* IEEE, 910–914.

[33] Mahmoud Al-Ayyoub, Safa Bani Essa, and Izzat Alsmadi. 2015. Lexicon-based sentiment analysis of Arabic tweets. *Int. J. Soc. Netw. Min.* 2, 2 (2015), 101–114.

[34] Mahmoud Al-Ayyoub, Abed Allah Khamaiseh, Yaser Jararweh, and Mohammed N Al-Kabi. 2018. A comprehensive survey of arabic sentiment analysis. *Inf. Process. Manage.* 56, 2 (2018), 320–342.

[35] Mahmoud Al-Ayyoub, Aya Nuseir, Kholoud Alsmearat, Yaser Jararweh, and Brij Gupta. 2017. Deep learning for Arabic NLP: A survey. *J. Comput. Sci.* 26, 1 (2017), 522–531.

[36] Mahmoud Al-Ayyoub, Aya Nuseir, Ghassan Kanaan, and Riyad Al-Shalabi. 2016. Hierar-chical classifiers for multi-way sentiment analysis of arabic reviews. *Int. J. Adv. Comput. Sci. Appl.* 7, 2 (2016), 531–539.

[37] Sadam Al-Azani and El-Sayed El-Alfy. 2018. Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks. In *Proceedings of the 2018 International Conference on Computing Sciences and Engineering (ICCSE'18).* IEEE, 1–6.

[38] Sadam Al-Azani and El-Sayed M El-Alfy. 2017. Hybrid deep learning for sentiment polarity determination of arabic microblogs. In *Proceedings of the International Conference on Neural Information Processing*. Springer, 491–500.

[39] Sadam Al-Azani and El-Sayed M El-Alfy. 2017. Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short arabic text. *Proc. Comput. Sci.* 109 (2017), 359–366.

[40] Sadam Al-Azani and El-Sayed M El-Alfy. 2018. Combining emojis with Arabic textual features for sentiment classification. In *Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS'18).* IEEE, 139–144.

[41] Heitham Al-Hajieh, Keith Redhead, and Timothy Rodgers. 2011. Investor sentiment and calendar anomaly effects: A case study of the impact of ramadan on islamic middle eastern markets. *Res. Int. Bus. Finan.* 25, 3 (2011), 345–356.

[42] Omar Al-Harbi. 2017. Using objective words in the reviews to improve the colloquial arabic sentiment analysis. *Int. J. Nat. Lang. Comput.* 6, 3 (2017), 1–14.

[43] Waad A. Al-Harbi and Ahmed Emam. 2015. Effect of Saudi dialect preprocessing on Arabic sentiment analysis. *Int. J. Adv. Comput. Technol.* 8, 2 (2015), 91–99.

[44] Lamia Al-Horaibi and Muhammad Badruddin Khan. 2016. Sentiment analysis of Arabic tweets using semantic resources. *Int. J. Comput. Inf. Sci.* 12, 2 (2016), 149.

[45] Mohammed Al-Kabi, Mahmoud Al-Ayyoub, Izzat Alsmadi, and Heider Wahsheh. 2016. A prototype for a standard arabic sentiment analysis corpus.*Int. Arab J. Inf. Technol.* 13, 1A (2016), 163–170.

[46] Mohammed Al-Kabi, Noor M. Al-Qudah, Izzat Alsmadi, M. Dabour, and H. Wahsheh. 2013. Arabic/English sentiment analysis: An empirical study. In *Proceedings of the 4th International Conference on Information and Communication Systems (ICICS'13)*. 23–25.

[47] Mohammed Al-Kabi, Amal Gigieh, Izzat Alsmadi, Heider Wahsheh, and Mohamad Haidar. 2013. An opinion analysis tool for colloquial and standard Arabic. In *Proceedings of the 4th International Conference on Information and Communication Systems (ICICS'13)*. 23–25.

[48] Mohammed N. Al-Kabi, Nawaf A. Abdulla, and Mahmoud Al-Ayyoub. 2013. An analytical study of Arabic sentiments: Maktoob case study. In *Proceedings of the 2013 8th International Conference for Internet Technology and Secured Transactions (ICITST'13)*. IEEE, 89–94.

[49] Mohammed N. Al-Kabi, Areej A. Al-Qwaqenah, Amal H. Gigieh, Kholoud Alsmearat, Mahmoud Al-Ayyoub, and Izzat M. Alsmadi. 2016. Building a standard dataset for Arabie sentiment analysis: Identifying potential annotation pitfalls. In *Proceedings of the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA'16)*. IEEE, 1–6.

[50] Mohammed N. Al-Kabi, Izzat M. Alsmadi, Rawan T. Khasawneh, and Heider A. Wahsheh. 2016. Evaluating social context in Arabic opinion mining. *Int. Arab J. Inf. Technol.* 15, 6 (2016), 974–982.

[51] Mohammed N. Al-Kabi, Amal H. Gigieh, Izzat M. Alsmadi, Heider A. Wahsheh, and Mohamad M. Haidar. 2014. Opinion mining and analysis for Arabic language. *Int. J. Adv. Comput. Sc. Appl.* 5, 5 (2014), 181–195.

[52] Tareq Al-Moslmi, Mohammed Albared, Adel Al-Shabi, Nazlia Omar, and Salwani Abdullah. 2017. Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. *J. Inf. Sci.* 44, 3 (2017), 345–362.

[53] Tareq Al-Moslmi, Nazlia Omar, Salwani Abdullah, and Mohammed Albared. 2017. Approaches to cross-domain sentiment analysis: A systematic literature review. *IEEE Access* 5 (2017), 16173–16192.

[54] Ayesha Jumaa Salem Al Mukhaiti, Sanjeera Siddiqui, and Khaled Shaalan. 2017. Dataset built for Arabic sentiment analysis. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*. Springer, 406–416.

[55] Ahmed Y. Al-Obaidi and Venus W. Samawi. 2016. Opinion mining: Analysis of comments written in arabic colloquial. In *Proceedings of the World Congress on Engineering and Computer Science*, Vol. 1.

[56] Salha Al-Osaimi and Khan Muhammad Badruddin. 2014. Role of emotion icons in sentiment classification of Arabic tweets. In *Proceedings of the 6th International Conference on Management of Emergent Digital Ecosystems*. ACM, 167–171.

[57] Qasem A. Al-Radaideh and Ghufran Y. Al-Qudah. 2017. Application of rough set-based feature selection for Arabic sentiment analysis. *Cogn. Comput.* 9, 4 (2017), 436–445.

[58] Qasem A. Al-Radaideh and Laila M. Twaiq. 2014. Rough set theory for Arabic sentiment classification. In *Proceedings of the 2014 International Conference on Future Internet of Things and Cloud (FiCloud'14)*. IEEE, 559–564.

[59] Khalid Al-Rowaily, Muhammad Abulaish, Nur Al-Hasan Haldar, and Majed Al-Rubaian. 2015. BiSAL–A bilingual sentiment analysis lexicon to analyze dark web forums for cyber security. *Dig. Invest.* 14 (2015), 53–62.

[60] Hamed Al-Rubaiee, Renxi Qiu, Khalid Alomar, and Dayou Li. 2016. Sentiment analysis of Arabic tweets in e-learning. *J. Comput. Sci.* 12, 11 (Jan. 2016), 553–563. DOI: https://doi.org/10.3844/jcssp.2016.553.563

[61] Hamed Al-Rubaiee, Renxi Qiu, and Dayou Li. 2016. Identifying mubasher software products through sentiment analysis of Arabic tweets. In *Proceedings of the 2016 International Conference on Industrial Informatics and Computer Systems (CIICS'16)*. IEEE, 1–6.

[62] Rania Al-Sabbagh and Roxana Girju. 2012. A supervised POS tagger for written Arabic social networking corpora. In *Proceedings of the Konferenz zur Verarbeitung natürlicher Sprache*. 39–52.

[63] Ahmad Al-Sallab, Ramy Baly, Hazem Hajj, Khaled Bashir Shaban, Wassim El-Hajj, and Gilbert Badaro. 2017. AROMA: A recursive deep learning model for opinion mining in arabic as a low resource language. *ACM Transac. Asian Low-Res. Lang. Inf. Process.* 16, 4 (2017), 25.

[64] Ahmad A. Al Sallab, Ramy Baly, Gilbert Badaro, Hazem Hajj, Wassim El Hajj, and Khaled B. Shaban. 2015. Deep learning models for sentiment analysis in arabic. In *Proceedings of the Annual Meeting on Natural Language Processing (ANLP'15) Workshop*, Vol. 9.

[65] Huda Al-Sarhan, Majd Al-So'ud, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2016. Framework for affective news analysis of arabic news: 2014 gaza attacks case study. In *Proceedings of the 2016 7th International Conference on Information and Communication Systems (ICICS'16)*. IEEE, 327–332.

[66] Bashar Al Shboul, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2015. Multi-way sentiment classification of arabic reviews. In *Proceedings of the 2015 6th International Conference on Information and Communication Systems (ICICS'15)*. IEEE, 206–211.

[67] Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Huda Nimer Al-Sarhan, and Yaser Jararweh. 2016. An aspect-based sentiment analysis approach to evaluating arabic news affect on readers. *J. Univ. Comput. Sci.* 22, 5 (2016), 630–649.

[68] Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yaser Jararweh, and Omar Qawasmeh. 2018. Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Journal of Information Processing & Management* 56, 2 (2018), 308–319.

[69] Mohammad Al Smadi, Islam Obaidat, Mahmoud Al-Ayyoub, Rami Mohawesh, and Yaser Jararweh. 2016. Using enhanced lexicon-based approaches for the determination of aspect categories and their polarities in arabic reviews. *Int. J. Inf. Technol. Web Eng.* 11, 3 (2016), 15–31.

[70] Mohammad Al-Smadi, Omar Qawasmeh, Mahmoud Al-Ayyoub, Yaser Jararweh, and Brij Gupta. 2017. Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *J. Comput. Sci.* 27 (2017), 386–393.

