Text Summarization: A Brief Review

Chapter · January 2020

DOI: 10.1007/978-3-030-34614-0_1

4 authors, including:

Laith Abualigah
Al al-Bayt University
551 PUBLICATIONS **19,759** CITATIONS
SEE PROFILE

Hamzeh Mohammd Alabool
Saudi Electronic University
27 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Mohammad Shehab
Amman Arab University
65 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Text Summarization: A Brief Review

Laith Abualigah, Mohammad Qassem Bashabsheh, Hamzeh Alabool and Mohammad Shehab

Abstract Text Summarization is the process of creating a summary of a certain document that contains the most important information of the original one, the purpose of it is to get a summary of the main points of the document. Abstractive summarization of multi-documents aims to generate a concentrated version of the document while keeping the main information. Due to the massive amount of data these days, the importance of summarization arose. Finally, this paper collects the most recent and relevant research in the field of the text summarization to study and analysis for future research. It will be significant by giving a new direction to who are interested in this domain in the future.

Keywords Text summarization · Documents and processing · Main information · Arabic text summarization

1 Introduction

There is a huge amount of data surfacing digitally, therefore the importance of developing a punctuate procedure to shorten long texts immediately while keeping the main idea of it is necessary [1]. Summarization also helps shorten the time needed for reading, fasten the search for information and help to get the most amount of information on one topic [2, 3].

The central object of computerized text summarization is decreasing the reference text into a smaller version maintaining its knowledge alongside with its meaning. Several descriptions of text summarization are provided, for example [4] explained the report as text that is generated from one or more documents that communicate relevant knowledge in the first text, and that is no higher than half of the primary text(s) and usually significantly more limited than that.

L. Abualigah (✉) · M. Q. Bashabsheh · H. Alabool · M. Shehab
Faculty of Computer Sciences and Informatics,
Amman Arab University, Amman 11953, Jordan
e-mail: lx.89@yahoo.com

M. Abd Elaziz et al. (eds.), *Recent Advances in NLP: The Case of Arabic Language*, Studies in Computational Intelligence 874,
https://doi.org/10.1007/978-3-030-34614-0_1

As of late, content mining has turned into a fascinating exploration field because of the colossal measure of the existing content on the web [5]. Content mining is a basic field with regards to information digging for finding fascinating examples with regards to literary information. Inspecting and extricating of such data designs from enormous datasets is considered as an essential procedure [6, 7]. Many review studies were led to utilize different content digging strategies for unstructured datasets. It has been seen that the far-reaching overview considers in the Arabic setting were ignored. Belkebir and Guessoum [8] gave a wide survey of different examinations identified with Arabic content mining with more spotlight on the Holy Quran, assessment investigation, and web records. Moreover, the combination of the exploration issues and philosophies of the reviewed investigations will help the content mining researchers in seeking after their future examinations.

The study [9] determined text summarization as a small but true description of the contents of a text and according to [10], text summarization can be described as the manner of providing a shorter display of the most powerful information of a source or multiple references of information according to special demands. Text summarization techniques can be classified based on various models as explained in Fig. 1.

Text Summarization is the process of creating a summary of a certain document that contains the most important information of the original one, the purpose of it is to get a brief summary of the main points of the document [11]. Abstractive summarization of multi-documents aims to generate a concentrated version of the

