International Journal of Information Technology and Language Studies (IJITLS)

Arabic text summarization approaches: A comparison study

Hani S. AlGhanem and **Rashan H. Ajamiah**
Hanismg@hotmail.com; roshan_4@hotmail.com

Faculty of Engineering & IT, The British University in Dubai, Dubai, UAE

Abstract. Text summarization is considered one of the essential parts of the NLP area, as it gets attention since the '50s of the last century. Although it has evolved rapidly in the last decades for Latin languages, Arabic text summarization is still a green area for researchers. Many algorithms can be used to generate Arabic text summarization. The analysis shows that the best algorithm for single document summarization in the Arabic language is Arabic summarization using the clustering technique with word rooting capability. The unique algorithm for multi-document summarization is Text Summarization using the Centrality Concept. A detailed literature review covers Text summarization in general and Arabic text summarization in specific and its challenges.

Keywords: *NLP, ANLP, Arabic Text Summarization, Automatic Text Summarization, ATS.*

1. Introduction

Recently Automatic Text summarization has excellent attention. Because of massive text generated daily in different formats on the internet (Chia-Hui Chang et al. 2006), the existing text data is available in electronic format either over the internet or on organizational or personal computers.

As people need more time to read the entire text of many documents with the same subject to get the main idea, text summarization came to overcome this issue, as it saves time by generating a shorter text edition of the text containing the same ideas. Also, it saves the cost comparing with human expert summarization.

Information Retrievals (IR), Machine Translation (MT), Questions and answers (QA), and Text Summarization, considered an application of natural language processing (Bataineh and Bataineh, 2009). There are many pieces of Research in NLP for Latin languages comparing to Arabic langue, especially in the summarization area. Imam et al. (2013) addressed a shortage in studying the Arabic language in NLP, especially in the summarization area. Using automatic summarization reduces the cost compared with human experts to do the summarization. Some efforts have been made from research in Arabic Text Summarization based on various algorithms and approaches; this paper will briefly summarize text summarization and evolution of text summarization in general and focus on ATS approaches and provide a comparative study. Mhamdi, Al-Emran, Salloum (2018) assert the use of Arabic text summarization as one application of using text mining, which increases the provided text's usability.

2. Literature Review

This section will cover the main topics of text summarization, including definition, details about techniques and approaches, the evolution of Automatic Text Summarization, and focusing on Arabic text summarization approaches and comparison with their methods and output results.

2.1. Text summarization

Text summarization is defined as a process to generate a short version of the original text that contains the same idea as the original Text (ALGULIEV & ALIGULIYEV 2009). while Jusoh (2018) describe text summarization as a process to understand the main idea of a source text and generate a shorter version of the text, usually it has done by extracting sentences from the original text based on linguistics and statistical methods, one of the most complex problems that face the summarization and NLP, in general, is the ambiguity.

Radev, Hovy, and McKeown (2002) "text that is produced from one or more texts that convey important information in the original text(s) and that is no longer than half of the original text(s) and usually significantly less than that.". It can also be defined as "To take an original article, understand it and pack it neatly into a nutshell without loss of substance or clarity, presents a challenge which many have felt worth taking up for the joys of achievement alone. These are the characteristics of an art form" (Ashworth 1973). Al-Taani and Al-Omour (2014) specify summarization categorized into multi-document and single-document summarization. Using the combination of statistics and machine learning will improve the summarization process. (PadmaPriya & Duraiswamy, 2012).

2.2. Types and approaches to Text Summarization

Nenkova (2005) expresses that there are two types of summarization, which are a single document and multi-document summarization. In contrast, ALGULIEV and ALIGULIYEV (2009) considered the summarization type as:

1) Extractive that selects some of the original text sentences to present the same idea of the Text-based on some centrality, Similarity, or clustering. Clustering is a process to find the distribution, grouping, and patterns. Clustering can be done by making a presentation for all elements and then calculating the distance between the elements. After that, determine the criteria for optimization, finally do the optimization.

2) Abstractive can generate new sentences that are not part of the original text using deep NLP processing, and this also supported by Mohamed and Hariharan (2018) as they addressed that summarizations are classified into two categories; the first category is based on selecting some sentences from the original text that has the same idea, while the second category is writing a short version of the text using new sentences which do not appear in the original Text; So summarization can be categorized either by the number of document it summarizes or the approach used to generate the script. by studying 42 systems from 1958 to 2007, the summary the output of summarizations system is categorized as extracts, indicative abstracts, capsule overview, abstracts, personalized extracts, personalized summaries, headlines, and abridged extracts (Lloret, 2008).

Lloret (2008) explains the stages of text summarization as 1) understanding the original text, 2)transfer it into a summary, 3) generate summary text from the second stage. Also, he classifies the approaches of text summarization depending on how deep the summarization system dealing with the text, and this came as

1) Shallow level or surface-level depends on a shallow characteristic such as Thematic based on the number of reparation words in the text after eliminating the stop words. Location choosing the sentence to be part of summarization based on its location in the paragraph, as the opening statement, Background focus on the statement that has more words related to the title, or Cue words which include a list of words that express that the sentence is essential.