[71] Mohammad Al-Smadi, Omar Qawasmeh, Bashar Talafha, and Muhannad Quwaider. 2015. Human annotated arabic dataset of book reviews for aspect based sentiment analysis. In *Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud'15)*. IEEE, 726–730.

[72] Mohammad Al-Smadi, Omar Qwasmeh, Bashar Talafha, Mahmoud Al-Ayyoub, Yaser Jararweh, and Elhadj Benkhelifa. 2016. An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study. In *Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST'16)*. IEEE, 98–103.

[73] Mohammad Al-Smadi, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2018. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int. J. Mach. Learn. Cybernet.* 55, 2 (2018), 1–13.

[74] Afnan A. Al-Subaihin, Hend S. Al-Khalifa, and AbdulMalik S. Al-Salman. 2011. A proposed sentiment analysis tool for modern arabic using human-based computing. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*. ACM, 543–546.

[75] Afnan S. Al-Subaihin and Hend S. Al-Khalifa. 2014. A system for sentiment analysis of colloquial Arabic using human computation. *Sci. World J.* 2014 (2014), 1–8.

[76] H. Al Suwaidi, Tariq Rahim Soomro, and K. Shaalan. 2016. Sentiment analysis for emiriti dialects in Twitter. *Sindh Univ. Res. J.* 48, 4 (2016).

[77] Nora Al-Twairesh, Hend Al-Khalifa, and AbdulMalik Al-Salman. 2014. Subjectivity and sentiment analysis of Arabic: Trends and challenges. In *Proceedings of the 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA'14)*. IEEE, 148–155.

[78] Nora Al-Twairesh, Hend Al-Khalifa, and AbdulMalik Al-Salman. 2015. Towards analyzing Saudi tweets. In *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15)*. IEEE, 114–117.

[79] Nora Al-Twairesh, Hend Al-Khalifa, AbdulMalik Al-Salman, and Yousef Al-Ohali. 2017. AraSenTi-Tweet: A corpus for Arabic sentiment analysis of Saudi tweets. *Proc. Comput. Sci.* 117 (2017), 63–72.

[80] Nora Al-Twairesh, Hend Al-Khalifa, and Abdulmalik AlSalman. 2016. AraSenTi: Large-scale twitter-specific arabic sentiment lexicons. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 697–705.

[81] Nora Al-Twairesh, Hend Al-Khalifa, AbdulMalik Alsalman, and Yousef Al-Ohali. 2018. Sentiment analysis of Arabic tweets: Feature engineering and a hybrid approach. https://arxiv.org/abs/1805.08533 (2018).

[82] Ayman Al Zaatari, Rim El Ballouli, Shady Elbassuoni, Wassim El-Hajj, Hazem M. Hajj, Khaled B. Shaban, Nizar Habash, and Emad Yahya. 2016. Arabic corpora for credibility analysis. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'16)*.

[83] Ahmed M. I. AlAsmar. 2016. *Feature Based Approach in Arabic Opinion Mining Using Ontology*. Ph.D. Dissertation. The Islamic University–Gaza.

[84] Abdulaziz M. Alayba, Vasile Palade, Matthew England, and Rahat Iqbal. 2017. Arabic language sentiment analysis on health services. In *Proceedings of the 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR'17)*. IEEE, 114–118.

[85] Abdulaziz M. Alayba, Vasile Palade, Matthew England, and Rahat Iqbal. 2018. A combined CNN and LSTM model for arabic sentiment analysis. In *Proceedings of the International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 179–191.

[86] Abdulaziz M. Alayba, Vasile Palade, Matthew England, and Rahat Iqbal. 2018. Improving sentiment analysis in Arabic using word representation. *arXiv preprint arXiv:1803.00124* (2018).

[87] Lamya Albraheem and Hend S. Al-Khalifa. 2012. Exploring the problems of sentiment analysis in informal Arabic. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*. ACM, 415–418.

[88] Haifa K. Aldayel and Aqil M. Azmi. 2016. Arabic tweets sentiment analysis–a hybrid scheme. *J. Inf. Sci.* 42, 6 (2016), 782–797.

[89] Samah Alhazmi, William Black, and John McNaught. 2013. Arabic SentiWordNet in relation to SentiWordNet 3.0. *IJCL. 2180* 1266, 4 (2013), 1.

[90] Sarah Alhumoud, Tarfa Albuhairi, and Wejdan Alohaideb. 2015. Hybrid sentiment analyser for Arabic tweets using R. In *Proceedings of the 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K'15)*. IEEE, 417–424.

[91] Sarah Alhumoud, Tarfa Albuhairi, and Mawaheb Altuwaijri. 2015. Arabic sentiment analysis using WEKA a hybrid learning approach. In *Proceedings of the 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K'15)*. IEEE, 402–408.

[92] Sarah O. Alhumoud, Mawaheb I. Altuwaijri, Tarfa M. Albuhairi, and Wejdan M. Alohaideb. 2015. Survey on Arabic sentiment analysis in twitter. *Int. Sci. Ind.* 9, 1 (2015), 364–368.

[93] Abdullah M. Alkadri and Abeer M. ElKorany. 2016. Semantic feature based arabic opinion mining using ontology. *Int. J. Adv. Comput. Sci. Appl.* 7, 5 (2016), 577–583.

[94] Khalid AlKhatib, Abdullateef Rabab'ah, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2016. On the use of arabic tweets to predict stock market changes in the arab world. *Int. J. Adv. Comput. Sci. Appl.* 7, 5 (2016), 560–566.

[95] Ghaith Abdulsattar A. Jabbar Alkubaisi, Siti Sakira Kamaruddin, and Husniza Husni. 2018. Stock market classification model using sentiment analysis on Twitter based on hybrid naive bayes classifiers. *Comput. Inf. Sci.* 11, 1 (2018), 52.

[96] Yousif Almas and Khurshid Ahmad. 2007. A note on extracting sentiments in financial news in English, Arabic & Urdu. In *Proceedings of the 2nd Workshop on Computational Approaches to Arabic Script-based Languages*. 1–12.

[97] Latifah Almuqren, Arwa Alzammam, Shahad Alotaibi, Alexandra Cristea, and Sarah Alhumoud. 2017. A review on corpus annotation for Arabic sentiment analysis. In *Proceedings of the International Conference on Social Computing and Social Media*. Springer, 215–225.

[98] Latifah Almuqren and Alexandra I. Cristea. 2016. Framework for sentiment analysis of Arabic text. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*. ACM, 315–317.

[99] Anwar Alnawas and Nursal Arıcı. 2018. The corpus based approach to sentiment analysis in modern standard Arabic and Arabic dialects: A literature review. *Polit. Derg.* 21, 2 (2018), 461–470.

[100] Khaled Mohammad Alomari, Hatem M. ElSherif, and Khaled Shaalan. 2017. Arabic tweets sentimental analysis using machine learning. In *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 602–610.

[101] Saud Alotaibi and Charles Anderson. 2017. Word clustering as a feature for Arabic sentiment classification. *Int. J. Educ. Manage. Eng.* 7, 1 (2017), 1–13.

[102] Saud S. Alotaibi and Charles W. Anderson. 2016. Extending the knowledge of the Arabic sentiment classification using a foreign external lexical source. *Int. J. Nat. Lang. Comput.* 5, 3 (2016), 1–11.

[103] Sana Alowaidi, Mustafa Saleh, and Osama Abulnaja. 2017. Semantic sentiment analysis of Arabic texts. *Int. J. Adv. Comput. Sci. Appl.* 8, 2 (2017), 256–262.

[104] Areeb Alowisheq, Sarah Alhumoud, Nora alTwairesh, and Tarfa Albuhairi. 2016. Arabic sentiment analysis resources: A survey. In *Proceedings of the 8th International Conference on Social Computing and Social Media (SCSM'16) Held as Part of HCI International 2016*. Springer, 267–278.

[105] Mohammed Alshahrani, Fuxi Zhu, Ahmed Sameh, Lin Zheng, and Summaya Mumtaz. 2018. Evaluating the influence of Twitter on the Saudi Arabian stock market indicators. In *Proceedings of the 5th International Symposium on Data Mining Applications*. Springer, 113–132.

[106] Kholoud Alsmearat, Mohammed Shehab, Mahmoud Al-Ayyoub, Riyad Al-Shalabi, and Ghassan Kanaan. 2015. Emotion analysis of arabic articles and its impact on identifying the author's gender. In *Proceedings of the 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA'15)*. IEEE, 1–6.

[107] A. Aziz Altowayan and Ashraf Elnagar. 2017. Improving Arabic sentiment analysis with sentiment-specific embeddings. In *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data'17)*. IEEE, 4314–4320.

[108] A. Aziz Altowayan and Lixin Tao. 2016. Word embeddings for Arabic sentiment analysis. In *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data'16)*. IEEE, 3820–3825.

[109] Ghadah Alwakid, Taha Osman, and Thomas Hughes-Roberts. 2017. Challenges in sentiment analysis for Arabic social networks. *Proc. Comput. Sci.* 117 (2017), 89–100.

[110] Mohamed A. Aly and Amir F. Atiya. 2013. LABR: A large scale Arabic book reviews dataset. In *Proceedings of the Annual Meeting of the Association of Computer Linguistics (ACL'13)*. 494–498.

[111] Matheus Araujo, Julio Reis, Adriano Pereira, and Fabricio Benevenuto. 2016. An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. ACM, 1140–1145.

[112] Adel Assiri, Ahmed Emam, and Hmood Al-Dossari. 2016. Saudi twitter corpus for sentiment analysis. *Int. J. Comput. Electr. Autom. Contr. Inf. Eng.* 10, 2 (2016), 242–245.

[113] Adel Assiri, Ahmed Emam, and Hmood Al-Dossari. 2017. Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. *J. Inf. Sci.* 44, 2 (2017).