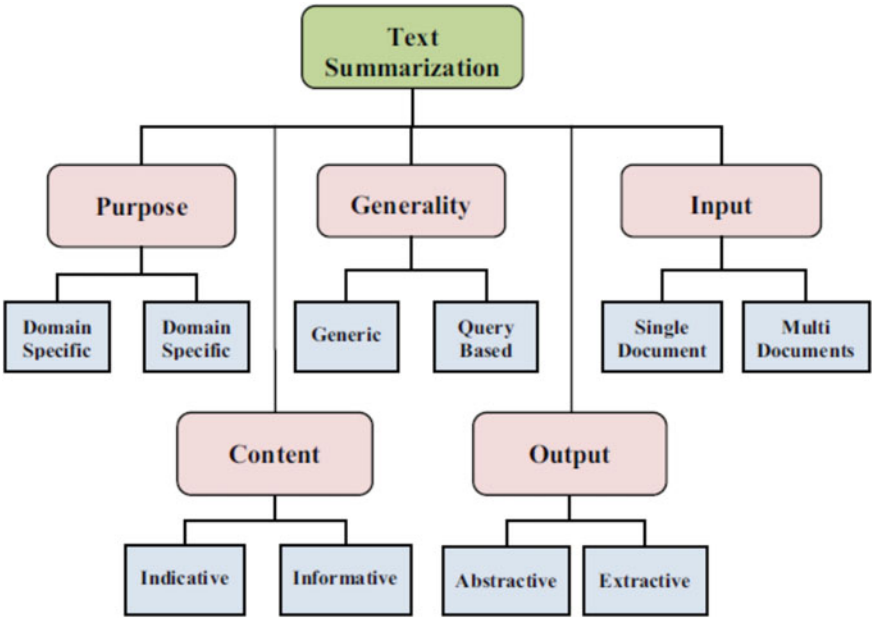

Fig. 1 Text summarization techniques [4]

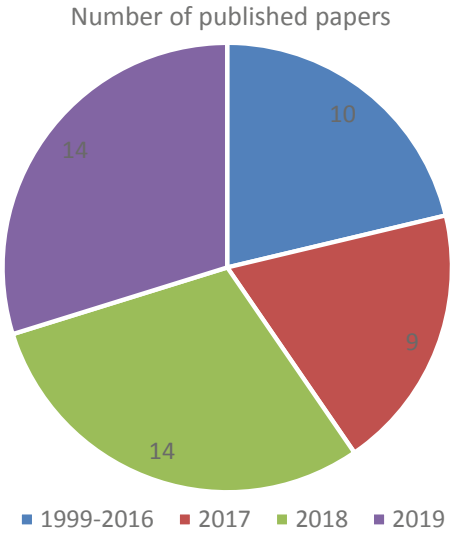

Fig. 2 The number of papers in the domain of text summarization per year in between (1999–2019), which taken in this study

document while keeping the main information [4]. Due to the massive amount of data these days, the importance of summarization arose [12].

Text summarization is a recent learning topic that caught attention rapidly, as research increase, we are hoping to witness a breakthrough that will affect this by providing a punctual method in summarizing long texts. In this paper, we provide a comprehensive review of text summarization techniques to draw attention to their importance in handling large data and to help researchers to use them to solve problems. Figure 2 shows the number of papers in the domain of text summarization per year in between (1999–2019), which taken in this study.

This paper is organized into three sections: Sect. 1 presented the introduction and organization of this paper for the text summarization domain. Section 2 Related Works shows the most related research in that domain. In Sect. 3 the related works are discussed and overviewed the used techniques. Finally, Sect. 3 concluded the paper to find the possible direction for the future researchers who are interested in the text summarization field.

2 Related Works

In this section, there are several papers in text summarization have been reviewed as following.

2.1 General Text Summarization

Because of the huge addition of information on the web, removing the most significant information as a reasonable brief would be profitable for specific clients. Accordingly, there is an enormous energy concerning the age of programmed content synopsis structures to establish abstracts consequently from the content, web, and interpersonal organization messages related with their satellite substance. This review features, just because, how the swarm intelligence (SI) advancement procedures are performed to settle the content summarization task proficiently [13–16].

Also, a persuading avocation regarding why SI, particularly Ant Colony Optimization (ACO), has been introduced [17]. Lamentably, three sorts of content synopsis undertakings utilizing SI show bit using in the writing when stood out from the other summarization strategies as AI and hereditary calculations, despite the way that there are genuinely encouraging results of the SI techniques. Then again, it has been seen that the summarization task with different kinds has not been formalized as a multi-target streamlining task previously, in spite of that there are numerous goals which can be considered. Additionally, the SI was not utilized before to help the continuous synopsis draws near. Subsequently, another model has been proposed to be satisfactory for accomplishing numerous targets and to fulfill the constant needs. In the long run, this investigation will enthuse analysts to further consider the different sorts of SI when tackling the synopsis assignments, especially, in the short content outline field.

There is a huge amount of data surfacing digitally, therefore the importance of developing a punctuate procedure to shorten long texts immediately while keeping the main idea of it is necessary [12]. Summarization also helps shorten the time needed for reading, fasten the search for information and help to get the most amount of information on one topic [18]. Text summarization is a fascinating learning topic that caught attention rapidly, as research increase, they are hoping to witness a breakthrough that will affect this by providing a punctual method in summarizing long texts [12].

The surge of information available through the internet and social networks and information technologies make the need of summarization more urgent, especially with the massive amount of data that is being spread due to the knowledge transfer among its users, which makes it difficult to differentiate between the right information from the wrong ones. To surmount the issues of data blast, a tool that can summarize these massive amounts of information has become a need. The procedure of summarization decreases the effort and time required to distinguish the most notable and important sentences. Generally, a summary can be described as content that is made from at least one message that passes on the most essential data in the first text while being adequately short [19].