2) Entity-level includes the relations between entities by checking the Similarity between words stems, the distance between words, Re-occurrence with the same meaning, Relationship between words, parsed tree, or meaning/ logical relations.

3) Speech level that focuses on the entire arrangement and main idea of the text, which depend on text format subjects that text covers, text coherence, and structure

ALGULIEV and ALIGULIYEV (2009) describe three phases for text summarization “analysis, transformation, and synthesis”.

In the early stages of text summarization Baxendale (1958) introduce the propositional method, which is suitable for scientific papers as he analysis around two hundred paragraphs and he find out that the basic sentences are the first and the last sentences as the essential info are located in it, this is a reasonable but straightforward approach that time. In the same period, Luhn (1958) summarizes technical text by using stemming, removing the stop words. Frequency of content terms, which is the sentences which include words for more frequent in a document, and chose sentence with a high concentration of most noticeable words, each sentence take a score based on the square of the number of the critical word in the sentence divided by the number of the words of the sentence. The high ranking sentences will be considered in the summary paragraph. Edmundson (1969) focus on technical document and uses a combination of features of 1)position and frequency. 2) Cue words like (“conclusion”, ”last of all”, ”significant”) that take extra points to consider the sentence in summary .3) and document structure are considered like titles or headings, or any sentence came under it.

A competitive approach was introduced by DeJong (1979, 1982) and named Frump knowledge-based summarization. It automatically processes a group of news articles and find the type of scenarios it provides and check with the slots that need to be filled, then use a slot filler to generate the sentences. There were around 50 sketchy scripts based on manually selected keywords, so it is challenging to apply them to other domains.

Brandow, Mitze, and Rau (1995) introduce another new summarization system called Automatic News Extraction System (ANES), which scans more than 20 thousand documents and words selected based on term frequency multiply inverse of document frequency. Also, he uses sentence-based features; his evaluation for his output has been done by asking people to rank the summaries, whether it is acceptable or not. In the same period, Kupiec, Pedersen, and Chen (1995) bring in the first trainable machine learning technique for sentence extraction. This extracts 20% of the input, and their collection is 188 documents from scientific journals using Naive Bayesian classifier and some new features sentence length, presence of uppercase words, thematic words, and 26 manually fixed phrases, and sentence position in paragraph. Result shows for 25% of summaries give 84% precision and 74% improvement over lead-based for smaller summaries. McKeown and Radev (1995) start working on summarization for multi-document, they use the approach use knowledge-base and information extraction (Message Understanding Conference -MUC- template), and text generation to generate a sentence that contains the slots fillers for MUC Template.

Salton et al. (1997) worked graph-based summarization techniques and using a corpus of the encyclopedia as they present all Phrases as nodes in the graph and connect nodes using links if the sentence’s contents overlap, which means having a lexical similarity above a threshold. The more links for a node, the more important is the phrase and better chances to be part of the summary. Jing et al. (1998) use the lexical chain, which is a sequence of semantically related words; the Relationship between words is categorized into three levels, the summary sentences to be chosen based on high scoring chain depending on the length and homogeneity index.

Ultra summarization or headline summarization, using the OCELOT system to do Gisting summarization, and the idea to find the gist that maximizes the probability the gist given the document (Berger & Mittal, 2000). Radev, Jingand Budzikowska (2000) provided a general-purpose framework for extractive summarization based on salience and named Mead, which can handle six different languages, using the centroid-based method, which works for both single and multi-document summarization. Also, it uses other features in addition to Centroids, like position and length.

Random walks method in single and multiple document summarization using Lexical centrality. If a sentence is likely to be visited during a random walk process from a similarity graph corresponding to all the sentences in the document, then that sentence is worth included in the summary. There are some steps required to achieve this, first is to present the text as a graph, then to use graph centrality metric to determine the top sentences, then to use graph clustering to segment the sentence to cover all the themes, finally do the random walk (Erkan & Radev, 2004).

2.3. Text Summarization evaluation

Extrinsic and Intrinsic are considered a type of summary evaluation. Extrinsic techniques or task-based is enabling to make the same decision using the summary as with the full text, but in less time. on the other hand, Intrinsic techniques compare the output summary with a gold-standard summary, and usually by calculating the precision, recall, and the harmonic between precision and recall.

Gholamrezazadeh, Salehi, and Gholamzadeh (2009) assert that recall and precision are used to compare the system's generated summary with human experts' gold standard. They show that recall is equal to correct divided by the correct and missed sentences, while precision is equal to correct divided by the correct and wrong sentences.

As per Al-Khawaldeh and Samawi (2015), Recall, Precision, and the harmonic between precision and recall F-Measure calculated using the following formulas:

$$\text{Recall} = \frac{\text{system - human choice overlap}}{\text{sentences chosen by human}}$$

$$\text{Precision} = \frac{\text{system - human choice overlap}}{\text{sentences chosen by system}}$$

$$F - measure = \frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})}$$

Haboush et al. (2012). The evaluation for text summarization can be measured using precision and recall. Precision is the percentage of the number of relevant sentences retrieved by the system divided by all the system's sentences. In comparison, recall is the percentage of the number of relevant sentences retrieved by the system divided by all the sentences generated by experts manually.