[114] Adel Assiri, Ahmed Emam, and Hmood Aldossari. 2015. Arabic sentiment analysis: A survey. *Int. J. Adv. Comput. Sci. Appl.* 6, 12 (2015), 75–85.

[115] Sasi Atia and Khaled Shaalan. 2015. Increasing the accuracy of opinion mining in Arabic. In *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15)*. IEEE, 106–113.

[116] Hunaida Awwad and Adil Alpkocak. 2016. Performance comparison of different lexicons for sentiment analysis in Arabic. In *Proceedings of the 2016 3rd European Network Intelligence Conference (ENIC'16)*. IEEE, 127–133.

[117] Aqil M. Azmi and Samah M. Alzanin. 2014. Aara'—A system for mining the polarity of Saudi public opinion through e-newspaper comments. *J. Inf. Sci.* 40, 3 (2014), 398–410.

[118] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'10)*, Vol. 10. 2200–2204.

[119] Gilbert Badaro, Ramy Baly, Rana Akel, Linda Fayad, Jeffrey Khairallah, Hazem Hajj, Wassim El-Hajj, and Khaled Bashir Shaban. 2015. A light lexicon-based mobile application for sentiment mining of Arabic tweets. In *Proceedings of the Annual Conference on Natural Language Processing (ANLP'15) Workshop 2015*. 18.

[120] Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, and Wassim El-Hajj. 2014. A large scale Arabic sentiment lexicon for Arabic opinion mining. In *Proceedings of the Annual Conference on Natural Language Processing (ANLP'14)* 165 (2014).

[121] Gilbert Badaro, Obeida El Jundi, Alaa Khaddaj, Alaa Maarouf, Raslan Kain, Hazem Hajj, and Wassim El-Hajj. 2018. EMA at SemEval-2018 task 1: Emotion mining for Arabic. In *Proceedings of the 12th International Workshop on Semantic Evaluation*. 236–244.

[122] Gilbert Badaro, Hazem Hajj, Wassim El-Hajj, and Lama Nachman. 2013. A hybrid approach with collaborative filtering for recommender systems. In *Proceedings of the 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC'13)*. IEEE, 349–354.

[123] Gilbert Badaro, Hazem Hajj, Ali Haddad, Wassim El-Hajj, and Khaled Bashir Shaban. 2014. A multiresolution approach to recommender systems. In *Proceedings of the 8th Workshop on Social Network Mining and Analysis*. ACM, 9.

[124] Gilbert Badaro, Hazem Hajj, Ali Haddad, Wassim El-Hajj, and Khaled Bashir Shaban. 2014. Recommender systems using harmonic analysis. In *Proceedings of the 2014 IEEE International Conference on Data Mining Workshop (ICDMW'14)*. IEEE, 1004–1011.

[125] Gilbert Badaro, Hussein Jundi, Hazem Hajj, and Wassim El-Hajj. 2018. EmoWordNet: Automatic expansion of emotion lexicon using English WordNet. In *Proceedings of the 7th Joint Conference on Lexical and Computational Semantics*. 86–93.

[126] Gilbert Badaro, Be, Ramy Baly, Hazem Hajj, Nizar Habash, Wassim El-hajj, and Khaled Shaban. 2014. An efficient model for sentiment classification of Arabic tweets on mobiles. In *Proceedings of the Qatar Foundation Annual Research Conference*.

[127] Hitham Abo Bakr, Khaled Shaalan, and Ibrahim Ziedan. 2008. A hybrid approach for converting written Egyptian colloquial dialect into diacritized Arabic. In *Proceedings of the 6th International Conference on Informatics and Systems (INFOS'08)*.

[128] Georgios Balikas and Massih-Reza Amini. 2016. TwiSE at SemEval-2016 task 4: Twitter sentiment classification. *arXiv preprint arXiv:1606.04351* (2016).

[129] Ramy Baly, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash, and Khaled Bashir Shaban. 2017. A characterization study of Arabic Twitter data with a benchmarking for state-of-the-art opinion mining models. In *Proceedings of the Arabic Natural Language Processing Workshop (WANLP'17) (co-located with EACL 2017)*. 110.

[130] Ramy Baly, Gilbert Badaro, Ali Hamdi, Rawan Moukalled, Rita Aoun, Georges El-Khoury, Ahmad Al Sallab, Hazem Hajj, Nizar Habash, Khaled Shaban, et al. 2017. OMAM at SemEval-2017 task 4: Evaluation of English state-of-the-art sentiment analysis models for Arabic and a new topic-based model. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. Association for Computational Linguistics, 601–608.

[131] Ramy Baly, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Khaled Bashir Shaban, and Wassim El-Hajj. 2017. Comparative evaluation of sentiment analysis methods across Arabic dialects. *Proc. Comput. Sci.* 117 (2017), 266–273.

[132] Ramy Baly, Hazem Hajj, Nizar Habash, Khaled Bashir Shaban, and Wassim El-Hajj. 2017. A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in Arabic. *ACM Trans. Asian Low-Res. Lang. Inf. Process.* 16, 4 (2017), 23.

[133] Ramy Baly, Roula Hobeica, Hazem Hajj, Wassim El-Hajj, Khaled Bashir Shaban, and Ahmad Al-Sallab. 2016. A meta-framework for modeling the human reading process in sentiment analysis. *ACM Trans. Inf. Syst.* 35, 1 (2016), 7.

[134] Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2010. Multilingual subjectivity: Are more languages better? In *Proceedings of the 23rd International Conference on Computational Linguistics.* Association for Computational Linguistics, 28–36.

[135] Anoud Bani-Hani, Munir Majdalawieh, and Feras Obeidat. 2017. The creation of an Arabic emotion ontology based on E-Motive. *Proc. Comput. Sci.* 109 (2017), 1053–1059.

[136] Amira Barhoumi, Yannick Estève1 Chafik Aloulou, and Lamia Hadrich Belguith. 2017. Document embeddings for Arabic sentiment analysis. *Language Processing and Knowledge Management* 1988 (2017).

[137] Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. 2008. International sentiment analysis for news and blogs. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'08).*

[138] Amine Bayoudhi, Hatem Ghorbel, and Lamia Hadrich Belguith. 2015. Sentiment classification of Arabic documents: Experiments with multi-type features and ensemble algorithms. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation.* 196–205.

[139] Amine Bayoudhi, Hatem Ghorbel, Houssem Koubaa, and Lamia Hadrich Belguith. 2015. Sentiment classification at discourse segment level: Experiments on multi-domain Arabic corpus.*J. Lang. Technol. Comp. Ling.* 30, 1 (2015), 1–24.

[140] Yonatan Belinkov and James Glass. 2016. A character-level convolutional neural network for distinguishing similar languages and dialects. *arXiv preprint arXiv:1609.07568* (2016).

[141] Mariam Biltawi, Ghazi Al-Naymat, and Sara Tedmori. 2017. Arabic sentiment classification: A hybrid approach. In *Proceedings of the 2017 International Conference on New Trends in Computing Sciences (ICTCS'17).* IEEE, 104–108.

[142] Mariam Biltawi, Wael Etaiwi, Sara Tedmori, Amjad Hudaib, and Arafat Awajan. 2016. Sentiment classification techniques for Arabic language: A survey. In *Proceedings of the 2016 7th International Conference on Information and Communication Systems (ICICS'16).* IEEE, 339–346.

[143] William Black, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum. 2006. Introducing the Arabic wordnet project. In *Proceedings of the 3rd International WordNet Conference.* 295–300.

[144] Rihab Bouchlaghem, Aymen Elkhelifi, and Rim Faiz. 2016. A machine learning approach for classifying sentiments in Arabic tweets. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics.* ACM, 24.

[145] Rihab Bouchlaghem, Aymen Elkhelifi, and Rim Faiz. 2016. Sentiment analysis in Arabic Twitter posts using supervised methods with combined features. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics.* Springer, 320–334.

[146] Naaima Boudad, Rdouan Faizi, Rachid Oulad Haj Thami, and Raddouane Chiheb. 2017. Sentiment analysis in Arabic: A review of the literature. *Ain Shams Eng. J.* 9, 4 (2017), 2479–2490.

[147] Naaima Boudad, Rdouan Faizi, Rachid Oulad Haj Thami, and Raddouane Chiheb. 2017. Sentiment classification of Arabic tweets: A supervised approach. *J. Mobile Multimedia* 13, 3&4 (2017), 233–243.

[148] Belgacem Brahimi, Mohamed Touahria, and Abdelkamel Tari. 2016. Data and text mining techniques for classifying Arabic tweet polarity.*J. Dig. Inf. Manage.* 14, 1 (2016).

[149] Tim Buckwalter. 2002. Arabic Morphological Analyzer (AraMorph). Linguistic Data Consortium, Philadelphia.

[150] T. Buckwalter. 2004. *Buckwalter Arabic Morphological Analyzer (BAMA) Version 2.0. Linguistic Data Consortium (LDC) Catalogue Number LDC2004L02.* Technical Report. ISBN1-58563-324-0.

[151] Erik Cambria, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. 2016. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING'16).* 2666–2677.

[152] Nan Cao, Lu Lu, Yu-Ru Lin, Fei Wang, and Zhen Wen. 2015. Socialhelix: Visual analysis of sentiment divergence in social media. *J. Vis.* 18, 2 (2015), 221–235.

[153] Yanqing Chen and Steven Skiena. 2014. Building sentiment lexicons for all major languages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 383–389.

[154] Marc Cheong and Vincent C. S. Lee. 2011. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Inf. Syst. Front.* 13, 1 (2011), 45–59.

[155] Walid Cherif, Abdellah Madani, and Mohamed Kissi. 2015. A new modeling approach for Arabic opinion mining recognition. In *Proceedings of the 2015 Intelligent Systems and Computer Vision (ISCV'15)*. IEEE, 1–6.