Over the last two decades, the text summarization task has gained more significance due to the vast amount of online information, and its capability to extract helpful data and information in a manner that could be effectively taken care of people and utilized for a heap of purposes, including skilled systems for text appraisal.

This paper exhibits a programmed procedure for content evaluation that depends on fuzzy principles on an assortment of extracted information to locate the most critical data in the surveyed texts; the consequently delivered summaries of these texts are contrasted with reference summaries made by domain specialists. In contrast to different writing methods in the literature, this method summarizes manuscripts by researching correlated highlights to diminish dimensionality, and thus the number of fuzzy standards utilized for text summarization. Thus, the proposed methodology for content summarization with a moderately small number of fuzzy rules can profit advancement and utilization of future expert systems able to evaluate writing [20].

Strategies to automatically summarize, link and evaluate information have become more imperative as progressively substantial literary datasets have been gathered and made accessible by an assortment of (digital libraries (DLs), Virtual Learning Environments (VLEs), social media). Recently, Natural Language Processing (NLP) and Information Extraction (IE) techniques have been proposed to figure the comparability between free writings at low computational expenses [21].

Grammatical feature tagging is the procedure of consequently deciding the best possible linguistic tag or syntactic classification of a word contingent upon its specific situation. Part-of-Speech tagging is a fundamental advance in most Natural Language Processing (NLP) applications, for example, content summarization, question answering, data extraction and data recovery. A productive labeling approach for the Arabic language is proposed in [17] utilizing Bee Colony Optimization calculation. The issue is spoken to as a chart and a novel procedure is proposed to dole out scores to potential labels of a sentence, at that point the honey bees locate the best arrangement way. The proposed methodology is assessed utilizing KALIMAT corpus, which comprises of 18 M words. Trial results demonstrated that the proposed methodology accomplished 98.2% of exactness contrasted with 98, 97.4 and 94.6% for Hybrid, Hidden Markov Model and Rule-Based strategies separately. Moreover, the proposed methodology decided every one of the labels introduced in the corpus while the referenced methodologies can recognize just three labels. Other optimization techniques can be used [22–27].

Text summarization creates summaries from input documents while keeping striking data. It is an imperative task and can be connected to a few true applications. Numerous strategies have been proposed to take care of the content summarization issue [28, 29]. There are two principles for the text summarization systems: extractive and abstractive. Extractive summarization creates synopsis by choosing remarkable sentences or expressions from the source content, while abstractive strategies reword and rebuild sentences to form the summary. They focus around abstractive summarization in this work as it is increasingly adaptable and hence can create progressively different summaries [30].

Recent neural system ways to deal with an outline are generally either sentence-extractive, choosing a lot of sentences as the summary, or abstractive, creating the summary from a seq 2seq model. In this work [31], they present a neural model for single-record summary dependent on joint extraction and pressure. Following later fruitful extractive models, they outline the summarization issues as a progression of local choices. This model picks sentences from the report and after that chooses

which of a set of compression options to apply to each chosen sentence. They compute this arrangement of discrete compression rules dependent on syntactic constituency parses; however, the proposed methodology is measured and it could utilize for any accessible source of compressions. For learning, they build oracle extractive-compressive summaries that reflect vulnerability over the proposed model's decision sequence, and then learn both of these parts together with this supervision. Test results on the CNN/Daily Mail and New York Times datasets demonstrate that this model accomplishes the innovative execution on substance determination assessed by ROUGE. Besides, human and manual assessment demonstrate that the proposed model's yield, for the most part, remains.

Recent days, a large portion of the client inquiries are of the complex in nature because of an increased need for dynamic information [32]. The overwhelming questions are replied with a summarization of applicable sentences from the web archives. These can be responded to by network question answering (CQA, for example, Quora, stack flood, Yahoo! Answers and so forth for different client levels utilizing AI methods. Answers can be created from different sources and summarized from a list of sentences [33] proposes a novel summarization technique which centers around programmed machine-produced summaries. Content summaries are the method utilized for condensing content with special, pertinent highlights in compression ranges.