There are many different criteria to evaluate the summarization. Some are the right length, faithful to the original document, providing salience information, written incorrect grammar, not redundant, well-formed, structured, and coherent. Some of those are not specific for summarization and are related to text generation, machine translation, and questions and answers. The ideal evaluation of summarization focuses on the compression ratio, what percentage of the original document is presented in summary, and retention ratio, what is the percentage of information of the original document presented in summary.

Nenkova (2005) says that The Document Understanding Conference (DUC) used the coverage criteria to test the result by checking what has covered statements from system summarization in human summarization. ALGULIEV and ALIGULIYEV (2009) explain that Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is considered one of the most accurate tools to measure the summary quality. Evaluation of text summarization can be done either by human checking or by one of these systems BE, SEE, or ROUGE (Lloret, 2008)

ROUGE technique, which Lin (2004) introduced, is an automatic matrix easy to prototype; Rouge is based on BLEU, introduced by Papineni et al. (2002) matrix used for machine translation. In BLEU, you get a higher score if the summary output has high precision, while in Rouge, you get a higher score if the output has a high recall, so ROUGE-n is a measure of n-gram overlap between the generated summary and a set of references summaries, while ROUGE-L uses longest common subsequence instead of n-gram overlap. Radev, Jing, and Budzikowska (2000) address another matrix related to Rouge, precision, and recall called Relative utility, which can tell there are multiple correct summaries .and calculated as the percentage of ideal utility covered by system summary and can compare the result with more than one judge. Nenkova and Passonneau (2004) establish a pyramid method used for multi-document summarization based on Semantic content units (SCU) that deal with the different facts that are realized using a different formulation.

As per Lin (2004), ROUGE-N is computed as follows:

$$= \frac{\sum_{S \in \{ReferemceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)}$$

2.4. Arabic text summarization and challenges

Haboush et al. (2012) assert that Arabic summarization can be done in the same way as Latin language summarization and using the same algorithms.

Mohamed and Hariharan (2018), in their experiment of different types of inputs, including text in different languages, tables, equations, graphs, and images, conclude that increasing the compression ratio will increase the generated summary accuracy. Lokhande et al. (2015(use NLP functions, and both Expectation-Maximization Clustering, Hierarchical clustering algorithms) generate the summary. Comparing the difference of performance between two clustering methods, they run the two tests using the same text, and results show that Hierarchical clustering is needed less time and less space than Expectation-Maximization Clustering. However, Expectation-Maximization Clustering shows a better result in summarization accuracy. AL-Khawaldeh (2015) focuses on how the negation words affecting the summarization in the Arabic language, and elimination of negation word as considering it stop words will give the opposite meaning of the sentence; so determining the polarity of the verbs and understanding the negation words gives better results comparing by neglecting the negation words.

It is considered challenging to do Arabic text summarization because there are many challenges such as missing of the diacritic for most written Arabic articles, which required a sophisticated analysis to know the right diacritic, which help to get the right meaning of the word and sentence (Azmi, and AlShenaifi, 2014) and (Meselhi et al.,2014). AL-Khawaldeh (2019) explains why's make the Arabic language harder than other Latin languages due to complicated morphology, few NLP tools compared with other languages, and not having the first capital letter indicates for Names. Also, the complication of Arabic language analysis is that it contains many forms of adding letters that can be reached to four letters and sometimes more. Additionally, the writing of letters is changed depending on its position of the word (Kanaan, Al-shalabi, and Sawalha, 2003). Regarding the stop words and stemming functionality, there is a complexity which required extra effort to make a high-quality root stemming process (Al-Zahrani, Mathkour, and Abdalla, 2015) and (Awajan, 2007). Furthermore, many writers use their local speaking and wording, which is different from one region to another, as one word might have a different meaning in different regions (Saleh, 2017).

3. Methodology

AlGhanem et al. (2020) describe the systematic literature review as an excellent way to explorer all literature related to a specific topic without biases, as it provides an in-depth knowledge of what is published. Also, the search for published papers comes in an organized way, searching for a specific keyword. Additionally, it determines what condition should include the research paper and what condition to exclude the research paper.

3.1. Inclusion/exclusion criteria

In this research, we set a list of conditions that considered inclusion and exclusion criteria as shown in table 1 below

Inclusion	Exclusion
It should be related to Arabic text summarization.	Articles published before 2010.
Should have evaluation using recall, precision, and F measure.	Articles the use of another corpus
Using single or multi-document summarization	Articles that do not mention the type of summarization (single or multiple document summarization)

Table 1. Inclusion and exclusion criteria

3.2. Data sources and search criteria

For data collection, a search engine of an academic published paper provided by google was used at https://scholar.google.com/. The used key wards were “Arabic”+ “text” + “summarization”. and the year of publication was also determined not older than 2010.