[156] Walid Cherif, Abdellah Madani, and Mohamed Kissi. 2015. Towards an efficient opinion measurement in Arabic comments. *Proc. Comput. Sci.* 73 (2015), 122–129.

[157] Walid Cherif, Abdellah Madani, and Mohamed Kissi. 2016. A combination of low-level light stemming and support vector machines for the classification of Arabic opinions. In *Proceedings of the 2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA'16)*. IEEE, 1–5.

[158] Walid Cherif, Abdellah Madani, and Mohamed Kissi. 2016. A hybrid optimal weighting scheme and machine learning for rendering sentiments in tweets. *Int. J. Intell. Eng. Inf.* 4, 3-4 (2016), 322–339.

[159] David Chiang, Mona T. Diab, Nizar Habash, Owen Rambow, and Safiullah Shareef. 2006. Parsing Arabic dialects. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL'06)*.

[160] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning*. ACM, 160–167.

[161] Mohammed Dahab and Salma Al Asmari. 2015. Towards discovering the sentiment holder in Arabic text. *Int. J. Comput. Appl.* 111, 16 (2015).

[162] Abdelghani Dahou, Shengwu Xiong, Junwei Zhou, Mohamed Houcine Haddoud, and Pengfei Duan. 2016. Word embeddings and convolutional neural network for arabic sentiment classification. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING'16): Technical Papers*. 2418–2427.

[163] Bouras Dalila, Amroune Mohamed, and Hakim Bendjanna. 2018. A review of recent aspect extraction techniques for opinion mining systems. In *Proceedings of the 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP'18)*. IEEE, 1–6.

[164] Jan Deriu, Maurice Gonzenbach, Fatih Uzdilli, Aurelien Lucchi, Valeria De Luca, and Martin Jaggi. 2016. SwissCheese at SemEval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. In *Proceedings of the International Workshop on Semantic Evaluation at the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (SemEval@ NAACL-HLT'16)*. 1124–1128.

[165] Mona Diab. 2009. Second generation AMIRA tools for Arabic processing: Fast and robust tokenization, POS tagging, and base phrase chunking. In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools*, Vol. 110.

[166] Mona Diab, Kadri Hacioglu, and Daniel Jurafsky. 2004. Automatic tagging of Arabic text: From raw text to base phrase chunks. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL'04): Short Papers*. Association for Computational Linguistics, 149–152.

[167] Mona T. Diab, Mohamed Al-Badrashiny, Maryam Aminian, Mohammed Attia, Heba Elfardy, Nizar Habash, Abdelati Hawwari, Wael Salloum, Pradeep Dasigi, and Ramy Eskander. 2014. Tharwa: A large scale dialectal Arabic-standard Arabic-english lexicon. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14)*. 3782–3789.

[168] Dongxu Duan, Weihong Qian, Shimei Pan, Lei Shi, and Chuang Lin. 2012. VISA: A visual sentiment analysis system. In *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction*. ACM, 22–28.

[169] Kevin Duh and Katrin Kirchhoff. 2005. POS tagging of dialectal Arabic: A minimally supervised approach. In *Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages*. Association for Computational Linguistics, 55–62.

[170] R. M. Duwairi, Nizar A. Ahmed, and Saleh Y. Al-Rifai. 2015. Detecting sentiment embedded in Arabic social media–a lexicon-based approach. *J. Intell. Fuzzy Syst.* 29, 1 (2015), 107–117.

[171] Rehab Duwairi and Mahmoud El-Orfali. 2014. A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. *J. Inf. Sci.* 40, 4 (2014), 501–513.

[172] R. M. Duwairi, Raed Marji, Narmeen Sha'ban, and Sally Rushaidat. 2014. Sentiment analysis in Arabic tweets. In *Proceedings of the 2014 5th International Conference on Information and Communication Systems (ICICS'14)*. IEEE, 1–6.

[173] Rehab M. Duwairi. 2015. Sentiment analysis for dialectical Arabic. In *Proceedings of the 2015 6th International Conference on Information and Communication Systems (ICICS'15)*. IEEE, 166–170.

[174] Rehab M. Duwairi, Mosab Alfaqeh, Mohammad Wardat, and Areen Alrabadi. 2016. Sentiment analysis for Arabizi text. In *Proceedings of the 2016 7th International Conference on Information and Communication Systems (ICICS'16)*. IEEE, 127–132.

[175] Rehab M. Duwairi and Mohammad A. Alshboul. 2015. Negation-aware framework for sentiment analysis in Arabic reviews. In *Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud'15)*. IEEE, 731–735.

[176] Rehab M. Duwairi and Islam Qarqaz. 2016. A framework for Arabic sentiment analysis using supervised classification. *Int. J. Data Min. Model. Manage.* 8, 4 (2016), 369–381.

[177] A. Eesee and N. Omar. 2016. A hybrid method for Arabic educational sentiment analysis. *J. Appl. Sci.* 16, 5 (2016), 216–222.

[178] Samhaa R. El-Beltagy. 2016. Niletmrg at semeval-2016 task 7: Deriving prior polarities for arabic sentiment terms. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 486–490.

[179] Samhaa R. El-Beltagy. 2016. NileULex: A phrase and word level sentiment lexicon for egyptian and modern standard Arabic. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'16)*.

[180] Samhaa R. El-Beltagy. 2017. WeightedNileULex: A scored Arabic sentiment lexicon for improved sentiment analysis. *Language Processing, Pattern Recognition and Intelligent Systems. Special Issue on Computational Linguistics, Speech& Image Processing for Arabic Language.* World Scientific Publishing Co.

[181] Samhaa R. El-Beltagy and Ahmed Ali. 2013. Open issues in the sentiment analysis of Arabic social media: A case study. In *Proceedings of the 2013 9th International Conference on Innovations in Information Technology (IIT'13)*. IEEE, 215–220.

[182] Samhaa R. El-Beltagy, Mona El Kalamawy, and Abu Bakr Soliman. 2017. NileTMRG at SemEval-2017 task 4: Arabic sentiment analysis. *arXiv preprint arXiv:1710.08458* (2017).

[183] Samhaa R. El-Beltagy, Talaat Khalil, Amal Halaby, and Muhammad Hammad. 2016. Combining lexical features and a supervised learning approach for Arabic sentiment analysis. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 307–319.

[184] Alaa El-Halees. 2011. Arabic opinion mining using combined classification approach. *Naif Arab University for Security Sciences.* http://hdl.handle.net/20.500.12358/25202

[185] Alaa El-Halees. 2014. Mining changes of opinions expressed by students to improve course evaluation. In *Proceedings of the International ARAB Conference on Information Technology (ACIT'14)*, 72–79.

[186] Alaa El-Halees and Ahmed Al-Asmar. 2017. Ontology based Arabic opinion mining. *J. Inf. Knowl. Manage.* 16, 03 (2017), 1750028.

[187] Alaa M. El-Halees. 2017. Arabic opinion mining using distributed representations of documents. In *Proceedings of the 2017 Palestinian International Conference on Information and Communication Technology (PICICT'17)*. IEEE, 28–33.

[188] Ayman El-Kilany, Amr Azzam, and Samhaa R El-Beltagy. 2018. Using deep neural networks for extracting sentiment targets in Arabic tweets. In *Intelligent Natural Language Processing: Trends and Applications*. Springer, 3–15.

[189] Nagwa El-Makky, Khaled Nagi, Alaa El-Ebshihy, Esraa Apady, Omneya Hafez, Samar Mostafa, and Shimaa Ibrahim. 2014. Sentiment analysis of colloquial Arabic tweets. In *Proceedings of the ASE BigData/SocialInformatics/PASSAT/BioMedCom 2014 Conference*. 1–9.

[190] Mazen El-Masri, Nabeela Altrabsheh, and Hanady Mansour. 2017. Successes and challenges of Arabic sentiment analysis research: A literature review. *Soc. Netw. Anal. Min.* 7, 1 (2017), 54.

[191] Mazen El-Masri, Nabeela Altrabsheh, Hanady Mansour, and Allan Ramsay. 2017. A web-based tool for Arabic sentiment analysis. *Proc. Comput. Sci.* 117 (2017), 38–45.

[192] Nadine El-Naggar, Yasser El-Sonbaty, and Mohamad Abou El-Nasr. 2017. Sentiment analysis of modern standard Arabic and Egyptian dialectal Arabic tweets. In *Proceedings of the Computing Conference 2017*. IEEE, 880–887.

[193] Mohamed Elarnaoty, Samir AbdelRahman, and Aly Fahmy. 2012. A machine learning approach for opinion holder extraction in Arabic language. *arXiv preprint arXiv:1206.1011* (2012).

[194] R. M. Elawady, S. Barakat, H. M. El-Bakry, and N. Elrashidy. 2015. Sentiment analysis for Arabic and English datasets. *Int. J. Innov. Comput. Inf. Contr.* 15, 1 (2015).

[195] Rasheed M. Elawady, Sherif Barakat, and Nora M. Elrashidy. 2014. Different feature selection for sentiment classification. *Int. J. Inf. Sci. Intell. Syst.* 3, 1 (2014), 137–150.

[196] Heba Elfardy, Mohamed Al-Badrashiny, and Mona Diab. 2014. AIDA: Identifying code switching in informal Arabic text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. 94.

[197] Heba Elfardy and Mona T. Diab. 2012. Token level identification of linguistic code switching. In *Proceedings of the International Conference on Computational Linguistics (COLING'12)*. 287–296.

[198] Tarek Elghazaly, Amal Mahmoud, and Hesham A. Hefny. 2016. Political sentiment analysis using twitter data. In *Proceedings of the International Conference on Internet of things and Cloud Computing*. ACM, 11.

[199] Mohamed Elhawary and Mohamed Elfeky. 2010. Mining Arabic business reviews. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops (ICDMW'10)*. IEEE, 1108–1113.