The learning model is prepared with benchmark datasets 20 newsgroup and DUC2001 using machine learning algorithms. The examinations are completed and checked with standard measurements, for example, ROUGE for the results [34]. Primary categorization of text summarization techniques is based on the type of summary generated. It can either be of extractive or abstractive type. Generating abstractive summary is cumbersome as it gives a summary with sentences different from the original document, though the meaning of information is preserved. On the other hand, Extractive text summarization uses sentences from the document to provide a condensed form of the document that is in simple terms, it is the subset of the actual document. Most of the studies on text summarization are on extractive techniques [35].

Because of countless records accessible on the web, messages and advanced libraries, archive grouping is turning into a significant assignment incredibly required. It is ordinarily accomplished in the wake of performing highlight choice that comprises of choosing proper highlights to improve the grouping exactness. The vast majority of the component determination put together content order strategies depend with respect to building a term-frequency converse record frequency include vector, which is not typically proficient. Likewise, various archive order studies are centered on the English language. Garg and Saini [36] proposed an Arabic Text Classification technique, which is not seriously examined because of the multifaceted nature of the Arabic language. Another firefly calculation-based component choice strategy is proposed. This calculation has been effectively connected to various combinatorial issues. In any case, it has not been associated with highlight choice idea to manage Arabic Text Classification. To approve this method, Support Vector Machine classifier is utilized just as three assessment measures including exactness, review, and F-measure. Besides, probes OSAC genuine dataset alongside an examination

with the innovative strategies are performed. The proposed strategy accomplishes an exactness worth equivalent to 0.994. The outcomes affirm the effectiveness of the proposed highlight choice technique in improving Arabic Text Classification precision.

The objective of the individual in summarizing any text is to get objective and concise ideas around it. It depending on the audience that wrote the text for it, in addition to the size of the original text, and there are many reasons that urge the individual to summarize a text: Summary for study, where the individual works to write important points in the text in order to learn the required material quickly. Summarizing the text helps readers read and understand it. The abstract is also used to write papers and academic papers to review and summarize the information written about the subject. There are grounds to consider when summarizing any text: Mention all important concepts and information in the text, and do not mention any non-important information. Do not repeat any information, delete all duplicate information. Replace difficult terms with simpler terms. Select and write the main idea in the text. Examples of summarizing the text (summarizing the text of the Indian language and the text on foreign languages [37].

2.2 Arabic Text Summarization

Arabic text clustering is an essential job for getting high-grade results with the common Information Retrieval (IR) systems particularly with the fast growth of the number of documents existing in the Arabic language. Documents clustering tries to automatically group related documents in one cluster working various similarity/distance criteria [38]. The length of the document often influences this job, useful knowledge on the documents is often followed by a large amount of noise, and since it is important to reduce this noise while having useful knowledge to increase the achievement of text documents clustering.

In this paper [39], they suggested assessing the impact of text summarization utilizing the Latent Semantic Analysis Model in order to determine problems indicated above, using 5 similarity/distance criteria without and with stemming. The experimental results show that the proposed method completely explains the problems of noisy knowledge and documents time, and thus significantly enhance clustering production.

Marie-Sainte and Alalyani [1] proposed the utilization of Firefly calculation for the extraction of summarization of single Arabic records. The proposed methodology is contrasted and two transformative methodologies that utilization hereditary calculations and concordance search. The EASC Corpus and the ROUGE toolbox are utilized for the assessment of the proposed methodology. Trial results demonstrated that the proposed methodology accomplished focused and significantly higher ROUGE scores in correlation with the two innovative draws near.

Content synopsis is the way toward creating a shorter variant of particular content. Programmed synopsis methods have been connected to different spaces, for example,

therapeutic, political, news, and legitimate areas demonstrating that adjusting area applicable highlights could improve the outline execution. Regardless of the presence of a lot of research work in the area-based synopsis in English and different dialects, such work is absent in Arabic because of the lack of existing information bases.

In this paper [4], a half-breed, single-report content summarization approach is introduced. The methodology consolidates area information, factual highlights, and hereditary calculations to extricate significant purposes of Arabic political records. The ASDKGA approach is tried on two corpora KALIMAT corpus and Essex Arabic Summaries Corpus (EASC). The Recall-Oriented Understudy for Gisting Evaluation (ROUGE) structure was utilized to analyze the naturally produced synopses by the ASDKGA approach with outlines created by people. Likewise, the methodology is thought about against three other Arabic content outline draws near. This proposed approach showed promising outcomes when abridging Arabic political records with a normal F-proportion of 0.605 at the pressure proportion of 40%.

The exponential development of online literary information set off the critical requirement for a compelling and useful asset that consequently gives the ideal substance in an abridged structure while safeguarding center data.