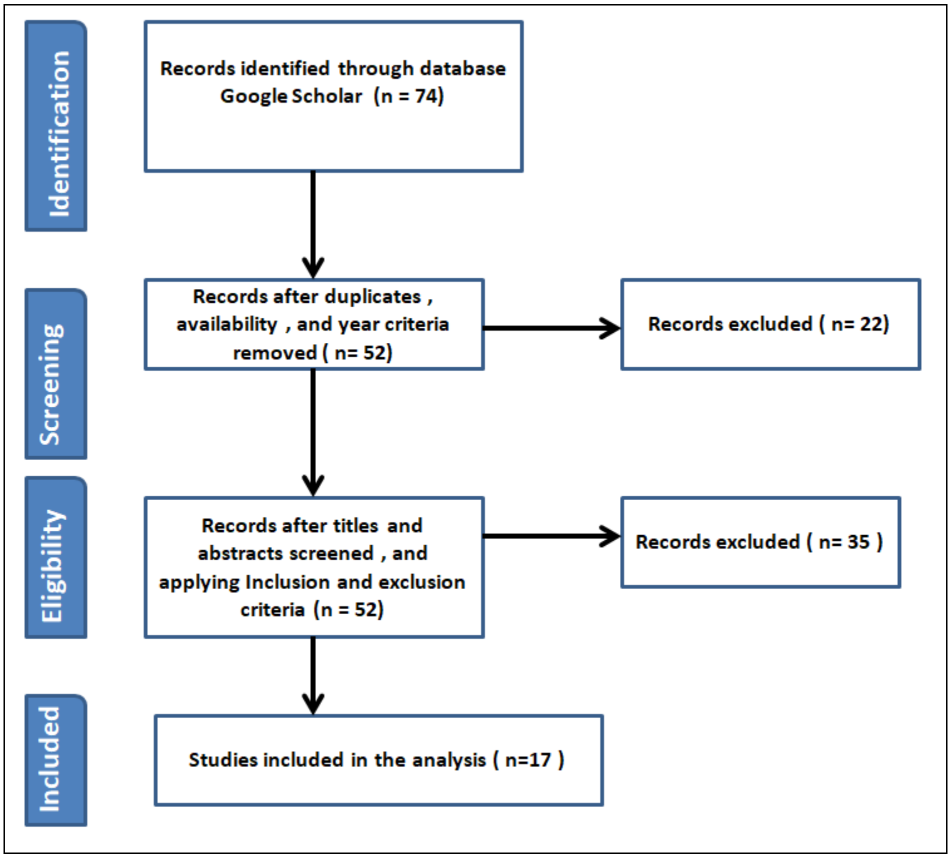

Figure 1. Process of selection papers

Figure 1 above shows the process of paper selection, which start by searching papers using the provided keywords; the initial results show that there are 74 papers related to Arabic text summarization; after removing the duplicated and non-availability of the full text of the papers, the remaining of 52 papers were considered. All of them were eligible to be introduced to the inclusion and exclusion criteria. After applying inclusion and exclusion criteria, only 19 papers remain. These papers are summarized in Table 3 below.

3.3. Quality assessment

There are different quality of the selected papers and quality matrix required to ensure strong evidence of the selected papers, the following list of question should be answered and scored for each paper:

1) Is the paper related to Arabic text summarization

2) Is the paper providing a clear idea and a specific scope?

3) Is there a clear description of the used algorithm?

4) Is the corpus mentioned clearly

5) Is the evaluation based on recall, precession, and F measure

All the above questions were applied to the selected paper to measure and ensure the excellent quality of selected papers, and the results are shown in Table 2.

Source	Q1	Q2	Q3	Q4	Q5	Total	%
P1	1	1	1	0	1	4	80%
P2	1	1	1	1	1	5	100%
P3	1	1	1	1	1	5	100%
P4	1	1	1	1	1	5	100%
P5	1	1	1	1	1	5	100%
P6	1	1	1	1	1	5	100%
P7	1	1	1	0	1	4	80%
P8	1	1	1	1	1	5	100%
P9	1	1	1	0	1	4	80%
P10	1	1	1	1	1	5	100%
P11	1	1	1	1	1	5	100%
P12	1	1	1	1	1	5	100%
P13	1	1	1	1	1	5	100%
P14	1	1	1	1	1	5	100%
P15	1	1	1	1	1	5	100%
P16	1	1	1	0	1	4	80%
P17	1	1	1	1	1	5	80%

Table 2. Quality assessment result

Results of quality assessment results show that the selected papers come with the range of 80% to 100%, which considered as a good quality to consider the paper in the analysis phase.

4. Discussion

The leading Arabic summarization techniques and approaches have been studied, and results checked to build the comparison between them based on the output result; Recall and Precision are gathered from published papers, and the harmonic between them are calculated, and they are listed as:

P1) Haboush et al. (2012), in studying Arabic summarization using the clustering technique, use word root instead of the word itself, as they use the same methodology of summarization stages except for the ranking phase is based on word root. The result shows an improvement of 14% in precision and 9% in a recall, with 75.5% and 78.7% values, respectively.