[200] AbdelRahim A. Elmadany and Walid Magdy Hamdy Mubarak. 2018. ArSAS: An Arabic speech-act and sentiment corpus of tweets. In *Proceedings of the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT 3)*. 20.

[201] Ashraf Elnagar. 2016. Investigation on sentiment analysis for Arabic reviews. In *Proceedings of the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA'16)*. IEEE, 1–7.

[202] Ashraf Elnagar and Omar Einea. 2016. Brad 1.0: Book reviews in arabic dataset. In *Proceedings of the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA'16)*. IEEE, 1–8.

[203] Ashraf Elnagar, Omar Einea, and Leena Lulu. 2017. Comparative study of sentiment classification for automated translated latin reviews into Arabic. In *Proceedings of the 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA'17)*. IEEE, 443–448.

[204] Ashraf Elnagar, Yasmin S. Khalifa, and Anas Einea. 2018. Hotel Arabic-reviews dataset construction for sentiment analysis applications. In *Intelligent Natural Language Processing: Trends and Applications*. Springer, 35–52.

[205] Abdeljalil Elouardighi, Hafdalla Hammia, and Mohcine Maghfour. 2017. Collecting and processing multilingual streaming tweets for sentiment analysis. In *Proceedings of the International Conference on Information Technology and Communication Systems*. Springer, 10–23.

[206] Abdeljalil Elouardighi, Mohcine Maghfour, and Hafdalla Hammia. 2017. Collecting and processing arabic facebook comments for sentiment analysis. In *Proceedings of the International Conference on Model and Data Engineering*. Springer, 262–274.

[207] Mohammed Elrazzaz, Shady Elbassuoni, Khaled Shaban, and Chadi Helwe. 2017. Methodical evaluation of arabic word embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 454–458.

[208] Hady ElSahar and Samhaa R. El-Beltagy. 2014. A fully automated approach for arabic slang lexicon extraction from microblogs. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 79–91.

[209] Hady ElSahar and Samhaa R. El-Beltagy. 2015. Building large arabic multi-domain resources for sentiment analysis. In *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing (CICLing'15)*. 23–34.

[210] Ramy Eskander and Owen Rambow. 2015. SLSA: A sentiment lexicon for standard Arabic. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'15)*. 2545–2550.

[211] Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR)* (2006).

[212] Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem Hajj. 2010. Sentence-level and document-level sentiment mining for arabic texts. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops (ICDMW'10)*. IEEE, 1114–1119.

[213] Noura Farra and Kathleen McKeown. 2017. Smarties: Sentiment models for arabic target entities. *arXiv preprint arXiv:1701.03434* (2017).

[214] Noura Farra, Kathy McKeown, and Nizar Habash. 2015. Annotating targets of opinions in arabic using crowdsourcing. In *Proceedings of the 2nd Workshop on Arabic Natural Language Processing*. 89–98.

[215] Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Commun. ACM* 56, 4 (2013), 82–89.

[216] Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.

[217] Ramy Georges Baly, Me, Gilbert Badaro, Hazem Hajj, Nizar Habash, Wassim El Hajj, and Khaled Shaban. 2014. Semantic model representation for human's pre-conceived notions in Arabic text with applications to sentiment mining. In *Proceedings of the Qatar Foundation Annual Research Conference*. ITPP1075.

[218] Al-Sukkar Ghadeer, Ibrahim Aljarah, and Hamad Alsawalqah. 2017. Enhancing the Arabic sentiment analysis using different preprocessing operators. In *Proceedings of the New Trends Inf. Technol. (NTIT'17)*, 113.

[219] Hazem Hajj Wassim El-Hajj Gilbert Badaro, Hussein Jundi and Nizar Habash. 2018. ArSEL: A large scale Arabic sentiment and emotion lexicon. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18)*.

[220] José-Angel González, Ferran Pla, and Lluís-F Hurtado. 2017. ELiRF-UPV at SemEval-2017 task 4: Sentiment analysis using deep learning. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. 723–727.

[221] Mario Graff, Eric S. Tellez, Hugo Jair Escalante, and Sabino Miranda-Jiménez. 2017. Semantic genetic programming for sentiment analysis. In *Results of the Numerical and Evolutionary Optimization Workshop (NEO'15)*. Springer, 43–65.

[222] Mario Graff, Eric S. Tellez, Sabino Miranda-Jiménez, and Hugo Jair Escalante. 2016. Evodag: A semantic genetic programming python library. In *Proceedings of the 2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC'16)*. IEEE, 1–6.

[223] Spence Green and Christopher D. Manning. 2010. Better Arabic parsing: Baselines, evaluations, and analysis. In *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, 394–402.

[224] Mourad Gridach, Hatem Haddad, and Hala Mulki. 2017. Empirical evaluation of word representations on Arabic sentiment analysis. In *Proceedings of the International Conference on Arabic Language Processing*. Springer, 147–158.

[225] Imane Guellil, Ahsan Adeel, Faical Azouaou, and Amir Hussain. 2018. Sentialg: Automated corpus annotation for algerian sentiment analysis. In *Proceedings of the International Conference on Brain Inspired Cognitive Systems*. Springer, 557–567.

[226] Nizar Habash, Ramy Eskander, and Abdelati Hawwari. 2012. A morphological analyzer for Egyptian Arabic. In *Proceedings of the 12th Meeting of the Special Interest Group on Computational Morphology and Phonology*. Association for Computational Linguistics, 1–9.

[227] Nizar Habash and Owen Rambow. 2005. Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 573–580.

[228] Nizar Habash, Owen Rambow, and Ryan Roth. 2009. MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR'09)*, Vol. 41. 62.

[229] Nizar Habash and Ryan M. Roth. 2009. CATiB: The Columbia Arabic treebank. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, 221–224.

[230] Nizar Habash and Fatiha Sadat. 2006. Arabic preprocessing schemes for statistical machine translation. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*. Association for Computational Linguistics, 49–52.

[231] Nizar Habash, Abdelhadi Soudi, and Timothy Buckwalter. 2007. On arabic transliteration. In *Arabic Computational Morphology*. Springer, 15–22.

[232] Nizar Y. Habash. 2010. Introduction to Arabic natural language processing. *Synth. Lect. Hum. Lang. Technol.* 3, 1 (2010), 1–187.

[233] Ali Hamdi, Khaled Shaban, and Anazida Zainal. 2016. A review on challenging issues in Arabic sentiment analysis. *J. Comput. Sci.* 12, 9 (2016), 471–481.

[234] Ali Hamdi, Khaled Shaban, and Anazida Zainal. 2018. Clasenti: A class-specific sentiment analysis framework. *ACM Trans. Asian Low-Res. Lang. Inf. Process.* 17, 4 (2018), 32.

[235] AL-Rubaiee Hamed, Renxi Qiu, Khalid Alomar, and Dayou Li. 2018. Techniques for improving the labelling process of sentiment analysis in the Saudi stock market. *Int. J. Adv. Comput. Technol.* 9, 3 (2018), 34–43.

[236] AL-Rubaiee Hamed, Renxi Qiu, and Dayou Li. 2015. Analysis of the relationship between Saudi twitter posts and the Saudi stock market. In *Proceedings of the 2015 IEEE 7th International Conference on Intelligent Computing and Information Systems (ICICIS'15)*. IEEE, 660–665.

[237] Alaa El-Dine Ali Hamouda and Fatma El-zahraa El-taher. 2013. Sentiment analyzer for arabic comments system. *Int. J. Adv. Comput. Sci. Appl.* 4, 3 (2013).

[238] Safa Ben Hamouda and Jalel Akaichi. 2013. Social networks text mining for sentiment classification: The case of Facebook statuses updates in the Arabic Spring Era. *Int. J. Appl. Innov. Eng. Manage.* 2, 5 (2013), 470–478.

[239] Fouzi Harrag. 2014. Estimating the sentiment of arabic social media contents: A survey. In *Proceedings of the 5th International Conference on Arabic Language Processing*.

[240] Ahmed Hassan, Amjad Abu-Jbara, Rahul Jha, and Dragomir Radev. 2011. Identifying the semantic orientation of foreign words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2*. Association for Computational Linguistics, 592–597.

[241] Nourah F. Bin Hathlian and Alaaeldin M. Hafezs. 2016. Sentiment-subjective analysis framework for arabic social media posts. In *Proceedings of the Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT'16)*. IEEE, 1–6.

[242] Abdel Rahman Hedar and M. Doss. 2013. Mining social networks arabic slang comments. In *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM'13)*.

[243] David R. Heise. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. Springer Science & Business Media.

[244] Orland Hoeber, Larena Hoeber, Maha El Meseery, Kenneth Odoh, and Radhika Gopi. 2016. Visual Twitter analytics (vista) temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Inf. Rev.* 40, 1 (2016), 25–41.

[245] Amal Htait, Sébastien Fournier, and Patrice Bellot. 2017. LSIS at SemEval-2017 task 4: Using adapted sentiment similarity seed words for English and Arabic tweet polarity classification. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. 718–722.

[246] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 168–177.

[247] Hossam S. Ibrahim, Sherif M. Abdou, and Mervat Gheith. 2015. Automatic expandable large-scale sentiment lexicon of modern standard Arabic and colloquial. In *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15)*. IEEE, 94–99.