In this paper [5], a programmed, nonexclusive, and extractive Arabic single archive outlining strategy is proposed, which goes for delivering an adequately enlightening summarization. The proposed extractive strategy assesses each sentence dependent on a mix of measurable and semantic highlights in which a novel definition is utilized considering sentence significance, inclusion, and assorted variety. Further, two abridging procedures including score-based and managed AI were utilized to create the summarization and afterward help to utilize the structured highlights. They exhibit the viability of the proposed strategy through a lot of investigations under EASC corpus utilizing ROUGE measure. Contrasted with some current related work, the test assessment demonstrates the quality of the proposed technique as far as exactness, review, and F-score execution measurements.

Arabic Text report bunching is a significant perspective for giving theoretical route and perusing systems by arranging monstrous measures of information into few characterized groups. In any case, Words as the vector are utilized for grouping techniques is regularly inadmissible as it overlooks connections between significant terms. Bunch examination isolates information into gatherings on groups for improved comprehension or summarization. Grouping has a long history and numerous strategies created in measurements, information mining, design acknowledgment and different fields. This examination proposes three approaches; Unsupervised, Semi Supervised procedures and Semi Supervised with dimensionality decrease to develop a grouping-based classifier for Arabic content archives.

Utilizing k-means, steady k-means Threshold + k-means and k- means with dimensionality decrease, after report preprocessing evacuating stop words and gets the root for each term in each record [40]. At that point, apply a term weighting technique to get the heaviness of each term concerning its record. At that point, apply a similitude measure technique to each report and its likeness with different archives. In addition, utilizing F-measure, entropy and bolster vector machine for figure precision. The datasets are online dynamic datasets that are portrayed by its

accessibility and believability on the web. Arabic language is a difficult dialect when connected in a deduction-based calculation. In this way, choosing the fitting dataset is a chief factor in such research. The precision of those techniques contrasted and different methodologies and the proposed strategies shows better exactness and less blunders for new arrangement experiments. Taking into account that the measurement decrease procedure is exceptionally delicate because expanding the proportion of decrease can crush significant terms.

Text summarization is one of the most testing and troublesome undertakings in normal language handling, and computerized reasoning even more largely. Different methodologies have been proposed in the writing. Content synopsis is ordered into two classes: extractive content outline and abstractive content summarization. Most by far of work in the writing pursued the extractive methodology, likely because of the intricacy of the abstractive one. As far as we could know, the work displayed here is the main work on Arabic that handles both the extractive and abstractive angles. To be sure, while the writing needs synopsis structures that permit the mix of different tasks inside a similar framework. It additionally gives a component that permits the task of the appropriate activity to each part of the source content, which is to be condensed, and this is accomplished in an iterative procedure.

Mosa et al. [12] examined an element based synopsis for digging sentiment for Arabic audits so as to produce an outline that co contains a lot of positive surveys to a specific element, just as negative audits for a similar component. This examination fundamentally relies upon the Natural Language Processing, beginning with extricating an element for a particular area, at that point estimation arrangement and a while later it condenses these audits as per the highlights.

Conventional Arabic content outline frameworks depend on pack of-words portrayal, which includes meager and high-dimensional information. In this manner, dimensionality decrease is enormously expected to expand the intensity of highlights separation. In this paper, they present another strategy for ATS utilizing variation auto-encoder (VAE) model to take in an element space from high-dimensional information.

Alami et al. [40] investigated a few information portrayals, for example, term frequency (tf), tf-idf and both neighborhood and worldwide vocabularies. All sentences are positioned by the inert portrayal created by the VAE. They examine the effect of utilizing VAE with two summarization draws near, chart-based, and question-based methodologies. Analyses on two benchmark datasets explicitly intended for ATS demonstrate that the VAE utilizing tf-idf portrayal of worldwide vocabularies gives an increasingly discriminative component space and improves the review of different models. Test results affirm that the proposed technique prompts preferred execution over a large portion of the best in class extractive synopsis approaches for both charts based and inquiry-based outline draws near.

Computerized summarization help handle the consistently developing volume of data coasting around. There are two general classes: concentrate and conceptual. In the previous, they hold the more significant sentences pretty much in their unique structure, while the last requires a combination of numerous sentences as well as summarizing. This is a more testing errand than concentrate summarization.

