Pertanika J. Sci. & Technol. 29 (1): 445 - 470 (2021)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Multilingual Sentiment Analysis: A Systematic Literature Review

Nur Atiqah Sia Abdullah[1]* and Nur Ida Aniza Rusli[2]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[2]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kampus Kuala Pilah, Pekan Parit Tinggi, 72000 Kuala Pilah, Negeri Sembilan, Malaysia*

ABSTRACT

With the explosive growth of social media, the online community can freely express their opinions without disclosing their identities. People with hidden agendas can easily post fake opinions to discredit target products, services, politicians, or organizations. With these big data, monitoring opinions and distilling their sentiments remain a formidable task because of the proliferation of diverse sites with a large volume of opinions that are portrayed in multilingual. Therefore, this paper aims to provide a systematic literature review on multilingual sentiment analysis, which summarises the common languages supported in multilingual sentiment analysis, pre-processing techniques, existing sentiment analysis approaches, and evaluation models that have been used for multilingual sentiment analysis. By following the systematic literature review, the findings revealed, most of the models supported two languages, and English is seen as the most used language in sentiment analysis studies. None of the reviewed literature has catered the combination of languages for English, Chinese, Malay, and Hindi language on multilingual sentiment analysis. The common pre-processing techniques for the multilingual domain are tokenization, normalization, capitalization, N-gram, and machine translation. Meanwhile, the sentiment analysis classification techniques for multilingual sentiment are hybrid sentiment analysis, which includes localized

ARTICLE INFO

Article history:
Received: 08 September 2020
Accepted: 07 December 2020
Published: 22 January 2021

DOI: https://doi.org/10.47836/pjst.29.1.25

E-mail addresses:
atiqah684@uitm.edu.my (Nur Atiqah Sia Abdullah)
idaaniza@uitm.edu.my (Nur Ida Aniza Rusli)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

language analysis, unsupervised topic clustering, and then followed by multilingual sentiment analysis. In terms of evaluation, most of the studies used precision, recall, and accuracy as the benchmark for the results.

Keywords: Machine learning, machine translation, multilingual sentiment analysis, opinion mining, pre-processing, sentiment analysis

INTRODUCTION

There is a well-known English saying "*The pen is mightier than the sword*" written by a novelist Edward Bulwer-Lytton in 1839 (as in Sykes et al., 2018), that emphasizes how the freedom of speech (including written and oral communications) has generally been a powerful tool than a weapon due to its capability to influence, persuade and control the society or situation (Mäntylä et al., 2018; Jing & Murugesan, 2018). The freedom of speech has opened opportunities for people to publicly voice out their feelings and opinions through various communication mediums without restrictions, which sometimes cause more damage and violated the right of free expressions.

Sentiment analysis is a method or process of detecting and extracting a given subject such as opinion and attitudes from written and spoken language. In general, sentiment analysis is about the ability to determine the sentiment of a topic and classify the overall polarity of the topic sentence in positive, negative, or neutral (Kang & Park, 2014). Sentiment analysis has been a popular research area over the past decade. It is gaining even more importance over time due to the emerging use of the internet and social media such as social networking sites; Twitter, forums, and blogs. Capturing public opinions about political issues, social events, products preferences, or services that they have used are valuable for understanding the concerns and to influence the decision-making process.

However, this sentiment analysis is facing an issue where the written opinions are often mixed with several languages which leads to the difficulties in fully capturing the text messages and consequently making the polarity of the text becomes harder to classify (Dashtipour et al., 2016; Devika et al., 2016). Hence, multilingual sentiment analysis is proposed to enhance the sentiment classification of texts in multiple languages. Recently, there has been considerable interest in multilingual sentiment analysis. Numerous methods and automatic tools have been developed to extract relevant information from various sources.

The purpose of this paper is to review different combination of strategies to develop a multilingual sentiment analysis that includes preprocessing techniques, sentiment analysis methods, and evaluation model that have been applied in the existing proposed models.

This paper is organized as follows. Section 2 briefly explains the concept of sentiment analysis. Section 3 presents the systematic literature review methods and processes. Section 4 and Section 5 present the results and discussions. Finally, Section 6 concludes the paper.

SENTIMENT ANALYSIS

The following sections describe the concept of sentiment analysis including pre-processing, sentiment analysis classification techniques, and evaluation models for multilingual sentiment analysis.

Pre-processing

Texts, especially in blogs, Twitter, and online chats are known to comprise various spelling errors, slang words, and multilingual words. Thus, pre-processing is important to remove irrelevant part of the texts, and to transform into a readable format to extract the sentiment. Some of pre-processing techniques in sentiment analysis are listed as follows (Dashtipour et al., 2016; Devika et al., 2016; Yadav & Elchuri, 2013).

Tokenization. A process of splitting text into words, phrases or other important parts called tokens. Tokens are separated by whitespace, punctuation marks, and line breaks; and characters such as punctuation marks are usually removed during the tokenization process. Tokenization is considered relatively easy compared to other preprocessing techniques.

Stopword Removal. A process of discarding words that do not have significant meaning such as ‘a’, ‘of’, and ‘is’.

Stemming. A process of identifying the root of a specific word. For example, stemming puts variation of words such as ‘greatly’, ‘greatest’, and ‘greater’ to the root word ‘great’.

N-gram Generation. A set of co-occurring words or letters taken from a body of text. The n-gram usually consists of bigram (n=2), for example, “honesty is”, “is the”, “the best”, or “best policy” and trigram (n=3) like “honesty is the”, “is the best” or “the best policy”.

Lemmatization. A process of converting words to its initial form. Unlike stemming, lemmatization considers the context of the texts and translates the word into its relevant structure. For example, lemmatization would map the word ‘caring’ to the form of the word ‘care’, whereas stemming would transform the word ‘caring’ to ‘car’.

POS (Part-of-speech) Tagging. A process of tagging a word in a text with its part of speech such as noun, verb, adverb, pronoun, preposition, and conjunction.

Noise Removal. A process of excluding noise such as HTML tags, keywords, scripts, or advertisements.

Normalization. A process of cleaning text and removing insignificant data such as word redundancy, spelling error, symbols, or tags.

Word Embedding/Text Vectors. A process to capture the similarities of words. In other words, it represents words in a coordinate system where related words, based on a corpus of relationships, are placed closer together. Word2Vec is the most common model for word embedding process.

Capitalization. A process of converting all letters to lowercase. Capitalization preprocessing technique is important to be employed especially for Twitter since Twitter users commonly use uppercase to express their emotions in texts.

Negation. A process of reversing the text polarity. A negation word can influence the structure of the whole sentence. When negation words such as ‘no’, ‘not’, and ‘never’. appear in a text, it is important to identify the scope of negation, as the presence of negations sometimes does not indicate the negative polarity.

Machine Translation. Several studies that adopt machine translation (to translate texts usually to English language) as a step to process texts and documents. Google Translate, Bing Translator, and Babylon translator are the most common machine translation tools used in sentiment analysis.

Sentiment Analysis Classification Techniques

Generally, the research on sentiment analysis is categorized into three approaches: machine learning, lexical based, and hybrid methods (Rajput & Solanki, 2016; Thakkar & Patel, 2015; Bahrainian & Dengel, 2013).

Machine Learning (ML). A method to teach a machine to learn and process data more efficiently. In this method, algorithms are used to train the computer to identify complex patterns, usually in big data and provide a decision based on the input given. These algorithms generally can be divided into two groups: supervised and unsupervised learning. Supervised learning is a type of system where data scientist or supervisor guides the algorithms to produce the aimed output. On the other hand, unsupervised learning can learn and recognize the pattern without human guidance. Support Vector Machine (SVM), Naive Bayesian (NB), Artificial Neural Network (ANN), and k-means algorithm are the common techniques used in ML (Sabbeh, 2018; Michie et al., 1994).