[248] Hossam S Ibrahim, Sherif M Abdou, and Mervat Gheith. 2015. Idioms-proverbs lexicon for modern standard Arabic and colloquial sentiment analysis. *arXiv preprint arXiv:1506.01906* (2015).
[249] Hossam S. Ibrahim, Sherif M. Abdou, and Mervat Gheith. 2015. MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis. In *Proceedings of the 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*. IEEE, 353–358.
[250] Hossam S. Ibrahim, Sherif M. Abdou, and Mervat Gheith. 2015. Sentiment analysis for modern standard Arabic and colloquial. *arXiv preprint arXiv:1505.03105* (2015).
[251] Mnahel A. Ibrahim and Naomie Salim. 2016. Aspect oriented sentiment analysis model of Arabic tweets. *Int. J. Comput. Sci. Trends Technol.* 4, 4 (2016), 342–353.
[252] Michael Nawar Ibrahim, Mahmoud N. Mahmoud, and Dina A. El-Reedy. 2016. Bel-Arabi: Advanced Arabic grammar analyzer. *Int. J. Soc. Sci. Human.* 6, 5 (2016), 341.
[253] Noah Iliinsky and Julie Steele. 2011. *Designing Data Visualizations: Representing Informational Relationships.* O'Reilly Media, Inc.
[254] Shimaa Ismail, Abdelwahab Alsammak, and Tarek Elshishtawy. 2016. A generic approach for extracting aspects and opinions of Arabic reviews. In *Proceedings of the 10th International Conference on Informatics and Systems.* ACM, 173–179.
[255] Maher Itani, Chris Roast, and Samir Al-Khayatt. 2017. Corpora for sentiment analysis of Arabic text in social media. In *Proceedings of the 2017 8th International Conference on Information and Communication Systems (ICICS'17)*. IEEE, 64–69.
[256] Maher Itani, Chris Roast, and Samir Al-Khayatt. 2017. Developing resources for sentiment analysis of informal Arabic text in social media. *Proc. Comput. Sci.* 117 (2017), 129–136.
[257] Maher M. Itani, Rached N. Zantout, Lama Hamandi, and Islam Elkabani. 2012. Classifying sentiment in arabic social networks: Naive search versus naive bayes. In *Proceedings of the 2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA'12)*. IEEE, 192–197.
[258] Mohammed Jabreel and Antonio Moreno. 2017. SiTAKA at SemEval-2017 task 4: Sentiment analysis in Twitter based on a rich set of features. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. 694–699.
[259] Brendan Jou, Margaret Yuying Qian, and Shih-Fu Chang. 2016. SentiCart: Cartography and geo-contextualization for multilingual visual sentiment. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval.* ACM, 389–392.
[260] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188* (2014).
[261] Jihen Karoui, Farah Banamara Zitoune, and Veronique Moriceau. 2017. SOUKHRIA: Towards an irony detection system for arabic in social media. *Proc. Comput. Sci.* 117 (2017), 161–168.
[262] Gehad S. Kaseb and Mona F. Ahmed. 2016. Arabic sentiment analysis approaches: An analytical survey. *Int. J. Sci. Eng. Res.* 7, 10 (2016), 712–723.
[263] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Gorg, Jorn Kohlhammer, and Guy Melançon. 2008. Visual analytics: Definition, process, and challenges. *Lecture Notes in Computer Science* 4950 (2008), 154–176.
[264] Khalid Khalifa and Nazlia Omar. 2014. A hybrid method using lexicon-based approach and naive Bayes classifier for Arabic opinion question answering. *J. Comput. Sci.* 10, 10 (2014), 1961–1968.
[265] Salam Khalifa, Sara Hassan, and Nizar Habash. 2017. A morphological analyzer for gulf Arabic verbs. In *Arabic Natural Language Processing Workshop (WANLP'17) (Co-located with EACL 2017)*. 35.
[266] Talaat Khalil, Amal Halaby, Muhammad Hammad, and Samhaa R. El-Beltagy. 2015. Which configuration works best? an experimental study on supervised Arabic twitter sentiment analysis. In *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15)*. IEEE, 86–93.
[267] Muhammad Taimoor Khan and Shehzad Khalid. 2016. Sentiment analysis for health care. In *Big Data: Concepts, Methodologies, Tools, and Applications.* IGI Global, 676–689.
[268] Rawan T. Khasawneh, Heider A. Wahsheh, Mohammed N. Al-Kabi, and Izzat M. Alsmadi. 2013. Sentiment analysis of arabic social media content: A comparative study. In *Proceedings of the 2013 8th International Conference for Internet Technology and Secured Transactions (ICITST'13)*. IEEE, 101–106.
[269] Rawan T. Khasawneh, Heider A. Wahsheh, Izzat M. Alsmadi, and Mohammed N. AI-Kabi. 2015. Arabic sentiment polarity identification using a hybrid approach. In *Proceedings of the 2015 6th International Conference on Information and Communication Systems (ICICS'15)*. IEEE, 148–153.
[270] H. Kilany, H. Gadalla, H. Arram, A. Yacoub, A. El-Habashi, and C. McLemore. 2002. Egyptian colloquial Arabic lexicon. *LDC Catalog Number LDC99L22* (2002).
[271] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).

[272] Yoosin Kim and Seung Ryul Jeong. 2015. Opinion-mining methodology for social media analytics. *KSII Trans. Internet Inf. Syst.* 9, 1 (2015).

[273] Svetlana Kiritchenko, Saif Mohammad, and Mohammad Salameh. 2016. Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 42–51.

[274] Svetlana Kiritchenko and Saif M. Mohammad. 2017. Capturing reliable fine-grained sentiment associations by crowdsourcing and best-worst scaling. *arXiv preprint arXiv:1712.01741* (2017).

[275] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*. Association for Computational Linguistics, 177–180.

[276] Moshe Koppel, Navot Akiva, Eli Alshech, and Kfir Bar. 2009. Automatically classifying documents by ideological and organizational affiliation. In *Proceedings of the IEEE International Conference on Intelligence and Security Informatics 2009 (ISI'09)*. IEEE, 176–178.

[277] Mohammed Korayem, David J. Crandall, and Muhammad Abdul-Mageed. 2012. Subjectivity and sentiment analysis of Arabic: A survey. In *Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications (AMLTA'12)*. Springer, 128–139.

[278] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing. In *Proceedings of the International Conference on Machine Learning*. 1378–1387.

[279] Farek Lazhar and Tlili Guiassa Yamina. 2010. Identification of opinions in Arabic newspapers. In *Proceedings of the 2010 International Conference on Machine and Web Intelligence (ICMWI'10)*. IEEE, 317–319.

[280] Farek Lazhar and Tlili Guiassa Yamina. 2012. Identification of opinions in arabic texts using ontologies. In *Proceedings of the Workshop on Ubiquitous Data Mining*. IEEE, 61–64.

[281] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng. 2014. News impact on stock price return via sentiment analysis. *Knowl.-Based Syst.* 69 (2014), 14–23.

[282] Shenghua Liu, Wenjun Zhu, Ning Xu, Fangtao Li, Xue-qi Cheng, Yue Liu, and Yuanzhuo Wang. 2013. Co-training and visualizing sentiment evolvement for tweet events. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 105–106.

[283] Mohamed Maamouri, Ann Bies, Tim Buckwalter, and Wigdan Mekki. 2004. The penn Arabic treebank: Building a large-scale annotated Arabic corpus. In *Proceedings of the NEMLAR Conference on Arabic Language Resources and Tools*, Vol. 27. 466–467.

[284] Mohamed Maamouri, Dave Graff, Basma Bouziri, Sondos Krouna, Ann Bies, and Seth Kulick. 2010. Standard Arabic morphological analyzer (SAMA) version 3.1. Linguistic Data Consortium, Catalog No.: LDC2010L01 (2010).

[285] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 142–150.

[286] Amal Mahmoud and Tarek Elghazaly. 2018. Using Twitter to monitor political sentiment for Arabic slang. In *Intelligent Natural Language Processing: Trends and Applications*. Springer, 53–66.

[287] Fawaz H. H. Mahyoub, Muazzam A. Siddiqui, and Mohamed Y. Dahab. 2014. Building an Arabic sentiment lexicon using semi-supervised learning. *J. King Saud Univ. Comput. Inf. Sci.* 26, 4 (2014), 417–424.

[288] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. 2011. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 227–236.

[289] Yuval Marton, Nizar Habash, and Owen Rambow. 2013. Dependency parsing of modern standard Arabic with lexical and inflectional features. *Comput. Ling.* 39, 1 (2013), 161–194.

[290] M'hamed Mataoui, Tadj Eddine Bendali Hacine, Imad Tellache, Abdelghani Bakhtouchi, and Omar Zelmati. 2018. A new syntax-based aspect detection approach for sentiment analysis in Arabic reviews. In *Proceedings of the 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP'18)*. IEEE, 1–6.

[291] M'hamed Mataoui, Omar Zelmati, and Madiha Boumechache. 2016. A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. *Res. Comput. Sci.* 110 (2016), 55–70.

[292] Salima Medhaffar, Fethi Bougares, Yannick Esteve, and Lamia Hadrich-Belguith. 2017. Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. In *Proceedings of the 3rd Arabic Natural Language Processing Workshop*. 55–61.

[293] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.

[294] Sabino Miranda-Jiménez, Mario Graff, Eric Sadit Tellez, and Daniela Moctezuma. 2017. INGEOTEC at SemEval 2017 task 4: A B4MSA ensemble based on genetic programming for Twitter sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. 771–776.

[295] Emad Mohamed, Behrang Mohit, and Kemal Oflazer. 2012. Annotating and learning morphological segmentation of egyptian colloquial Arabic. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'12)*. 873–877.

[296] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*. 1–17.

[297] Saif Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. Sentiment lexicons for Arabic social media. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'16)*.

[298] Saif M. Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decis. Supp. Syst.* 53, 4 (2012), 730–741.

[299] Saif M. Mohammad. 2015. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emot. Meas.* (2015), 201–238.

[300] Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242* (2013).

[301] Saif M. Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How translation alters sentiment. *J. Artif. Intell. Res.* 55 (2016), 95–130.

[302] Tagwa Abd Elatif Mohammed. 2016. Review of sentiment analysis for classification Arabic tweets. *Int. J. Emerg. Technol. Adv. Eng.* 6, 3 (2016), 47–53.

[303] Behrang Mohit, Alla Rozovskaya, Nizar Habash, Wajdi Zaghouani, and Ossama Obeid. 2014. The first QALB shared task on automatic text correction for Arabic. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP'14)*. 39–47.