In this paper [41], a novel conventional theoretical summarizer is displayed for a solitary report in the Arabic language. The framework begins by dividing the info content subject shrewd. At that point, each literary section is extractive abridged. Finally, they apply rule-based sentence decrease procedure. The RST-based extractive summarizer is an improved variant of the framework. By controlling the size of the extricated outline of each fragment, they can top the size of the last abstractive synopsis. Both summarizers, the improved extractive and the abstractive, were assessed. They tried to improve extractive summarizer on the equivalent dataset in the previously mentioned paper, utilizing the measures review, accuracy, and Rouge. The outcomes show detectable improvement in the exhibition, particularly the accuracy in shorter summarization. The abstractive summarizer was tried on a lot of 150 reports, creating synopses of sizes half, 40, 30 and 20% (of the first's statement check). Two human specialists who reviewed them out of the extreme score of five evaluated the outcomes. The normal score extended somewhere in the range of 4.53 and 1.92 for synopses at various granularities, with shorter outlines accepting the lower score. The trial results are empowering and exhibit the viability of the proposed methodology.

Supposition mining applications work with countless conclusion holders. This implies a summarization of assessments is significant so we can without much of a stretch translate holders' sentiments. This paper expects to give a component-based synopsis to Arabic audits. In this work [42], a framework is proposed utilizing Natural Language Processing (NLP) procedures, data extraction, and slant vocabularies. This gives clients to get to the feelings communicated in many surveys in a brief and valuable way. They begin with extricating highlight for a particular space, appointed opinion characterization to each component, and afterward abridged the audits. They led many trials to assess the proposed framework by utilizing information corpus from the lodging area. The exactness for conclusion mining they determined utilizing target assessment was 71.22%. They, likewise, connected emotional assessment for the synopsis age and it showed that the proposed framework accomplished an important proportion of 73.23% precision for positive outline and 72.46% exactness for a negative summarization.

3 Discussions and Overview

There are many challenges facing text summarization. Table 1 shows some details of the text summarization studies that conducted to handle the problem of text extraction and abstraction. Notice that the majority of reviewed studies focused on the problem of text extraction over abstraction. However, text abstraction is more difficult than extraction since it refers to the reformulated version of the original text. While text extraction studies aim to extract a copy of some sentences of the original text. In addition, the majority of the reviewed studies are proposed a singular method (e.g., Fuzzy-based method, Score-based method, Cluster-based method, Semantic-based method, Optimization-Based method, and Machine learning-based method) over proposing hybrid method.

Table 1 studies on methods for text summarization

References	Year	Output	Method	Language	Input Type	Validation	Compared with	Evaluation
[20]	2019	Extractive	Fuzzy method	Brazilian Portuguese texts	Multi documents	Experiment	Naive baseline, score, model and sentence	ROUGE
[38]	2012	Abstractive	Fuzzy method	English	Multi documents	Example	N/A	N/A
[32]	2018	Abstractive	Neural network-based methods/bottom-up attention	English	Single document	Experiment	Graph based attention neural model and	ROUGE
[19]	2013	Abstractive	Latent Semantic Analysis	Arabic	Single document	Experiment	N/A	N/A
[4]	2018	Extractive	Hybrid approach (domain knowledge, statistical features, and genetic algorithms)	Arabic	Single document	Experiment	Graph-based approach, statistical-based approach), and lexical cohesion and text entailment	Human and ROUGE
[34]	2019	Extractive	Hybrid approach (score-based method and machine-learning method.)	Arabic	Single document	Experiment	Score-based, cluster-based, semantic-based, optimization-based and machine learning based	ROUGE
[5]	2017	Extractive	β-Hill climbing search	English	Single document	Experiment	Original hill climbing search	P, R and F1

(continued)

Table 1 (continued)

References	Year	Output	Method	Language	Input Type	Validation	Compared with	Evaluation
[42]	2017	Abstractive	Rule-based sentence reduction technique	Arabic	Single document	Experiment	Ikhtasir approach	Human and ROUGE
[12]	2018	Extractive	Machine learning based	Arabic	Single document	Experiment	N/A	Human and P, R and F1
[10]	2019	Extractive	Firefly algorithm based feature selection method	Arabic	Single document	Experiment	InfoGain, OneR and TFIDF.	P, R and F1
[40]	2018	Extractive	Variation auto-ncoder (VAE) model	Arabic	Single document	Experiment	Latent semantic analysis (LSA), LexRank, Graph-based AE, TextRank, baseline Tf.ISF	Human and ROUGE

The hybrid method might achieve better results in terms of recall, precision, and F-measure than using one single method for handling text summarization. This is because the process of text summarization is a multi-dimensional problem that combines: segmentation and tokenization, word and sentence scoring, text evaluation, Cosine Similarity Matrix, identify the optimal sentence combination, which needs to be handled by different methods. Furthermore, most of the proposed methods are used to extract text from single documentation over multi documentation that could provide important information about the original text. Finally, two evaluation methods are used: human-based and automatic based (e.g., ROUGE). Studies should tack human evaluation in consideration since this approach is that it assesses coherence and informatively of the summary compared to the original text.