Lexicon Approach. A method that utilizes a predefined set of patterns, which is also known as sentiment dictionary or lexicon. Each data entry will be associated with sentiment orientation. For example, the word "great" is classified as positive sentiment word, and the word "bad" is classified as negative sentiment word. The sentiment classification for lexicon approach can be implemented either using dictionary-based or corpus-based approach. In dictionary-based approach, a dictionary which contains synonyms and antonyms of words are referred towards the opinion words from the texts. The dictionaries such as WordNet, SentiWordNet, SenticNet are usually used to classify the sentiment polarity of the words. Meanwhile, for corpus-based approach, the method works by relying on syntactic rules in large corpora. It provides a list of opinion words with relatively high precision of specific context (Hamouda & Rohaim, 2011; Esuli & Sebastiani, 2006).

Hybrid Method. The method combines the concept of machine learning and lexicon-based approach. The process generally started by analysing texts using lexicon-based approach. The produced results are then inserted into the machine learning as training data (Ardabili et al., 2019; Tsai & Wang, 2009).

Evaluation Methods

Evaluation methods are part of sentiment analysis proposals. There are many different evaluation models used for evaluating multilingual sentiment analysis. The most common evaluation models are as follows (Sokolova et al., 2006; Padmaja & Fatima, 2013):

Accuracy. Predictions of how often classifier makes correction prediction. It measures the ratio of correct predictions over the total number of instances evaluated.

$$Accuracy = \frac{Correct\ Prediction}{Total\ Number\ of\ Instance}$$

Precision. Calculates the exactness of a classifier the consistency of the results when the measurements are repeated. It measures the positive patterns that are correctly predicted from the total predicted patterns in a positive class.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall. Computes the number of positive class predictions made out of all positive examples in the dataset. It measures the fraction of positive patterns that are correctly classified.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

F-measure/ F1-score. Provides a single metric that balances both of precision and recall in one number. It measures the mean value between recall and precision value.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

SYSTEMATIC LITERATURE REVIEW

Appropriate guidelines have been followed to conduct this systematic literature review, particularly the guidelines for SLRs in Software Engineering by Kitchenham and Charters (2007). The systematic literature review includes the research questions, data sources and search strategy, study selection, inclusion/exclusion criteria, and quality assessment as shown in Figure 1.

Research Questions

The research question addressed by this study is "What is the best sentiment analysis approach for multilingual sentiment analysis specific to English, Malay and Chinese?

- What are the existing multilingual languages involved in sentiment analysis?
- Which pre-processing techniques are suitable to extract multilingual sentiment analysis?

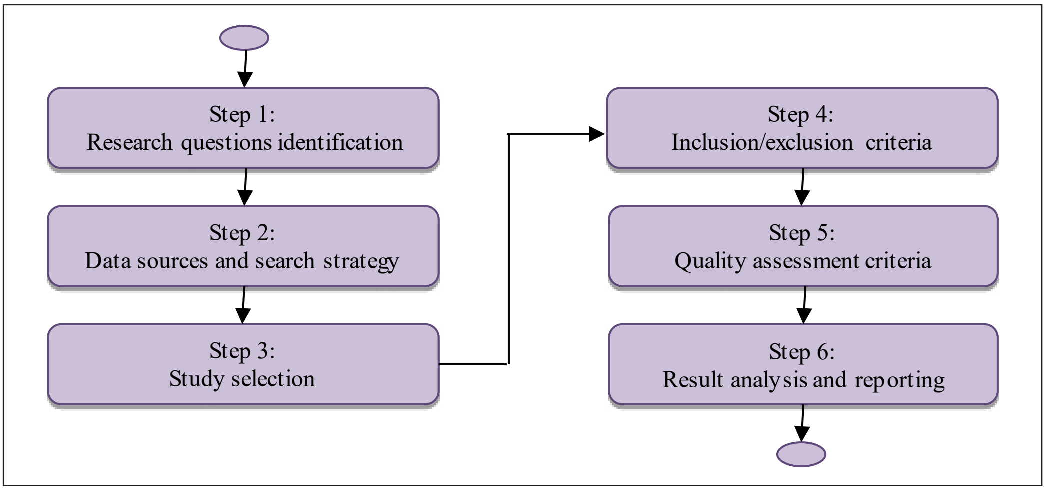

Figure 1. A systematic literature review flow for multilingual sentiment analysis

- What sentiment analysis classification methods are available for multilingual domain?
- How to evaluate the sentiment analysis classification methods in multilingual sentiment analysis?

Data Sources and Search Strategy

The planning stage also involved enumerating data sources which we searched for studies or previous works and used to define the query strings that would be executed on those sources. The following digital libraries were selected to carry out the search process of this review:

- IEEEXplore: http://ieeexplore.ieee.org/
- Science Direct: http://www.sciencedirect.com/.
- SpringerLink: http://www.springerlink.com/.
- ACM Digital Library: http://portal.acm.org/.
- Google Scholar: http://scholar.google.com/

We formulated the search terms using Population, Intervention, Contrast, and Outcome (Kitchenham & Charters, 2007). The following general search string was eventually used: "multilingual sentiment analysis" and ("pre-processing" or "lexicon-based" or "machine learning"). Table 1 shows that our initial searches elicited 336 articles.

Table 1
Search results

Digital Libraries	**Search Results**
IEEE Explore	64
Science Direct	57
Springer Link	66
ACM Digital Library	18
Google Scholar	131
Total	**336**

Study Selection

We obtained 336 articles in the first search process. Since many articles did not provide sufficient information to answer the research questions, we performed another filtration steps as follows:

- Step 1: remove the duplicated articles obtained by authors and/or different libraries.
- Step 2: apply inclusion and exclusion criteria to the candidate papers to avoid any irrelevant articles.
- Step 3: apply the quality assessment rules to include the qualified articles that give the best answers to the research questions.

Inclusion/Exclusion Criteria

After removing the duplicated articles, we obtained 248 articles. Next, we adopted the following inclusion/exclusion criteria. We went through the abstract and body of each paper to ensure their relevance according to these criteria. Defining inclusion and exclusion criteria helped to clarify the boundaries of the study.

The inclusion criteria:

- Primary studies published in journals, conferences, or workshop proceedings in the form of experiments, surveys, case studies, reports, and observation papers using multilingual sentiment analysis.
- Primary studies within the period from 2010 to 2019.

The exclusion criteria:

- Non-English publications
- Publications that did not include multilingual sentiment
- Informal studies (unknown conferences or journals)
- Articles that were irrelevant to the research questions

Quality Assessment

Once we had selected several works based on inclusion and exclusion criteria (51 articles), we assessed the quality of the research they presented. Six Quality Assessment (QA) questions (Indajat et al., 2016; Maita et al., 2015) had been defined to assess the quality of the research of each proposal and to provide a quantitative comparison between them. The scoring procedure used was Yes (Y) = 1, Partly (P) = 0.5 or No (N) = 0. The quality assessment questions defined in this systematic literature review, are as shown in Table 2.

Table 3 presents the results of the quality assessment score of each article. Each study could obtain a score ranging from zero to six points. Any study that awarded with a quality score three and lower was eliminated from the review. From the results, four articles were excluded since they did not satisfy the assessment criteria. There are 45 articles with grade four and higher that were considered as the resources for this review. The selected articles are listed in Table 4.

RESULTS

What are the Existing Multilingual Languages that involved in Sentiment Analysis?