P2) Waheeb and Husni (2014(they addressed the redundant problem, like sentences with the same meaning appear in summary, redundant removal can be done by checking the Similarity and clustering of the summarized sentences .this experiment was tested using Essex Arabic Summaries Corpus (EASC) which contain 157 articles in Arabic language and 765 summaries created by experts. After evaluating the final result, it shows the average precision and recalls a value of 60%.

P3) Imam et al. (2013) noticed that Machine Learning (ML) is not used in most summarization systems. So they introduce An Ontology-based Summarization System for Arabic Documents (OSSAD), which is built based on ML and decision tree. OSSAD shows higher values for precision 53% and Recall 47% comparing with other systems like Gen-Summ, and LSA-Summ.

P4) Al-Taani and Al-Omour (2014) make an experiment to summarize Arabic text depending on graph extractive. Also, they use NLP functions like n-gram and stemming; the summary generation is built based on short bath rout using sentence sequence, then to determine the Similarity between the extracted, this approach shows output values with Precision 50.2% and recalls around 47.2%. In conclusion, it shows that using the n-gram method is far from using the word or its stem.

P5) Algaphari, Ba-Alwi, and Moharram (2013) use the Centroid-Based Algorithm (CBA) and sentence centrality to create Arabic Text Summarization. They use four main sentence centrality based on graph representation, summarization based on Centroid, sentence salience concept, and Degree centrality. results show good performance with a precision of 64.6% and recall of 71.9%

p6) Fejer and Omar (2015) suggest an approach to do Arabic multi-document text summarization by combining both keyphrase extraction and clustering; by loading multi-document into the system, a text pre-processing is used to do the remove stop word, tokenization, and stemming. Then document clustering is started by doing hierarchal methods complete-linkage and single-linkage; as the last step in the clustering phase, a K-means method is used. After that, keyphrase extracting started and used noun-phrase extraction and ranking; the last phase is to filter and extract the essential sentences by doing sentence splitting, essential sentence extraction, and similar sentence filtering. Results show recall of 45.2%, precision 41.8%.

p7) Al-Khawaldeh and Samawi (2015) introduce Lexical Cohesion and Entailment based Segmentation for Arabic Text Summarization (LCEAS), which uses the following steps to generate Arabic document summary, starting by having the plain text converted into sentences. In the processing stage, removing stop words and doing the stemming, the next stage is to do word sense dis-ambiguity for the relevant words the system applies. The un-ambiguity words moved to lexical cohesion-based summarization; the meaningful sentences moved to the Entailment based text summarization stage; the non-entitled sentences are considered a final summary. The result shows a recall of 73.6% and a precision of 66.7%.

P8, P9, P10) Alami, Meknassi, and Rais (2015) studying the current state of art summarization systems and they mentioned three Arabic summarization system as 1) Word frequency; RST which mentioned in Ikhtasir — A user selected compression ratio Arabic text summarization system with a recall of 57%, and precision 72%. 2) Statistical features extraction; Bayesian classifier; Genetic Programming classifier, which is mentioned in An Optimized Dual Classification System for Arabic Extractive Generic Text Summarization with recall 72.5% and precision 49%. 3) RST, which mentioned in Arabic text summarization using Rhetorical Structure Theory with recall 26% and precision 34%.

P11) Al-Abdallah, and Al-Taani(2017) use in their research the particle swarm optimization algorithm to generate Arabic text summarization depending on a single document approach. His study shows the recall of 54.44%, precision of 58,82%, and F measure of 55.32%.

P12) Waheeb et al. (2020) use Clustering and Word2Vec to Reduce Redundancy algorithm in their study to summarize Arabic text from EASC corpus and results show 69.5%, 60%,64.4% for recall, precision, and f Measure prospectively.

P13) Al-Abdallah, and Al-Taani, (2019) apply Firefly Algorithm to extract Arabic text summarization from the EASC corpus, and results show 60.14%, 57.32%,57.52% for recall, precision, and f Measure prospectively.

P14) Graph-Based Extractive algorithm were used in Elbarougy, Behery, and KHATIB,(2020) study, as they apply it into single document summarization on ESAC. Results show the recall of 73.8%, precision of 82.67%, and F measure of 76.37%.

P15) in Elbarougy, Behery, and El Khatib,(2020) the use Extractive Arabic Text Summarization Using Modified PageRank Algorithm to generate a short version of provided single document from EASC corpus, results shows values of 72.94% 68.75% 67.99% for recall, precision, and f Measure prospectively.

P16) Abu Nada et al.(2020) describe the use of AraBERT Model Using Extractive Text Summarization Approach to summarize a single document, and results show 39% 90% 54% for recall, precision, and f Measure prospectively.

P17) Jaradat, and Al-Taani, (2016) use a Hybrid-based algorithm on a single document of EASC corpus, and results show 57.13%,56.58%,54.76% for recall, precision, and f Measure prospectively.