[304] Ayman Mohamed Mostafa. 2017. An automatic lexicon with exceptional-negation algorithm for Arabic sentiments using supervised classification. *J. Theor. Appl. Inf. Technol.* 95, 15 (2017).

[305] Ayman Mohamed Mostafa. 2017. An evaluation of sentiment analysis and classification algorithms for Arabic textual data. *Int. J. Comput. Appl.* 158, 3 (2017).

[306] Asmaa Mountassir, Houda Benbrahim, and Ilham Berrada. 2012. Addressing the problem of unbalanced data sets in sentiment analysis. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR'12)*. 306–311.

[307] Asmaa Mountassir, Houda Benbrahim, and Ilham Berrada. 2012. A cross-study of sentiment classification on Arabic corpora. In *Research and Development in Intelligent Systems XXIX*. Springer, 259–272.

[308] Asmaa Mountassir, Houda Benbrahim, and Ilham Berrada. 2012. An empirical study to address the problem of unbalanced data sets in sentiment classification. In *Proceedings of the 2012 IEEE International Conference onSystems, Man, and Cybernetics (SMC)*. IEEE, 3298–3303.

[309] Asmaa Mountassir, Houda Benbrahim, and Ilham Berrada. 2012. Some methods to address the problem of unbalanced sentiment classification in an arabic context. In *Proceedings of the 2012 Colloquium in Information Science and Technology (CIST)*. IEEE, 43–48.

[310] Ahmed Mourad and Kareem Darwish. 2013. Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs. In *WASSA and Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (WASSA@ NAACL-HLT'13)*. 55–64.

[311] Hala Mulki, Hatem Haddad, Mourad Gridach, and Ismail Babaoğlu. 2017. Tw-StAR at SemEval-2017 task 4: Sentiment classification of Arabic tweets. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*. 664–669.

[312] Manal Mustafa, A. Shakour AlSamahi, and Alaa Hamouda. 2017. New avenues in Arabic sentiment analysis. *Int. J. Sci. Eng. Res.* 8, 2 (2017), 907–915.

[313] Mahmoud Nabil, Mohamed Aly, and Amir Atiya. 2014. LABR: A large scale Arabic sentiment analysis benchmark. *arXiv preprint arXiv:1411.6718* (2014).

[314] Mahmoud Nabil, Mohamed A. Aly, and Amir F. Atiya. 2015. ASTD: Arabic sentiment tweets dataset. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'15)*. 2515–2519.

[315] Hassan Najadat, Amnah Al-Abdi, and Yasmeen Sayaheen. 2018. Model-based sentiment analysis of customer satisfaction for the Jordanian telecommunication companies. In *Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS'18)*. IEEE, 233–237.

[316] Dhekra Najar and Slim Mesfar. 2014. Political monitoring and opinion mining for standard Arabis texts. In *Proceedings of the International NooJ Conference*.

[317] Dhekra Najar and Slim Mesfar. 2017. Opinion mining and sentiment analysis for Arabic on-line texts: Application on the political domain. *Int. J. Speech Technol.* 20, 3 (2017), 575–585.

[318] Vu Dung Nguyen, Blesson Varghese, and Adam Barker. 2013. The royal birth of 2013: Analysing and visualising public sentiment in the uk using twitter. In *Proceedings of the 2013 IEEE International Conference on Big Data.* IEEE, 46–54.

[319] Aya Nuseir, Mahmoud Al-Ayyoub, Mohammed Al-Kabi, Ghasan Kanaan, and Riyad Al-Shalabi. 2017. Improved hierarchical classifiers for multi-way sentiment analysis. *Int. Arab J. Inf. Technol.* 14, 4A (2017), 654–661.

[320] Islam Obaidat, Rami Mohawesh, Mahmoud Al-Ayyoub, Al-Smadi Mohammad, and Yaser Jararweh. 2015. Enhancing the determination of aspect categories and their polarities in arabic reviews using lexicon-based approaches. In *Proceedings of the 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT'15).* IEEE, 1–6.

[321] Nazlia Omar, Mohammed Albared, Adel Qasem Al-Shabi, and Tareq Al-Moslmi. 2013. Ensemble of classification algorithms for subjectivity and sentiment analysis of Arabic customers' reviews. *Int. J. Advance. Comput. Technol.* 5, 14 (2013), 77.

[322] Shereen Oraby, Yasser El-Sonbaty, and Mohamad Abou El-Nasr. 2013. Exploring the effects of word roots for arabic sentiment analysis. In *Proceedings of the 6th International Joint Conference on Natural Language Processing.* 471–479.

[323] Shereen Oraby, Yasser El-Sonbaty, and Mohamad Abou El-Nasr. 2013. Finding opinion strength using rule-based parsing for Arabic sentiment analysis. In *Proceedings of the Mexican International Conference on Artificial Intelligence.* Springer, 509–520.

[324] Mai Oudah and Khaled Shaalan. 2017. NERA 2.0: Improving coverage and performance of rule-based named entity recognition for Arabic. *Nat. Lang. Eng.* 23, 3 (2017), 441–472.

[325] Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retriev.* 2, 1–2 (2008), 1–135.

[326] Arfath Pasha, Mohamed Al-Badrashiny, Mona T. Diab, Ahmed El Kholy, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow, and Ryan Roth. 2014. MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14)*, Vol. 14. 1094–1101.

[327] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP*'14), Vol. 14. 1532–1543.

[328] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Al-Smadi Mohammad, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16).* 19–30.

[329] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval'14).*

[330] Martin Popel and Zdenek Zabokrtskỳ. 2010. TectoMT: Modular NLP framework. In *proceedings of the international conference on Natural Language Processing (NLP'10)*, 293–304.

[331] Ali Mustafa Qamar, Suliman A. Alsuhibany, and Syed Sohail Ahmed. 2017. Sentiment classification of twitter data belonging to saudi arabian telecommunication companies. *Int. J. Adv. Comput. Sci. Appl.* 1, 8 (2017), 395–401.

[332] Abdullateef M. Rabab'ah, Mahmoud Al-Ayyoub, Yaser Jararweh, and Mohammed N. Al-Kabi. 2016. Evaluating sentistrength for arabic sentiment analysis. In *Proceedings of the 2016 7th International Conference on Computer Science and Information Technology (CSIT'16).* IEEE, 1–6.

[333] Limia Rahamatallah, Eltayeb Abuelyaman, and Wafaa Mukhtar. 2014. Constructing opinion mining model of sudanese telecom products. Master Thesis at Sudan University of Science and Technology (SUST).

[334] Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. Sarcasm detection on twitter: A behavioral modeling approach. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining.* ACM, 97–106.

[335] Ramy Baly, Alaa Khaddaj, Hazem Hajj, Wassim El-Hajj, Khaled Bashir Shaban. 2018. ArSentD-LEV: A multi-topic corpus for target-based sentiment analysis in Arabic levantine tweets. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18).*

[336] Eshrag Refaee. 2017. Sentiment analysis for micro-blogging platforms in Arabic. In *Proceedings of the International Conference on Social Computing and Social Media.* Springer, 275–294.

[337] Eshrag Refaee and Verena Rieser. 2014. An Arabic Twitter corpus for subjectivity and sentiment analysis. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14).* 2268–2273.

[338] Eshrag Refaee and Verena Rieser. 2014. Can we read emotions from a smiley face? Emoticon-based distant supervision for subjectivity and sentiment analysis of arabic twitter feeds. In *Proceedings of the 5th International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data (LREC'14).*

[339] Eshrag Refaee and Verena Rieser. 2014. Evaluating distant supervision for subjectivity and sentiment analysis on Arabic twitter feeds. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP'14).* 174–179.

[340] Eshrag Refaee and Verena Rieser. 2014. Subjectivity and sentiment analysis of Arabic twitter feeds with limited resources. In *Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools Workshop Programme*. 16.

[341] Eshrag Refaee and Verena Rieser. 2015. Benchmarking machine translated sentiment analysis for Arabic tweets. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*. 71–78.

[342] Eshrag Refaee and Verena Rieser. 2016. iLab-edinburgh at SemEval-2016 task 7: A hybrid approach for determining sentiment intensity of Arabic Twitter phrases. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 474–480.

[343] Internet ICT Telecommunications Reports. [n. d.]. Internet World Users by Language. Top 10 Languages. retrieved June 30, 2017 from https://www.internetworldstats.com/stats7.htm.

[344] Elaine Cristina Resende, and Aron Culotta. 2015. A demographic and sentiment analysis of e-cigarette messages on Twitter. In *Proceedings of the Computational Health Sciences Workshop at the 6th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB'15).*

[345] Sandra Rizkallah, Amir Atiya, Hossam ElDin Mahgoub, and Momen Heragy. 2018. Dialect versus MSA sentiment analysis. In *Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications*. Springer, 605–613.

[346] Christian Rohrdantz, Ming C. Hao, Umeshwar Dayal, Lars-Erik Haug, and Daniel A. Keim. 2012. Feature-based visual sentiment analysis of text document streams. *ACM Trans. Intell. Syst. Technol.* 3, 2 (2012), 26.

[347] Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17).* 502–518.

[348] Sebastian Ruder, Parsa Ghaffari, and John G. Breslin. 2016. A hierarchical model of reviews for aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02745* (2016).

[349] Sebastian Ruder, Parsa Ghaffari, and John G. Breslin. 2016. Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02748* (2016).

[350] Mohammed Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and José M. Perea-Ortega. 2011. Bilingual experiments with an arabic-english corpus for opinion mining. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*. 740–745.

[351] Mohammed Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and José M. Perea-Ortega. 2011. OCA: Opinion corpus for Arabic. *J. Assoc. Inf. Sci. Technol.* 62, 10 (2011), 2045–2054.

[352] Motaz Saad, David Langlois, and Kamel Smaïli. 2013. Comparing multilingual comparable articles based on opinions. In *Proceedings of the 6th Workshop on Building and Using Comparable Corpora*. 105–111.