A summary of results for research question RQ1 is presented in Tables 5 and 6. The results in Table 5 shows 31% (n=14) of selected studies used two languages in their proposed model, 27% (n=12) used three languages, 18% (n=8) used four languages, 13% (n=6) used five languages, 7% (n=3) used seven languages, and only 4% (n=2) used eight languages in their proposed method.

Table 2
Quality assessment checklist

Item	Assessment Criteria	Score	Description
QA1	Is there a clear statement in research aims?	0	No, aims are not described
		0.5	Partially, aims are described but unclearly
		1	Yes, aims are well described and clear
QA2	Does the pre-processing/ feature used in the study is clearly described?	0	No, the pre-processing/feature are not described
		0.5	Partially, the pre-processing/feature are described but unclearly
		1	Yes, the pre-processing/feature are well described and clear
QA3	Does the study present a detailed description of the approach (classifier/ techniques)?	0	No, details are missing
		0.5	Partially, if you want to use the approach, you need to read the references
		1	Yes, the approach can be used with presented details
QA4	Does the study present a detailed evaluation of the approach?	0	No, evaluation is missing
		0.5	Partially, evaluation process is described but unclearly
		1	Yes, the evaluation process is well described and clear
QA5	Is there a comparison with other approach?	0	No, comparison with other approach is missing
		0.5	Partially, comparison is described but unclearly
		1	Yes, the comparison with other approach is well described and clear
QA6	Is there a clear statement of the findings?	0	No, findings are not described
		0.5	Partially, findings are described but unclearly
		1	Yes, aims are well described and clear

Table 3
Quality assessment score

No	Author(s)	QA1	QA2	QA3	QA4	QA5	QA6	Total Score
1	Pessutto et al., 2018	1	0.5	1	1	1	1	**5.5**
2	Pustulka-Hunt et al., 2018	1	1	1	1	1	1	**6**
3	Vilares et al., 2018	1	0	1	1	0	1	**4**
4	Wehrmann et al., 2018	1	1	1	1	1	1	**6**
5	Vīksna and Jēkabsons, 2018	1	0	0	0	0	1	**2**
6	Bhargava and Sharma, 2017	1	1	1	1	0	1	**6**
7	Becker et al., 2017a	1	1	1	1	1	1	**6**
8	Lo et al., 2017b	1	1	1	1	0	1	**5**
9	Tellez et al., 2017	1	1	1	1	1	1	**6**
10	Vilares et al., 2017	1	1	1	1	0	1	**5**
11	Lo et al., 2017a	1	0	0	0	0	1	**2**
12	Becker et al. 2017b	1	1	1	1	1	1	**6**
13	Karima and Smaili, 2016	1	1	1	1	0	1	**5**
14	Lu and Mori, 2017	1	1	1	1	1	1	**6**
15	Kaity and Balakrishnan, 2017	1	0	1	0	0	0	**2**
16	Patel et al., 2017	1	1	1	1	1	1	**6**
17	Rosenthal et al., 2017	1	0	1	1	0	1	**4**

Table 3 (*continue*)

No	Author(s)	QA1	QA2	QA3	QA4	QA5	QA6	Total Score
18	Al-Shabi et al., 2017	1	0.5	1	1	1	1	**5.5**
19	Kaur et al., 2017	1	1	1	0	0	0	**3**
20	Deriu et al., 2017	1	1	1	1	1	1	**6**
21	Lo et al., 2016	1	1	1	1	1	1	**6**
22	Saravia et al., 2016	1	1	1	1	1	1	**6**
23	Araujo et al., 2016	1	0	1	1	1	1	**5**
24	Zhou et al., 2016	1	1	1	1	0	1	**5**
25	Pappas et al., 2016	1	0.5	1	1	0	1	**4.5**
26	Argueta et al., 2016	1	1	1	1	1	1	**6**
27	Shalunts and Backfried, 2016	1	0	1	0	1	1	**4**
28	Dadoun and Olssson, 2016	1	0.5	1	1	0	1	**4.5**
29	Balahur and Perea-Ortega, 2015	1	1	1	1	1	1	**6**
30	Nowson et al., 2015	1	1	1	1	0	1	**5**
31	Vilares et al., 2015	1	1	1	1	0	1	**5**
32	Shaluns and Backfried, 2015	1	1	1	1	1	1	**6**
33	Lin et al., 2014b	1	1	1	1	1	1	**6**
34	Balahur and Turchi, 2014	1	1	1	1	1	1	**6**
35	Cruz et al., 2014	1	1	1	1	1	1	**6**
36	Balahur et al., 2014	1	0.5	1	1	0	1	**4.5**
37	Abdel-Hady et al., 2014	1	0.5	1	1	0	1	**4.5**
38	Lin et al., 2014a	1	1	1	1	1	1	**6**
39	Erdmann et al., 2014	1	0.5	1	1	0	1	**4.5**
40	Volkova et al., 2013	1	0	1	1	0	1	**4**
41	Balahur and Turchi, 2013	1	1	1	1	0	1	**5**
42	Saad et al., 2013	1	1	1	1	1	1	**6**
43	Demirtas and Pechenizkiy, 2013	1	1	1	1	1	1	**6**
44	Balahur and Turchi, 2012a	1	1	1	1	1	1	**6**
45	Balahur and Turchi, 2012b	1	1	1	1	1	1	**6**
46	Tromp and Pechenizkiy, 2011	1	1	1	1	0	1	**5**
47	Cui et al., 2011	1	1	1	1	1	1	**6**
48	Gînscă et al., 2011	1	1	1	1	0	1	**5**
49	Steinberger et al., 2011	1	0	1	1	0	1	**4**

Table 4
List of selected articles

Study ID	Author (s)	Digital Library	Year
S1	Pessutto et al.	IEEE	**2018**
S2	Pustulka-Hunt et al.	IEEE	**2018**
S3	Vilares et al.	IEEE	**2018**
S4	Wehrmann et al.	Google Scholar	**2018**
S5	Bhargava, and Sharma	IEEE	**2017**
S6	Becker et al.	Science Direct	**2017a**
S7	Lo et al.	Science Direct	**2017b**
S8	Tellez et al.	Science Direct	**2017**
S9	Vilares et al.	Science Direct	**2017**

Table 4 (*continue*)

Study ID	Author (s)	Digital Library	Year
S1	Pessutto et al.	IEEE	**2018**
S2	Pustulka-Hunt et al.	IEEE	**2018**
S3	Vilares et al.	IEEE	**2018**
S4	Wehrmann et al.	Google Scholar	**2018**
S5	Bhargava, and Sharma	IEEE	**2017**
S6	Becker et al.	Science Direct	**2017a**
S7	Lo et al.	Science Direct	**2017b**
S8	Tellez et al.	Science Direct	**2017**
S9	Vilares et al.	Science Direct	**2017**
S10	Becker et al.	Google Scholar	**2017b**
S11	Karima and Smaili	Google Scholar	**2016**
S12	Lu and Mori	Google Scholar	**2017**
S13	Patel et al.	Google Scholar	**2017**
S14	Rosenthal et al.	Google Scholar	**2017**
S15	Al-Shabi et al.	Google Scholar	**2017**
S16	Deriu et al.	ACM	**2017**
S17	Lo et al.	Science Direct	**2016**
S18	Saravia et al.	Springer Link	**2016**
S19	Araujo et al.	ACM	**2016**
S20	Zhou et al.	ACM	**2016**
S21	Pappas et al.	ACM	**2016**
S22	Argueta et al.	Google Scholar	**2016**
S23	Shalunts and Backfried	Google Scholar	**2016**
S24	Dadoun and Olssson	Google Scholar	**2016**
S25	Balahur and Perea-Ortega	Science Direct	**2015**
S26	Nowson et al.	Google Scholar	**2015**
S27	Vilares et al.	Google Scholar	**2015**
S28	Shalunts and Backfried	Springer Link	**2015**
S29	Lin et al.	IEEE	**2014b**
S30	Balahur and Turchi	Science Direct	**2014**
S31	Cruz et al.	Science Direct	**2014**
S32	Balahur et al.	Google Scholar	**2014**
S33	Abdel-Hady et al.	Google Scholar	**2014**
S34	Lin et al.	ACM	**2014a**
S35	Erdmann et al.	Springer Link	**2014**
S36	Volkova et al.	Google Scholar	**2013**
S37	Balahur and Turchi	Google Scholar	**2013**
S38	Saad et al.	Google Scholar	**2013**
S39	Demirtas and Pechenizkiy	ACM	**2013**
S40	Balahur and Turchi	ACM	**2012a**
S41	Balahur and Turchi	Google Scholar	**2012b**
S42	Tromp and Pechenizkiy	IEEE	**2011**
S43	Cui et al.	Springer Link	**2011**
S44	Gînscă et al.	ACM	**2011**
S45	Steinberger et al.	Google Scholar	**2011**