ID	Reference	Technique/description	Recall	Precision	F-Measure	Corpus	Multi/Single Documents
P1	Haboush et al. (2012)	Arabic summarization using clustering technique with word rooting capability.	78.7%	75.5%	77%	Not mentioned	Single Document
P2	Waheeb and Husni (2014)	Arabic Summarization Using Text Clustering to Reduce Redundancy	60%	60%	60%	Essex Arabic Summaries Corpus (EASC)	Multi-document
P3	Imam et al. (2013)	An Ontology-based Summarization System for Arabic Documents (OSSAD)	47%	53%	49.8%	Essex Arabic Summaries Corpus (EASC)	Single Document
P4	Al-Taani and Al-Omour (2014)	An Extractive graph-based arabic text summarization Approach	47.2%	50.2%	48.6%	Essex Arabic Summaries Corpus (EASC)	Single Documents
P5	Algaphari, Ba-Alwi, and Moharram (2013)	Text Summarization using Centrality Concept	71.9%	64.6%	68%	Arabic NEWSWIRE-a corpus	Multi-document
P6	Fejer and Omar (2015)	Automatic Multi-Document Arabic Text Summarization Using	45.2%	41.8%	43.4%	corpus DUC2002	Multi-document
P7	Al-Khawaldeh and Samawi (2015)	Lexical Cohesion and Entailment based Segmentation for Arabic Text Summarization (LCEAS)	73.6%	66.7%	69.9%	Arabic textual entailment (ArbTEDS)	Single Document
P8	Alami, Meknassi, and Rais (2015)	Word frequency; RST	57%	72%	66.2%	Not Mentioned	Single Document
P9	Alami, Meknassi, and Rais (2015)	Statistical features extraction; Bayesian classifier; Genetic Programming classifier	72.5%	49%	58.4%	manually labeled corpus	Single Document
p10	Alami, Meknassi, and Rais (2015)	RST	26%	24%	24.9%	Not Mentioned	Single Document a
P11	Al-Abdallah, and Al-Taani(2017)	particle swarm optimization algorithm	54.44%	58,82%	55.32%	EASC	Single Document
P12	Waheeb et al. (2020)	Clustering and Word2Vec to Reduce Redundancy	69.5%	60%	64.4%	EASC	Single Document
P13	Al-Abdallah, and Al-Taani, (2019)	Firefly Algorithm	60.14%	57.32%	57.52%	EASC	Single Document
P14	Elbarougy, Behery, and KHATIB, (2020)	Graph-Based Extractive	73.8%	82.67%	76.37%	EASC	Single Document
P15	Elbarougy, Behery, and El Khatib,(2020)	Extractive Arabic Text Summarization Using Modified PageRank Algorithm	72.94%	68.75%	67.99%	EASC	Single Document
P16	Abu Nada et al. (2020)	AraBERT Model Using Extractive Text Summarization Approach	39%	90%	54%	Not Mentioned	Single Document
P17	Jaradat, and Al-Taani, (2016)	Hybrid-based	57.13%	56.58%	54.76%	EASC	Single Document

Table 3. List of Arabic text summarization methods and approaches.

Table 3 shows the list of Arabic text summarization methods and approaches, and its evaluation value, including the reference and the published paper, brief about technique and approach, recall, precision, and F-Measure calculated for the give recall and precision, and the corpus that used to validate the test result. From the above table, only three methods are used for multi-document summarization, and they are presented in below Figure 2.

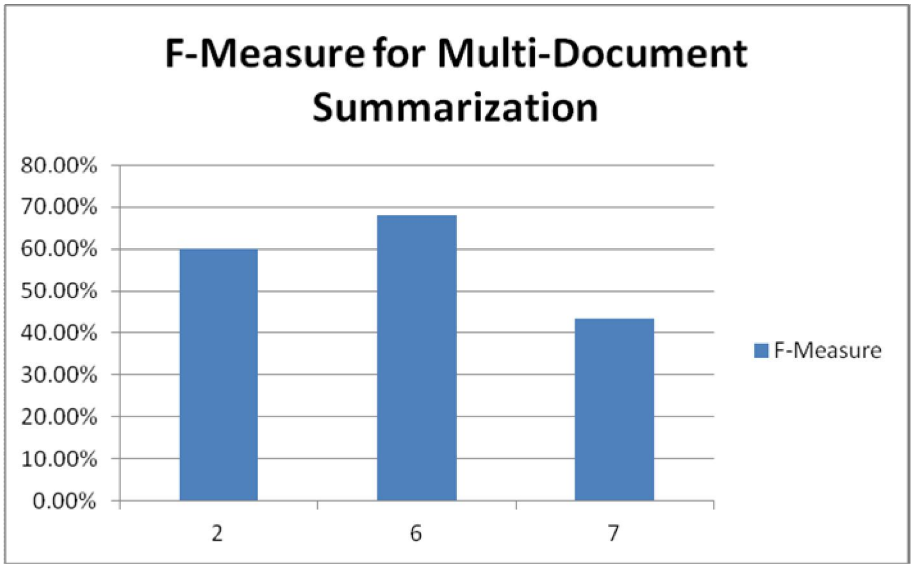

Figure 2. F-Measure for multi-document summarization

Figure 2 shows X-axis as a reference to the approach in Table 3 and the y-axis as the percentage of F-Measure, which means that the text summarization using the centrality concept provides the best result of Arabic text summarization for multi-document. However, all of the approaches listed depend on clustering; the centrality concept based on graph representation shows a higher F-Measure with a value of 68%, while other methods did not exceed 60%.For single Arabic document summarization below Figure 3