4 Conclusions

Due to the tremendous increment of data on the web, extracting the most important data as a conceptual brief would be valuable for certain users. Because of the large volume of data deployed in digital space, there is a need to find a way to shorten texts and provide clear summaries. Summarizing the texts is still active in several research and needs further research and development in summarizing the texts Due to the huge increase in data on the web, extracting the most important data as a conceptual summary will be useful to some researchers. In this paper, we drew the reader to the latest and most important data problems and the need to summarize the texts and explained the utility of short texts while preserving the original texts. This paper provides a summary of future research findings in this area. Finally, the optimization way can be used to deal with different problems.

References

1. S.L. Marie-Sainte, N. Alalyani, Firefly algorithm-based feature selection for Arabic text classification. J. King Saud Univ.-Comput. Inf. Sci. (2018)
2. L.M. Abualigah, A.T. Khader, M.A. Al-Betar, Multi-objectives-based text clustering technique using K-mean algorithm, in *2016 7th International Conference on Computer Science and Information Technology (CSIT)* (IEEE, 2016), pp. 1–6
3. A.M. Azmi, N.I. Altmami, An abstractive Arabic text summarizer with user-controlled granularity. Inf. Process. Manage. **54**(6), 903–921 (2018)
4. D.R. Radev, E. Hovy, K. McKeown, Introduction to the special issue on summarization. Comput. Linguist. **28**(4), 399–408 (2002)
5. A. Qaroush, I.A. Farha, W. Ghanem, M. Washaha, E. Maali, An efficient single document Arabic text summarization using a combination of statistical and semantic features. J. King Saud Univ.-Comput. Inf. Sci. (2019)
6. J. Xu, G. Durrett, Neural extractive text summarization with syntactic compression. arXiv preprint arXiv:1902.00863 (2019)
7. L.M.Q. Abualigah, Feature selection and enhanced krill herd algorithm for text document clustering. studies in computational intelligence (2019)

8. R. Belkebir, A. Guessoum, TALAA-ATSF: a global operation-based arabic text summarization framework, *Intelligent Natural Language Processing: Trends and Applications* (Springer, Cham, 2018), pp. 435–459
9. H. Saggion, Automatic summarization: an overview. Revue française de linguistique appliquée **13**(1), 63–81 (2008)
10. R.Z. Al-Abdallah, A.T. Al-Taani, Arabic text summarization using firefly algorithm, in *2019 Amity International Conference on Artificial Intelligence (AICAI)*. (IEEE, 2019), pp. 61–65
11. X. Carreras, L. Màrquez, Introduction to the CoNLL-2004 shared task: Semantic role labeling, in *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004* (2004), pp. 89–97
12. M.A. Mosa, A.S. Anwar, A. Hamouda, A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms. Knowl.-Based Syst. **163**, 518–532 (2019)
13. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, Hybrid clustering analysis using improved krill herd algorithm. Appl. Intell. (2018)
14. L.M. Abualigah, A.T. Khader, M.A. Al-Betar, O.A. Alomari, Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Syst. Appl. **84**, 24–36 (2017)
15. L.M. Abualigah, A.T. Khader, M.A. AlBetar, E.S. Hanandeh, Unsupervised text feature selection technique based on particle swarm optimization algorithm for improving the text clustering, in *Eai International Conference on Computer Science and Engineering* (2017)
16. L.M. Abualigah, A.T. Khader, M.A. Al-Betar, Z.A.A. Alyasseri, O.A. Alomari, E.S. Hanandeh, Feature selection with β-hill climbing search for text clustering application, in *2017 Palestinian International Conference on Information and Communication Technology (PICICT)* (IEEE, 2017), pp. 22–27
17. A. Alhasan, A.T. Al-Taani, POS tagging for arabic text using bee colony algorithm. Procedia Comput. Sci. **142**, 158–165 (2018)
18. R. Barzilay, M. Elhadad, Using lexical chains for text summarization. Adv. Autom. Text Summ. 111–121 (1999)
19. A. Ibrahim, T. Elghazaly, Improve the automatic summarization of Arabic text depending on Rhetorical Structure Theory, in *2013 12th Mexican International Conference on Artificial Intelligence* (IEEE, 2013), pp. 223–227
20. F.B. Goularte, S.M. Nassar, R. Fileto, H. Saggion, A text summarization method based on fuzzy rules and applicable to automated assessment. Expert Syst. Appl. **115**, 264–275 (2019)
21. R. Nallapati, F. Zhai, B. Zhou, Summarunner: a recurrent neural network based sequence model for extractive summarization of documents, in *Thirty-First AAAI Conference on Artificial Intelligence* (2017)
22. L.M.Q. Abualigah, E.S. Hanandeh, Applying genetic algorithms to information retrieval using vector space model. Int. J. Comput. Sci., Eng. Appl. **5**(1), 19 (2015)
23. L.M. Abualigah, A.T. Khader, Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. J. Supercomput. **73**(11), 4773–4795 (2017)
24. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A new feature selection method to improve the document clustering using particle swarm optimization algorithm. J. Comput. Sci. (2017)
25. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A combination of objective functions and hybrid krill herd algorithm for text document clustering analysis. Eng. Appl. Artif. Intell. (2018)
26. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A novel weighting scheme applied to improve the text document clustering techniques, *Innov. Comput., Optim. Its Appl.* (Springer, Cham, 2018), pp. 305–320
27. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A.H. Gandomi, A novel hybridization strategy for krill herd algorithm applied to clustering techniques. Appl. Soft Comput. **60**, 423–435 (2017)