Table 5
Number of languages supported in sentiment analysis

Number of Language	Study ID	Total	%
2 languages	S6, S7, S9, S11, S13, S14, S15, S24, S25, S27, S35, S38, S39, S42	14	**31%**
3 languages	S2, S12, S17, S18, S22, S28, S29, S30, S33, S36, S40, S44	12	**27%**
4 languages	S4, S5, S10, S16, S23, S26, S41, S43	8	**18%**
5 languages	S1, S3, S20, S31, S32, S37	6	**13%**
6 languages	-	0	**0**
7 languages	S19, S34, S45	3	**7%**
8 languages	S8, S21	2	**4%**

The detail languages supported in the selected studies are shown in Table 6. From the results, we can conclude that 91% (n=41) applied a combination of English with other languages, 13% (n=6) applied a combination of English, Chinese and other languages, and only one article (S17) that included English, Chinese and Malay on multilingual sentiment analysis. Meanwhile, two studies (S3 and S13) were using the combination of English, Hindi, and other languages.

Thus, from the literature review, we found out that none of the studies had catered the combination of languages for English, Chinese, Malay, and Hindi language on multilingual sentiment analysis. This is an important finding because based on the social environment in Malaysia, major races use languages like English, Malay, and Chinese language. Hindi language is not used in Malaysia. Most of the Indians in Malaysia use Tamil language. However, there seems to be a lag in Tamil language use among the younger generation. The younger generation tends to converse in Malay and English more than Tamil (Paramasivam & Farashaiyan, 2016).

Table 6
Languages used in the selected articles

Study ID	Target Languages
S1	English, Spanish, Dutch, Russian, and Turkish
S2	English, German, and French
S3	Spanish, Italian, Portuguese, Chinese, and Hindi
S4	English, Spanish, German, and Portuguese
S5	English, Spanish, German, and French
S6	English and Portuguese
S7	English and Chinese (dialect)
S8	English, Spanish, German, Italian, Portuguese, Russian, Arabic, and Swedish
S9	English and Spanish
S10	English, Spanish, German, and Portuguese
S11	English and Arabic
S12	English, Chinese, and Japanese

Table 6 (*continue*)

Study ID	Target Languages
S13	English and Hindi
S14	English and Arabic
S15	English and Arabic
S16	English, German, French, and Italian
S17	English, Chinese (dialect), and Malay
S18	English, Spanish, and French
S19	Spanish, German, French, Italian, Portuguese, Dutch, and Turkish
S20	English, Spanish, German, Portuguese, and Dutch
S21	English, Spanish, German, French, Italian, Chinese, Russian, and Turkish
S22	English, Spanish, and French
S23	English, Spanish, German, and Russian
S24	English and Swedish
S25	English and Spanish
S26	English, Spanish, Italian, and Dutch
S27	English and Spanish
S28	English, German, and Russian
S29	English, German, and French
S30	Spanish, German, and French
S31	English, Spanish, Catalan, Basque, and Galician
S32	English, Spanish, German, French, and Italian
S33	English, Spanish, and Portuguese
S34	English, Spanish, German, French, Italian, Chinese, and Dutch
S35	English and Japanese
S36	English, Spanish, and Russian
S37	English, Spanish, German, French, and Italian
S38	English and Arabic
S39	English and Turkish
S40	Spanish, German, and French
S41	English, Spanish, German, and French
S42	English and Dutch
S43	English, Spanish, German, and Portuguese
S44	English and Romanian
S45	English, Spanish, German, French, Italian, Czech, and Hungarian

Which Pre-processing Techniques are Suitable to Extract Multilingual Sentiment Analysis?

A quantitative summary of the results for research questions RQ2 is shown in Table 7. The results presented in Table 7 reveal that machine translation (49%) and tokenization (42%) were the most common preprocessing techniques in multilingual sentiment analysis;

Table 7
Pre-processing technique in sentiment analysis

Processing	Study ID	Total	%
Tokenization	S1, S2, S4, S6, S12, S15, S16, S17, S18, S20, S22, S24, S25, S26, S32, S34, S37, S43, S44	19	**42%**
Stop word removal	S5, S7, S8, S11, S15, S29, S31, S43	8	**18%**
Stemming	S5, S8, S11, S15, S28	5	**11%**
N-gram	S2, S9, S12, S13, S15, S26, S27, S30, S32, S33, S38, S39, S40, S41, S44	15	**33%**
Lemmatization	S9, S11, S20, S26, S27, S31, S44	7	**16%**
POS tagging	S5, S9, S11, S20, S26, S27, S42, S43	8	**18%**
Noise removal	S7, S35	2	**4%**
Normalization	S12, S17, S18, S22, S25, S26, S32, S37, S43, S44	10	**22%**
Word Embedding / Text Vectors	S1, S4, S5, S10, S12, S16, S21, S30	8	**18%**
Capitalization	S1, S2, S12, S16, S17, S25, S28, S37, S44	9	**20%**
Negation	S8, S28	2	**4%**
Machine translation	S3, S13, S15, S19, S20, S21, S23, S24, S25, S26, S30, S32, S33, S34, S35, S36, S37, S39, S40, S41, S43, S45	22	**49%**

this was followed by n-gram (33%), normalization (22%) and capitalization (20%). Next, POS tagging (18%), word embedding/text vectors (18%), lemmatization (16%), stemming (11%), noise removal (4%), and negation (4%).

Table 8 shows several pre-processing techniques used for English, Chinese, Malay, and Hindi language. Among the articles that had focused on English language, tokenization was the most pre-processing technique used for English language (n=19), followed by machine translation (n=18), n-gram (n=13), normalization (n=10) and capitalization (n=9). Meanwhile, in Chinese language, tokenization (n=3), machine translation (n=3), word embedding/text vector (n=2), normalization (n=2) and capitalization (n=2) were the most preprocessing techniques used in the proposals. S17, the only proposal that focused primarily on the Malay language, had adopted tokenization, normalization, and capitalization in their proposed model. Lastly, S13 used n-gram and machine translation to process Hindi text.

Based on Table 8, the pre-processing techniques, which are commonly used for multilingual sentiment analysis, include tokenization, normalization, and capitalization. These three pre-processing techniques are used to extract the sentiment in the multilingual languages like English, Chinese and Malay. However, for the combination of languages that include English, Chinese and Hindi, the suitable pre-processing techniques are N-gram and machine translation.