[353] Motaz Saad, David Langlois, and Kamel Smaïli. 2014. Building and modelling multilingual subjective corpora. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA).

[354] Khaled S. Sabra, Rached N. Zantout, Mohamad A. El Abed, and Lama Hamandi. 2017. Sentiment analysis: Arabic sentiment lexicons. In *Proceedings of the Conference on Sensors Networks Smart and Emerging Technologies (SENSET'17), 2017*. 1–4.

[355] Janan Ben Salamah and Aymen Elkhlifi. 2014. Microblogging opinion mining approach for kuwaiti dialect. In *Proceedings of the International Conference on Computing Technology and Information Management (ICCTIM'14).* The Society of Digital Information and Wireless Communication, 388–396.

[356] Mohammad Salameh, Saif Mohammad, and Svetlana Kiritchenko. 2015. Sentiment after translation: A case-study on arabic social media posts. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 767–777.

[357] Said A. Salloum, Ahmad Qasim AlHamad, Mostafa Al-Emran, and Khaled Shaalan. 2018. A survey of Arabic text mining. In *Intelligent Natural Language Processing: Trends and Applications*. Springer, 417–431.

[358] Wael Salloum and Nizar Habash. 2011. Dialectal to standard Arabic paraphrasing to improve Arabic-English statistical machine translation. In *Proceedings of the 1st Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*. Association for Computational Linguistics, 10–21.

[359] Younes Samih, Mohammed Attia, Mohamed Eldesouki, Hamdy Mubarak, Ahmed Abdelali, Laura Kallmeyer, and Kareem Darwish. 2017. A neural architecture for dialectal Arabic segmentation. In *Proceedings of the Arabic Natural Language Processing Workshop (WANLP'17)*. 46.

[360] Arno Scharl, David Herring, Walter Rafelsberger, Alexander Hubmann-Haidvogel, Ruslan Kamolov, Daniel Fischl, Michael Föls, and Albert Weichselbraun. 2017. Semantic systems and visual tools to support environmental communication. *IEEE Syst. J.* 11, 2 (2017), 762–771.

[361] SemEval-2017. [n. d.]. International Workshop on Semantic Evaluation. Retrieved September 9, 2017 from http://alt.qcri.org/semeval2017/index.php?id=tasks.
[362] Khaled Shaalan and Hafsa Raza. 2009. NERA: Named entity recognition for Arabic. *J. Assoc. Inf. Sci. Technol.* 60, 8 (2009), 1652–1663.
[363] Amira Shoukry and Ahmed Rafea. 2012. Sentence-level Arabic sentiment analysis. In *Proceedings of the 2012 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 546–550.
[364] Amira Shoukry and Ahmed Rafea. 2015. A hybrid approach for sentiment classification of Egyptian dialect tweets. In *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15)*. IEEE, 78–85.
[365] Amira Magdy Shoukry. 2013. Arabic sentence-level sentiment analysis. Master Thesis at the American University in Cairo. Retrieved from http://dar.aucegypt.edu/handle/10526/3536.
[366] Muazzam Ahmed Siddiqui, Mohamed Yehia Dahab, and Omar Abdullah Batarfi. 2015. Building a sentiment analysis corpus with multifaceted hierarchical annotation. *Int. J. Comput. Ling.* 6, 2 (2015), 11–25.
[367] Sanjeera Siddiqui, Azza Abdel Monem, and Khaled Shaalan. 2017. Towards improving sentiment analysis in Arabic. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*, Aboul Ella Hassanien, Khaled Shaalan, Tarek Gaber, Ahmad Taher Azar, and M. F. Tolba (Eds.). Springer International Publishing, Cham, 114–123.
[368] Sanjeera Siddiqui, Azza Abdel Monem, and Khaled Shaalan. 2018. Evaluation and enrichment of Arabic sentiment analysis. In *Intelligent Natural Language Processing: Trends and Applications*. Springer, 17–34.
[369] Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language processing*. Association for Computational Linguistics, 151–161.
[370] Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, Christopher Potts, et al. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'13)*, Vol. 1631. 1642.
[371] Abu Bakr Soliman, Kareem Eissa, and Samhaa R. El-Beltagy. 2017. Aravec: A set of arabic word embedding models for use in arabic nlp. *Proc. Comput. Sci.* 117 (2017), 256–265.
[372] Taysir Hassan Soliman, M. A. Elmasry, A. Hedar, and M. M. Doss. 2014. Sentiment analysis of Arabic slang comments on facebook. *Int. J. Comput. Technol.* 12, 5 (2014), 3470–3478.
[373] Robert Speer, Joshua Chin, and Catherine Havasi. 2017. ConceptNet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence (AAAI'17)*. 4444–4451.
[374] F. Steiner-Correa, M. I. Viedma-del Jesus, and A. G. Lopez-Herrera. 2017. A survey of multilingual human-tagged short message datasets for sentiment analysis tasks. *Soft Comput.* 22, 24 (2017), 8227–8242.
[375] Aleš Tamchyna and Kateřina Veselovská. 2016. Ufal at semeval-2016 task 5: Recurrent neural networks for sentence classification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 367–371.
[376] Duyu Tang, Bing Qin, and Ting Liu. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'15)*. 1422–1432.
[377] Samir Tartir and Ibrahim Abdul-Nabi. 2017. Semantic sentiment analysis in Arabic social media. *J. King Saud Univ.-Comput. Inf. Sci.* 29, 2 (2017), 229–233.
[378] Eric S. Tellez, Sabino Miranda-Jiménez, Mario Graff, Daniela Moctezuma, Ranyart R. Suárez, and Oscar S. Siordia. 2017. A simple approach to multilingual polarity classification in Twitter. *Pattern Recogn. Lett.* 94 (2017), 68–74.
[379] Mike Thelwall. 2013. Heart and soul: Sentiment strength detection in the social web with sentistrength. *Proc. CyberEmot.* 5 (2013), 1–14.
[380] Megan K. Torkildson, Kate Starbird, and Cecilia Aragon. 2014. Analysis and visualization of sentiment and emotion on crisis tweets. In *Proceedings of the International Conference on Cooperative Design, Visualization and Engineering*. Springer, 64–67.
[381] Imen Touati, Marwa Graja, Mariem Ellouze, and Lamia Hadrich Belguith. 2016. Arabic fine-grained opinion categorization using discriminative machine learning technique. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*. Springer, 104–113.
[382] Imen Touati, Marwa Graja, Mariem Ellouze, and Lamia Hadrich Belguith. 2016. Towards Arabic semantic opinion mining: Identifying opinion, polarity and intensity. In *Proceedings of the Mediterranean Conference on Pattern Recognition and Artificial Intelligence*. ACM, 131–136.
[383] Peter D. Turney and Michael L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *arXiv preprint cs/0212012* (2002).
[384] Heider A. Wahsheh, Mohammed N. Al-Kabi, and Izzat M. Alsmadi. 2013. SPAR: A system to detect spam in Arabic opinions. In *Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT'13)*. IEEE, 1–6.

[385] Changbo Wang, Zhao Xiao, Yuhua Liu, Yanru Xu, Aoying Zhou, and Kang Zhang. 2013. SentiView: Sentiment analysis and visualization for internet popular topics. *IEEE Trans. Hum.-Mach. Syst.* 43, 6 (2013), 620–630.

[386] Florence Ying Wang, Arnaud Sallaberry, Karsten Klein, Masahiro Takatsuka, and Mathieu Roche. 2015. SentiCompass: Interactive visualization for exploring and comparing the sentiments of time-varying twitter data. In *Proceedings of the 2015 IEEE Pacific Visualization Symposium (PacificVis'15).* IEEE, 129–133.

[387] Hao Wang, Ayman Hanafy, Mohamed Bahgat, Sara Noeman, Ossama S. Emam, and Vijay R. Bommireddipalli. 2015. A system for extracting sentiment from large-scale Arabic social data. In *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15).* IEEE, 71–77.

[388] Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X. Zhou, Weihong Qian, Lei Shi, Li Tan, and Qiang Zhang. 2010. Tiara: A visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 153–162.

[389] Andreas Weiler, Michael Grossniklaus, and Marc H. Scholl. 2015. *The Stor-e-Motion Visualization for Topic Evolution Tracking in Text Data Streams.*

[390] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing.* Association for Computational Linguistics, 347–354.

[391] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2017. Recent trends in deep learning based natural language processing. *arXiv preprint arXiv:1708.02709* (2017).

[392] La-sheng Yu and Sadeq al Baadani. 2018. A sentiment analysis approach based on Arabic social media platforms. In *proceedings of the International Conference on Mechanical, Electronic and Information Technology (ICMEIT'18).* 504–510.

[393] Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Ossama Obeid, Nadi Tomeh, Alla Rozovskaya, Noura Farra, Sarah Alkuhlani, and Kemal Oflazer. 2014. Large scale Arabic error annotation: Guidelines and framework. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14).* 2362–2369.

[394] Omar F. Zaidan and Chris Callison-Burch. 2014. Arabic dialect identification. *Comput. Ling.* 40, 1 (2014), 171–202.

[395] Taoufiq Zarra, Raddouane Chiheb, Rajae Moumen, Rdouan Faizi, and Abdellatif El Afia. 2017. Topic and sentiment model applied to the colloquial Arabic: A case study of maghrebi Arabic. In *Proceedings of the 2017 International Conference on Smart Digital Environment.* ACM, 174–181.

[396] Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F. Zaidan, and Chris Callison-Burch. 2012. Machine translation of Arabic dialects. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Association for Computational Linguistics, 49–59.

[397] Imed Zitouni, Jeffrey S. Sorensen, and Ruhi Sarikaya. 2006. Maximum entropy based restoration of Arabic diacritics. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, 577–584.


Received April 2018; revised September 2018; accepted November 2018