28. B. Ionescu, H. Müller, M. Villegas, A.G.S. de Herrera, C. Eickhoff, V. Andrearczyk, … Y. Ling, Overview of Image CLEF 2018: challenges, datasets and evaluation, in *International Conference of the Cross-Language Evaluation Forum for European Languages* (Springer, Cham, 2018), pp. 309–334
29. Z.A. Al-Sai, L.M. Abualigah, Big data and E-government: a review, in *2017 8th International Conference on Information Technology (ICIT)* (IEEE, 2017), pp. 580–587
30. H. Zhang, Y. Gong, Y. Yan, N. Duan, J. Xu, J. Wang, … M. Zhou (2019) Pretraining-based natural language generation for text summarization. arXiv preprint arXiv:1902.09243
31. K. Karpagam, A. Saradha, A framework for intelligent question answering system using semantic context-specific document clustering and Wordnet. Sādhanā **44**(3), 62 (2019)
32. Y. Wu, Pattern-enhanced topic models and relevance models for multi-document summarisation (Doctoral dissertation, Queensland University of Technology, 2018)
33. H. Froud, A. Lachkar, S.A. Ouatik, Arabic text summarization based on latent semantic analysis to enhance arabic documents clustering. arXiv preprint arXiv:1302.1612 (2013)
34. A. Kumar, A. Sharma, Systematic literature review of fuzzy logic based text summarization. Iran. J. Fuzzy Syst. 1706–3727 (2019)
35. A. Garg, J.R.A. Saini, Systematic and exhaustive review of automatic abstrac-tive text summarization for hindi language (2012)
36. A.K. Sangaiah, A.E. Fakhry, M. Abdel-Basset, I. El-henawy, Arabic text clustering using improved clustering algorithms with dimensionality reduction. Clust. Comput. 1–15 (2018)
37. S. Shashikanth, S. Sanghavi, Text summarization techniques survey on telugu and foreign languages (2019)
38. A. Wilbik, J.M. Keller, A fuzzy measure similarity between sets of linguistic summaries. IEEE Trans. Fuzzy Syst. **21**(1), 183–189 (2012)
39. Q.A. Al-Radaideh, D.Q. Bataineh, A hybrid approach for Arabic text summarization using domain knowledge and genetic algorithms. Cogn. Comput. **10**(4), 651–669 (2018)
40. N. Alami, N. En-nahnahi, S.A. Ouatik, M. Meknassi, Using unsupervised deep learning for automatic summarization of Arabic documents. Arab. J. Sci. Eng. **43**(12), 7803–7815 (2018)
41. A.M. El-Halees, D. Salah, Feature-based opinion summarization for arabic reviews, in *2018 International Arab Conference on Information Technology (ACIT)*. (IEEE, 2018), pp. 1–5
42. D.I. Saleh, feature-based opinion summarization for Arabic reviews. Featur.-Based Opin. Summ. Arab. Rev. (2017)