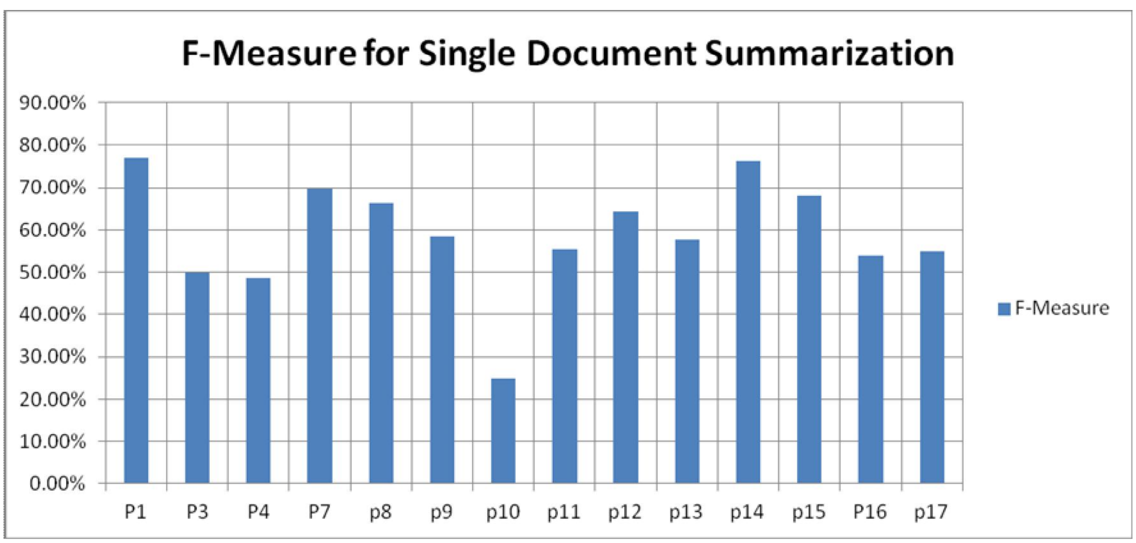

Figure 3. F-Measure for singe-document summarization

Figure 3 shows X-axis as a reference to the approach in Table 3. The y-axis the percentage of F-Measure, the best F-Measure value of 77% for Arabic summarization using the clustering technique with word rooting capability, the critical feature of reaching this heist value comparing by other methods is the usage of the words rooting before building the clusters. However, other approaches use the roots of the words as (LCEAS). It focuses on polarity and negation words and verbs, but the result of F-Measure did not exceed 70%.

5. Future Prospects

(LCEAS) has an excellent point regarding Arabic summarization as the negation, which might be considered as stop words, can give a different and opposite meaning of the sentence. While the root of words will remain the same, it is suggested for future research related to Arabic text summarization to combine the verbs' polarity and recognize the negation words in a stage before getting word roots, or words similarities give better results and more accuracy. Also, it suggested to restudy the approaches in table 3 and to check the performance of summarization speed and memory usage. Furthermore, to revalidate all the mentioned algorithms using the same corpus.

6. Conclusions

Text summarization is considered very useful for readers to understand the main idea of provided long text and saves time and effort. Although there are many studies related to text summarization of different languages, Arabic text summarization is considered a greenfield for exploring and doing new research. Many techniques and algorithms could be employed and applied to the Arabic text to generate a shorter version. Studies related to Arabic text summarization from the past ten years were collected and analyzed using a systematic literature review; various algorithms were used with different quality of the provided answer. This study shows the best algorithm used to provide higher accuracy and quality for Arabic text summarization for single and multi-values summarization.

By comparing the Arabic summarization methods, it found that there are several kinds of summarizations. One of them depends on the input document's number; as shown in the discussion, the best approach for single document summarization is Arabic summarization using the clustering technique with word rooting capability. Furthermore, for multi-document summarization, the best approach is Text Summarization using the Centrality Concept.

References

Abu Nada, A. M., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach.

Al-Abdallah, R. Z., & Al-Taani, A. T. (2017). Arabic single-document text summarization using particle swarm optimization algorithm. Procedia Computer Science, 117, 30-37.

Al-Abdallah, R. Z., & Al-Taani, A. T. (2019, February). Arabic text summarization using firefly algorithm. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 61-65). IEEE.

Alami, N., Meknassi, M., & Rais, N. (2015). Automatic texts summarization: Current state of the art. Journal of Asian Scientific Research, 5(1), 1-15.