Table 8
Number of preprocessing techniques used for English, Chinese, Malay and Hindi language

Pre-processing	English	Chinese	Malay	Hindi
Tokenization	S1, S2, S4, S6, S12, S15, S16, S17, S18, S20, S22, S24, S25, S26, S32, S34, S37, S43, S44	S12, S17, S34	S17	-
	Total: 19	**3**	**1**	**0**
Stopword removal	S5, S7, S8, S11, S15, S29, S31, S43	S7	-	-
	Total: 8	**1**	**0**	**0**
Stemming	S5, S8, S11, S15, S28	-	-	-
	Total: 5	**0**	**0**	**0**
N-gram	S2, S9, S12, S13, S15, S26, S27, S32, S33, S38, S39, S41, S44	S12	0	**S13**
	Total: 13	**1**	**-**	**1**
Lemmatization	S9, S11, S20, S26, S27, S31, S44	-	-	-
	Total: 7	**0**	**0**	**0**
POS tagging	S5, S9, S11, S20, S26, S27, S42, S43	-	-	-
	Total: 8	**0**	**0**	**0**
Noise removal	S7, S35	S7	-	-
	Total: 2	**1**	**0**	**0**
Normalization	S12, S17, S18, S22, S25, S26, S32, S37, S43, S44	S12, S17	S17	-
	Total: 10	**2**	**1**	**0**
Word embedding/ text vector	S1, S4, S5, S10, S12, S16, S21	S12, S21	-	-
	Total: 7	**2**	**0**	**0**
Capitalization	S1, S2, S12, S16, S17, S25, S28, S37, S44	S12, S17	S17	-
	Total: 9	**2**	**1**	**0**
Negation	S8, S28	-	-	-
	Total: 2	**0**	**0**	**0**
Machine translation	S13, S15, S20, S21, S23, S24, S25, S26, S32, S33, S34, S35, S36, S37, S39, S41, S43, S45	S3, S21, S34	-	**S13**
	Total: 18	**3**	**0**	**1**

What Sentiment Analysis Classification Methods are Available for Multilingual Domain?

Table 9 shows a quantitative summary of the results for research questions RQ3. From the results in Table 9, machine learning (51%, n=23) was the most common sentiment analysis technique for multi-language; followed by lexicon (38%, n=17) and hybrid technique (11%, n=5).

From the results in Table 10, we can conclude that machine learning was the leading sentiment analysis technique for English language (n=21), followed by lexicon (n=16) and hybrid technique (n=4). Meanwhile, machine learning, lexicon-based, and hybrid method were equally adopted by two articles as the sentiment analysis method for the Chinese

Table 9
Methods for multilingual sentiment analysis

Sentiment Analysis	Study ID	Total	%
Machine learning	S1, S2, S4, S6, S7, S8, S9, S10, S12, S13, S15, S18, S25, S26, S27, S29, S30, S32, S37, S39, S40, S41, S44	23	**51%**
Lexicon	S3, S11, S14, S16, S20, S21, S22, S23, S28, S31, S33, S35, S36, S38, S42, S43, S45	17	**38%**
Hybrid	S5, S17, S24, S19, S34	5	**11%**

Table 10
Summary of sentiment analysis methods for English, Chinese, Malay and Hindi language

Language	Sentiment Analysis	StudyID	Total
English	Machine Learning	S1, S2, S4, S6, S7, S8, S9, S10, S12, S13, S15, S18, S25, S26, S27, S29, S32, S37, S39, S41, S44	**21**
	Lexicon	S11, S14, S16, S20, S21, S22, S23, S28, S31, S33, S35, S36, S38, S42, S43, S45,	**16**
	Hybrid	S5, S17, S24, S34	**4**
Chinese	Machine learning	S7, S12	**2**
	Lexicon	S3, S21	**2**
	Hybrid	S17, S34	**2**
Malay	Machine Learning	-	**0**
	Lexicon	-	**0**
	Hybrid	S17	**1**
Hindi	Machine Learning	S13	**1**
	Lexicon	S3	**1**
	Hybrid	-	**0**

language. For Malay language, S17 had used the hybrid technique in the proposed model, while the Hindi language had adopted machine learning and lexicon-based approach.

Based on Table 10, the sentiment analysis classification technique for multilingual sentiment that involves English, Chinese, and Malay is hybrid sentiment analysis. The hybrid sentiment analysis processes include the localized language analysis, unsupervised topic clustering, and followed by the multilingual sentiment analysis. Meanwhile, for the combination of languages that includes English, Chinese and Hindi, the classification techniques can be either machine learning or lexicon-based techniques.

How to Evaluate the Sentiment Analysis Methods in Multilingual Sentiment Analysis?

A quantitative summary of the results for research questions RQ4 is presented in Table 11. The results described that accuracy (36%, n=16) was the most common evaluation model for multilingual sentiment analysis, followed by precision (22%, n=10). In contrast, recall

Table 11
Evaluation model for multilingual sentiment analysis

Evaluation Criteria	Study ID	Total	%
Precision	S2, S5, S7, S17, S18, S33, S35, S36, S44, S45	10	**22%**
Recall	S2, S5, S7, S14, S17, S32, S33,S36, S45	9	**20%**
F measure/ F1 score	S6, S10, S16, S32, S33, S36, S43	7	**16%**
Accuracy	S2, S4, S10, S12, S13, S14, S18, S22, S27, S29, S31, S34, S37, S38, S39, S43	16	**36%**

Table 12
Evaluation model for English, Chinese, Malay, Hindi and Arabic language

Language	Evaluation Model	Study ID	Total
English	Precision	S2, S5, S7, S17, S18, S33, S35, S36, S44, S45	**10**
	Recall	S2, S5, S7, S14, S17, S32, S33, S36, S45	**9**
	F measure/ F1 score	S6, S10, S16, S32, S33, S36, S43	**7**
	Accuracy	S2, S4, S10, S12, S13, S14, S18, S22, S27, S29, S31, S34, S37, S38, S39, S43	**16**
Chinese	Precision	S7, S17	**2**
	Recall	S7, S17	**2**
	F measure/ F1 score	-	**0**
	Accuracy	S12, S34	**2**
Malay	Precision	S17	**2**
	Recall	S17	**2**
	F measure/ F1 score	-	**0**
	Accuracy	-	**0**
Hindi	Precision	-	**0**
	Recall	-	**0**
	F measure/ F1 score	-	**0**
	Accuracy	S13	**1**

(20%, n=9) and F-measure (16%, n=7) were found to have the lowest number of articles used for the evaluation process.

Table 12 illustrates the evaluation model used for English, Chinese, Malay, and Hindi language. According to the results, we can summarize that accuracy is the most common evaluation model for English language (n=16). Instead, for Chinese language, precision and recall are employed to evaluate S7 and S17. Malay language in S17 had also used precision and recall model, while S13 had used accuracy model for the Hindi language.

Based on Table 12, to evaluate the sentiment analysis methods for multilingual sentiment in English, Chinese, and Malay, the evaluation models include precision and recall. For English, Chinese, and Hindi, the evaluation model is more on the accuracy.

From the literature, there is an issue in terms of the evaluation model selected for the analysis process. There is no specific approach to evaluate sentiment analysis specifically for multilingual. It means the evaluation model is selected without academic justification and the evaluation criteria are solely decided by the researcher (Alsaeedi, 2009). Although these evaluation models are broadly applied in practices, it is important to find a set of generic evaluation criteria in which they are capable to accommodate various languages without producing biases towards given datasets.

DISCUSSION

The growth of the internet and social media has given users to share their thoughts and opinions on all kinds of topic in different languages. Sentiment analysis in only one language could increase the risks of missing important information if the texts are written using a combination of other languages. Furthermore, most research on sentiment analysis focuses on text written in English, and there is a significant lack regarding the information sources for other languages. Consequently, most of the resources, such as sentiment lexicons and corpora, have been developed for the English language. An effective sentiment analysis approach should be able to handle a variety of languages so that it could easily detect the content or specific word in different languages and improves the overall classification of sentiment in the data.

While increasing effort has been made in creating resources for other formal languages, there are not many resources available when it comes to languages that are not commonly used in informal communication. It is well-known that different languages have their unique way of expression; for example, in S17, Singaporeans generally speak and write in English with some Chinese dialects and Malay language; which are certainly mixed with informal languages. Thus, it is important to note that, future research should not only cater to formal multilingual texts, but it should be possible to process texts in informal representation too.

Preprocessing method plays an important role to extract the relevant content and eliminate unnecessary words. From the analysis, machine translation is the most common preprocessing technique in multilingual sentiment analysis, followed by tokenization. The performance of machine translation systems such as Google Translate and Bing Translator has proven effective to provide accurate translation for most spoken languages. Machine translation, however, occasionally faced problems where it does not fully translate the texts, which can bring the risk of missing relevant content in texts (Balahur & Turchi, 2012a; Al-Kabi et al., 2013). Whereas tokenization is generally considered as an easy preprocessing technique compared to other techniques, however, it is important to note that tokenization process can be hard to implement to the text that does not have any whitespace or other characters. Some of the languages are classified as space-delimited; by means, the words are separated from each other's blank space, while languages such as Chinese, Japanese and Thai are referred as unsegmented word, where the languages do not have particular boundaries (Wang et al., 2017). Tokenization of unsegmented language can be difficult and would require additional technique or procedure. For example, instead of tokenized words using Chinese characters, S7 and S17 adapt the romanization of Chinese words to execute the tokenization process. Hence, researchers need to identify properly which preprocessing technique is suitable to process their target languages.

In terms of sentiment analysis technique, machine learning approaches for multi-language have been extensively studied due to their capability to adopt a variety of preprocessing techniques and features. Machine learning techniques like Naïve Bayes (NB), Support vector machine (SVM) and K-nearest neighbor (KNN) have been proven achieved great success in sentiment analysis. Among 23 literatures, S6, S13, S15, S29, S39 and S44 have adopted NB classifier in the proposal. The advantage of the Naïve Bayes classifier is that it requires a small amount of training data to estimate the parameters necessary for classification. However, empirical finding (Alsaleem, 2011; Hadi et al., 2010) indicated that SVM approaches are performed better compared to NB approach for multilingual sentiment analysis. Support vector machines (SVMs) are one of the classification methods that are well-known to be more accurate; thus, studies such as S6, S8, S12, S13, S15, S25, S30, S32, S37, S39, S40, and S41 have used SVM method in their proposal. Meanwhile, S13, S15, and S29 adopted KNN in their proposal. This method was said to be effective and easy to be implemented for multiple languages (Baro et al., 2019).

Choosing the right evaluation models is particularly important since the selection of the techniques can produce a potential positive or negative bias towards measuring the sentiment analysis characteristics. The analysis shows that accuracy is the most common method to measure the performance of sentiment analysis proposals due to its simplicity and straightforward process in generating the results. However, it should be noted that using accuracy model alone could be insufficient to ensure the results can be used solely as the indicator since it only yields a single number without describing the types of errors occur during the evaluation process. In addition, accuracy is highly affected by the imbalance number of instances in different classes (Dinsoreanu & Bacu, 2014; Al-Azani & El-Alfy, 2017). Thus, a combination of other evaluation models such as accuracy and precision, or accuracy and f-measure could draw the right conclusions on the performance of sentiment analysis models.

CONCLUSION

This paper has presented a systematic literature review of articles from 2010 to 2019, covering the aspects of common languages supported in multilingual sentiment analysis, pre-processing techniques, sentiment analysis approaches, and evaluation model that have been used to multilingual sentiment analysis. We have identified 45 primary studies that are related to the four research questions (RQs) in this review. A vast majority (31%) of the 45 articles include two languages multilingual sentiment analysis, and most of the studies (91%) introduced a combination of English with other languages (RQ1). The most preferred preprocessing technique for answering RQ2 is machine translation (49%), followed by tokenization (42%). Overall, 51% of articles used machine learning as the method in multilingual sentiment analysis (RQ3), and finally, it is found that most of articles (36%) are likely to use accuracy to evaluate their proposed method (RQ4).

ACKNOWLEDGEMENTS

The authors would like to thank Ministry of Higher Education (MOHE) Malaysia for sponsoring this paper under the grant of FRGS/1/2018/ICT04/UITM/02/9. A special appreciation for Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA for supporting the publication of this paper.

REFERENCES

Abdel-Hady, M., Mansour, R., & Ashour, A. (2014, August 24). Cross-lingual twitter polarity detection via projection across word-aligned corpora. In *Proceedings of WISDOM* (pp. 1-12). New York, USA.

Al-Azani, S., & El-Alfy, E. S. M. (2017). Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short arabic text. *Procedia Computer Science*, *109*, 359-366.

Al-Kabi, M. N., Hailat, T. M., Al-Shawakfa, E. M., & Alsmadi, I. M. (2013). Evaluating English to Arabic machine translation using BLEU. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *4*(1), 66-73.

Alsaeedi, A. (2019). EFTSA: Evaluation framework for Twitter sentiment analysis. *Journal of Software, 14*(1), 24-35. doi: 10.17706/jsw.14.1.24-35

Alsaleem, S. (2011). Automated Arabic text categorization using SVM and NB. *International Arab Journal of e-Technology*, *2*(2), 124-128.

Al-Shabi, A., Adel, A., Omar, N., & Al-Moslmi, T. (2017). Cross-lingual sentiment classification from English to Arabic using machine translation. *International Journal of Advanced Computer Science and Applications*, *8*(12), 434-440. doi: 10.14569/IJACSA.2017.081257

Araujo, M., Reis, J., Pereira, A., & Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing* (pp. 1140-1145). New York, USA: ACM. doi: https://doi.org/10.1145/2851613.2851817

Ardabili, S., Mosavi, A., & Várkonyi-Kóczy, A. R. (2019). Advances in machine learning modeling reviewing hybrid and ensemble methods. In *International Conference on Global Research and Education* (pp. 215-227). Cham, Switzerland: Springer. doi: https://doi.org/10.1007/978-3-030-36841-8_21

Argueta, C., Calderon, F. H., & Chen, Y. S. (2016). Multilingual emotion classifier using unsupervised pattern extraction from microblog data. *Intelligent Data Analysis*, *20*(6), 1477-1502. doi: 10.3233/IDA-140267

Bahrainian, S. A., & Dengel, A. (2013, December 3-5). Sentiment analysis and summarization of twitter data. In *2013 IEEE 16th International Conference on Computational Science and Engineering* (pp. 227-234). Sydney, Australia. doi: 10.1109/CSE.2013.44

Balahur, A., & Perea-Ortega, J. M. (2015). Sentiment analysis system adaptation for multilingual processing: The case of tweets. *Information Processing and Management*, *51*(4), 547-556. doi: https://doi.org/10.1016/j.ipm.2014.10.004

Balahur, A., & Turchi, M. (2012a, September 28). Comparative experiments for multilingual sentiment analysis using machine translation. In *Proceedings of the 1st International Workshop in Sentiment Discovery from Affective Data* (pp. 75-86). Bristol, UK.

Balahur, A., & Turchi, M. (2012b, July 12). Multilingual sentiment analysis using machine translation? In *Proceedings of the 3rd workshop in Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 52-60). Jeju, Republic of Korea.

Balahur, A., & Turchi, M. (2013, September 7-13). Improving sentiment analysis in twitter using multilingual machine translated data. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2013* (pp. 49-55). Hissar, Bulgaria.

Balahur, A., & Turchi, M. (2014). Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech and Language*, *28*(1), 56-75. doi: https://doi.org/10.1016/j.csl.2013.03.004

Balahur, A., Turchi, M., Steinberger, R., Ortega, J. M. P., Jacquet, G., Küçük, D., & El Ghali, A. (2014). Resource creation and evaluation for multilingual sentiment analysis in social media texts. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation 2014* (pp. 4265-4269). Reykjavik, Iceland: European Language Resources Association.

Baro, R. A., Pagudpud, M. V., Padirayon, L. M., & Dilan, R. E. (2019, February). Classification of project management tool reviews using machine learning-based sentiment analysis. In *IOP Conference Series: Materials Science and Engineering* (Vol. 482, No. 1, p. 012041). Bristol, UK: IOP Publishing. doi: https://doi.org/10.1088/1757-899X/482/1/012041

Becker, K., Moreira, V. P., & dos Santos, A. G. (2017a). Multilingual emotion classification using supervised learning: Comparative experiments. *Information Processing and Management*, *53*(3), 684-704. doi: https://doi.org/10.1016/j.ipm.2016.12.008

Becker, W., Wehrmann, J., Cagnini, H. E., & Barros, R. C. (2017b). An efficient deep neural architecture for multilingual sentiment analysis in twitter. In *Proceedings of the Thirtieth International Flairs Conference* (pp. 246-251). Palo Alto, California: AAAI Press.

Bhargava, R., & Sharma, Y. (2017, January 12-13). MSATS: Multilingual sentiment analysis via text summarization. In *Proceedings of 7th International Conference on Cloud Computing, Data Science and Engineering-Confluence 2017* (pp. 71-76). Noida, India. doi: 10.1109/CONFLUENCE.2017.7943126

Cruz, F. L., Troyano, J. A., Pontes, B., & Ortega, F. J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, *41*(13), 5984-5994. doi: https://doi.org/10.1016/j.eswa.2014.04.005

Cui, A., Zhang, M., Liu, Y., & Ma, S. (2011). Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis. In *Asia information retrieval symposium* (pp. 238-249). Heidelberg, Germany: Springer. doi: https://doi.org/10.1007/978-3-642-25631-8_22

Dadoun, M., & Olssson, D. (2016). *Sentiment classification techniques applied to swedish tweets investigating the effects of translation on sentiments from Swedish into English* (Degree Project). KTH Royal Institute of Technology, Stockholm, Sweden.

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., & Zhou, Q. (2016). Multilingual sentiment analysis: State of the art and independent comparison of techniques. *Cognitive computation*, *8*(4), 757-771. doi: https://doi.org/10.1007/s12559-016-9415-7

Demirtas, E., & Pechenizkiy, M. (2013). Cross-lingual polarity detection with machine translation. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining* (pp. 9-17). Chicago, USA: ACM. doi: https://doi.org/10.1145/2502069.2502078

Deriu, J., Lucchi, A., De Luca, V., Severyn, A., Müller, S., Cieliebak, M., … & Jaggi, M. (2017). Leveraging large amounts of weakly supervised data for multi-language sentiment classification. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1045-1052). Geneva, Switzerland: International World Wide Web Conferences Steering Committee. doi: https://doi.org/10.1145/3038912.3052611

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science*, *87*, 44-49. doi: https://doi.org/10.1016/j.procs.2016.05.124

Dinsoreanu, M., & Bacu, A. (2014, October 21-24). Unsupervised twitter sentiment classification. In *Proceedings of the International Conference on Knowledge Management and Information Sharing 2014* (pp. 220-227). Rome, Italy. doi: 10.5220/0005079002200227

Erdmann, M., Ikeda, K., Ishizaki, H., Hattori, G., & Takishima, Y. (2014). Feature based sentiment analysis of tweets in multiple languages. In *International Conference on Web Information Systems Engineering* (pp. 109-124). Cham, Switzerland: Springer. doi: https://doi.org/10.1007/978-3-319-11746-1_8

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of Language Resources and Evaluation (LREC 2006), 6*, 417-422.

Gînscă, A. L., Boroş, E., Iftene, A., TrandabĂţ, D., Toader, M., Corîci, M., & Cristea, D. (2011, June 24). Sentimatrix: Multilingual sentiment analysis service. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis* (pp. 189-195). Portland, Oregon, USA.

Hadi, W. E. M., Salam, M. A., & Al-Widian, J. A. (2010). Performance of NB and SVM classifiers in Islamic Arabic data. In *Proceedings of the 1st International Conference on Intelligent Semantic Web-Services and Applications* (pp. 1-6). New York, USA: ACM. doi: https://doi.org/10.1145/1874590.1874604

Hamouda, A., & Rohaim, M. (2011). Reviews classification using sentiwordnet lexicon. *The Online Journal on Computer Science and Information Technology (OJCSIT), 2*(1), 120-123.

Injadat, M., Salo, F., & Nassif, A. B. (2016). Data mining techniques in social media: A survey. *Neurocomputing, 214*, 654-670. doi: https://doi.org/10.1016/j.neucom.2016.06.045

Jing, T. W., & Murugesan, R. K. (2018). A theoretical framework to build trust and prevent fake news in social media using blockchain. In *International Conference of Reliable Information and Communication*

Technology (pp. 955-962). Cham, Switzerland: Springer. doi: https://doi.org/10.1007/978-3-319-99007-1_88

Kaity, M., & Balakrishnan, V. (2017, July 18). A multi-layered framework for building multilingual sentiment lexicons. In *Proceedings of the Postgraduate Research Excellence Symposium 2017* (pp. 29-34). Kuala Lumpur, Malaysia.

Kang, D., & Park, Y. (2014). Based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert Systems with Applications*, *41*(4), 1041-1050. doi: https://doi.org/10.1016/j.eswa.2013.07.101

Karima, A., & Smaili, K. (2016, September 29 - October 1). Measuring the comparability of multilingual corpora extracted from Twitter and others. In *Proceedings of the Tenth International Conference on Natural Language Processing* (HrTAL2016). Dubrovnik, Croatia.

Kaur, H., Mangat, V., & Krail, N. (2017). Dictionary based sentiment analysis of hinglish text. *International Journal of Advanced Research in Computer Science*, *8*(5), 816-822. doi: 10.26483/ijarcs.v8i5.3438

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report). Keele University, UK.

Lin, Z., Jin, X., Xu, X., Wang, W., Cheng, X., & Wang, Y. (2014a). A cross-lingual joint aspect/sentiment model for sentiment analysis. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management* (pp. 1089-1098). New York, USA: ACM. doi: https://doi.org/10.1145/2661829.2662019

Lin, Z., Jin, X., Xu, X., Wang, Y., Tan, S., & Cheng, X. (2014b, August 11-14). Make it possible: Multilingual sentiment analysis without much prior knowledge. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02* (pp. 79-86). Warsaw, Poland. doi: 10.1109/WI-IAT.2014.83

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2016). A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection. *Knowledge-Based Systems, 105*, 236-247. doi: https://doi.org/10.1016/j.knosys.2016.04.024

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2017a). Multilingual sentiment analysis: From formal to informal and scarce resource languages. *Artificial Intelligence Review*, *48*(4), 499-527. doi: https://doi.org/10.1007/s10462-016-9508-4

Lo, S. L., Chiong, R., & Cornforth, D. (2017b). An unsupervised multilingual approach for online social media topic identification. *Expert Systems with Applications*, *81*, 282-298. doi: https://doi.org/10.1016/j.eswa.2017.03.029

Lu, Y., & Mori, T. (2017). Deep learning paradigm with transformed monolingual word embeddings for multilingual sentiment analysis. *Computing Research Repository, 2017*, 1-10.

Maita, A. R. C., Martins, L. C., Lopez Paz, C. R., Peres, S. M., & Fantinato, M. (2015). Process mining through artificial neural networks and support vector machines: A systematic literature review. *Business Process Management Journal*, *21*(6), 1391-1415. doi: https://doi.org/10.1108/BPMJ-02-2015-0017

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis - A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, 16-32. doi: https://doi.org/10.1016/j.cosrev.2017.10.002

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning. *Neural and Statistical Classification, 13*(1994), 1-298.

Nowson, S., Perez, J., Brun, C., Mirkin, S., & Roux, C. (2015, September 8-11). XRCE personal language analytics engine for multilingual author profiling. In *Proceedings of the Working Notes of CLEF 2015 - Conference and Labs of the Evaluation Forum* (pp. 1412-1424). Toulouse, France.

Padmaja, S., & Fatima, S. S. (2013). Opinion mining and sentiment analysis-an assessment of peoples' belief: A survey. *International Journal of Ad hoc, Sensor & Ubiquitous Computing*, *4(1)*, 21-33. doi: 10.5121/ijasuc.2013.4102

Pappas, N., Redi, M., Topkara, M., Jou, B., Liu, H., Chen, T., & Chang, S. F. (2016). Multilingual visual sentiment concept matching. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (pp. 151-158). New York, USA: ACM. doi: https://doi.org/10.1145/2911996.2912016

Paramasivam, M., & Farashaiyan, A. (2016). Language change and maintenance of Tamil language in the multilingual context of Malaysia. *International Journal of Humanities and Social Science Invention, 5*(12), 55-60.

Patel, S., Nolan, B., Hofmann, M., Owende, P., & Patel, K. (2017). Sentiment analysis: Comparative analysis of multilingual sentiment and opinion classification techniques. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *11*(6), 565-571.

Pessutto, L. R. C., Vargas, D. S., & Moreira, V. P. (2018, December 3-6). Clustering multilingual aspect phrases for sentiment analysis. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 182-189). Santiago, Chile. doi: 10.1109/WI.2018.00-91

Pustulka-Hunt, E., Hanne, T., Blumer, E., & Frieder, M. (2018, August 27-29). Multilingual sentiment analysis for a swiss gig. In *2018 6th International Symposium on Computational and Business Intelligence* (ISCBI) (pp. 94-98). Basel, Switzerland. doi: 10.1109/ISCBI.2018.00028

Rajput, R., & Solanki, A. K. (2016). Review of sentimental analysis methods using lexicon-based approach. *International Journal of Computer Science and Mobile Computing*, *5*(2), 159-166.

Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation* (SemEval-2017) (pp. 502-518). Vancouver, Canada: Association for Computational Linguistics. doi: 10.18653/v1/S17-2088

Saad, M. K., Langlois, D., & Smaıli, K. (2013). Comparing multilingual comparable articles based on opinions. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora* (pp. 105-111). Sofia, Bulgaria: Association of Computational Linguistics.

Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of Advanced Computer Science and Applications, 9*(2), 273-281.

Saravia, E., Argueta, C., & Chen, Y. S. (2016). Unsupervised graph-based pattern extraction for multilingual emotion classification. *Social Network Analysis and Mining*, *6*(1), 1-21. doi: https://doi.org/10.1007/s13278-016-0403-4

Shalunts, G., & Backfried, G. (2015). SentiSAIL: Sentiment analysis in English, German, and Russian. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 87-97). Cham, Switzerland: Springer. doi: https://doi.org/10.1007/978-3-319-21024-7_6

Shalunts, G., & Backfried, G. (2016, October 9-13). Multilingual sentiment analysis on data of the Refugee crisis in Europe. In *Proceedings of the Fifth International Conference on Data Analytics 2016* (pp. 45-50). Venice, Italy.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *Australasian Joint Conference on Artificial Intelligence* (pp. 1015-1021). Heidelberg, Germany: Springer. doi: https://doi.org/10.1007/11941439_114

Steinberger, J., Lenkova, P., Kabadjov, M., Steinberger, R., & Van der Goot, E. (2011, September 12-14). Multilingual entity-centered sentiment analysis evaluated by parallel corpora. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* 2011 (pp. 770-775). Hissar, Bulgaria.

Sykes, L. M., Evans, W. G., Buchanan, G., Warren, N., & Fernandes, N. (2018). To pen or to probe. Prescribing versus treating, how to decide. *South African Dental Journal, 73*(1), 53-55.

Tellez, E. S., Miranda-Jiménez, S., Graff, M., Moctezuma, D., Suárez, R. R., & Siordia, O. S. (2017). A simple approach to multilingual polarity classification in Twitter. *Pattern Recognition Letters*, *94*, 68-74. doi: https://doi.org/10.1016/j.patrec.2017.05.024

Thakkar, H., & Patel, D. (2015). Approaches for sentiment analysis on twitter: A state-of-art study. *Computing Research Repository, 2015*, 1-8.

Tromp, E., & Pechenizkiy, M. (2011, December 11). Senticorr: Multilingual sentiment analysis of personal correspondence. In *2011 IEEE 11th International Conference on Data Mining Workshops* (pp. 1247-1250). Vancouver, Canada. doi: 10.1109/ICDMW.2011.152

Tsai, C. F., & Wang, S. P. (2009, March 18-20). Stock price forecasting by hybrid machine learning techniques. In *Proceedings of The International Multiconference of Engineers and Computer Scientists* (Vol. 1, No. 755, pp. 60-66). Hong Kong, China.

Vīksna, R., & Jēkabsons, G. (2018). Sentiment analysis in Latvian and Russian: A survey. *Applied Computer Systems, 23*(1), 45-51. doi: https://doi.org/10.2478/acss-2018-0006

Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2015, September 17). Sentiment analysis on monolingual, multilingual and code-switching twitter corpora. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 2-8). Lisboa, Portugal.

Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2017). Supervised sentiment analysis in multilingual environments. *Information Processing and Management*, *53*(3), 595-607. doi: https://doi.org/10.1016/j.ipm.2017.01.004

Vilares, D., Peng, H., Satapathy, R., & Cambria, E. (2018, November 18-21). BabelSenticNet: A commonsense reasoning framework for multilingual sentiment analysis. In *2018 IEEE Symposium Series on Computational Intelligence* (SSCI) (pp. 1292-1298). Bangalore, India. doi: 10.1109/SSCI.2018.8628718

Volkova, S., Wilson, T., & Yarowsky, D. (2013, October 18-21). Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1815-1827). Washington, USA.

Wang, X., Li, J., Yang, X., Wang, Y., & Sang, Y. (2017, October 21-23). Chinese text sentiment analysis using bilinear character-word convolutional neural networks. In *Proceedings of International Conference on Computer Science and Application Engineering* (CSAE 2017) (pp. 36-43). Shanghai, China. doi: 10.12783/dtcse/csae2017/17466

Wehrmann, J., Becker, W. E., & Barros, R. C. (2018). A multi-task neural network for multilingual sentiment classification and language detection on Twitter. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (pp. 1805-1812). New York, USA: ACM. doi: https://doi.org/10.1145/3167132.3167325

Yadav, V., & Elchuri, H. (2013, June 14-15). Serendio: Simple and practical lexicon-based approach to sentiment analysis. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 543-548). Atlanta, Georgia.

Zhou, Y., Demidova, E., & Cristea, A. I. (2016, April). Who likes me more?: Analysing entity-centric language-specific bias in multilingual Wikipedia. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing* (pp. 750-757). New York, USA: ACM. doi: https://doi.org/10.1145/2851613.2851858