Algaphari, G., Ba-Alwi, F. M., & Moharram, A. (2013). Text summarization using centrality concept. International Journal of Computer Applications, 79(1).

AlGhanem, H., Shanaa, M., Salloum, S., & Shaalan, K. (2020). The Role of KM in Enhancing AI Algorithms and Systems. Advances in Science, Technology and Engineering Systems Journal, 5(4), 388-396.

Alguliev, R., & Aliguliyev, R. (2009). Evolutionary algorithm for extractive text summarization. Intelligent Information Management, 1(02), 128.

AL-Khawaldeh, F. T. (2019). A study of the effect of resolving negation and sentiment analysis in recognizing text entailment for Arabic. arXiv preprint arXiv:1907.03871.

AL-Khawaldeh, F. T. (2019). Answer extraction for why Arabic questions answering systems: EWAQ. arXiv preprint arXiv:1907.04149.

AL-Khawaldeh, F. T., & Samawi, V. W. (2015). Lexical cohesion and entailment based segmentation for arabic text summarization (lceas). World of Computer Science & Information Technology Journal, 5(3).

Al-Radaideh, Q., & Afif, M. (2009). Arabic text summarization using aggregate similarity. In International Arab conference on information technology (ACIT2009), Yemen.

Al-Taani, A. T., & Al-Omour, M. M. (2014). An extractive graph-based Arabic text summarization approach. In The International Arab Conference on Information Technology.

Al-Zahrani, A. M., Mathkour, H., & Abdalla, H. I. (2015). PSO-Based Feature Selection for Arabic Text Summarization. J. UCS, 21(11), 1454-1469.

Ashworth, W. (1973). Abstracting as a fine art.

Awajan, A. (2007). Arabic text preprocessing for the natural language processing applications. Arab Gulf Journal of Scientific Research, 25(4), 179-189.

Azmi, A., & AlShenaifi, N. (2014). Handling "why" questions in Arabic. In The 5th International Conference on Arabic Language Processing (CITALA'14).

Bataineh, B. M., & Bataineh, E. A. (2009, July). An efficient recursive transition network parser for Arabic language. In Proceedings of the World Congress on Engineering (Vol. 2, pp. 1-3).

Baxendale, P. B. (1958). Machine-made index for technical literature—an experiment. IBM Journal of research and development, 2(4), 354-361.

Berger, A., & Mittal, V. O. (2000, October). Query-relevant summarization using FAQs. In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (pp. 294-301).

Brandow, R., Mitze, K., & Rau, L. F. (1995). Automatic condensation of electronic publications by sentence selection. Information Processing & Management, 31(5), 675-685.

Chang, C. H., Kayed, M., Girgis, M. R., & Shaalan, K. F. (2006). A survey of web information extraction systems. IEEE transactions on knowledge and data engineering, 18(10), 1411-1428.

DeJong, G. (1979). Prediction and substantiation: A new approach to natural language processing. Cognitive Science, 3(3), 251-273.

DeJong, G. (1982, August). Automatic Schema Acquisition in a Natural Language Environment. In AAAI (pp. 410-413).

Edmundson, H. P. (1969). New methods in automatic extracting. Journal of the ACM (JACM), 16(2), 264-285.

Elbarougy, R., Behery, G., & El Khatib, A. (2020). Extractive Arabic Text Summarization Using Modified PageRank Algorithm. Egyptian Informatics Journal, 21(2), 73-81.

Elbarougy, R., Behery, G., & KHATIB, A. E. (2020). Graph-Based Extractive Arabic Text Summarization Using Multiple Morphological Analyzers. Journal of Information Science & Engineering, 36(2).

Erkan, G., & Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research, 22, 457-479.

Fejer, H. N., & Omar, N. (2015). Automatic multi-document Arabic text summarization using clustering and keyphrase extraction. Journal of Artificial Intelligence, 8(1), 1.

Gholamrezazadeh, S., Salehi, M. A., & Gholamzadeh, B. (2009, December). A comprehensive survey on text summarization systems. In 2009 2nd International Conference on Computer Science and its Applications (pp. 1-6). IEEE.

Haboush, A., Al-Zoubi, M., Momani, A., & Tarazi, M. (2012). Arabic text summarization model using clustering techniques. World of Computer Science and Information Technology Journal (WCSIT) ISSN, 2221-0741.

Imam, I., Nounou, N., Hamouda, A., & Khalek, H. A. A. (2013). An ontology-based summarization system for arabic documents (ossad). International Journal of Computer Applications, 74(17), 38-43.

Jaradat, Y. A., & Al-Taani, A. T. (2016, April). Hybrid-based Arabic single-document text summarization approach using genatic algorithm. In 2016 7th International Conference on Information and Communication Systems (ICICS) (pp. 85-91). IEEE.

Jing, H., Barzilay, R., McKeown, K., & Elhadad, M. (1998, March). Summarization evaluation methods: Experiments and analysis. In AAAI symposium on intelligent summarization (pp. 51-59).

Jusoh, S. (2018). A STUDY ON NLP APPLICATIONS AND AMBIGUITY PROBLEMS. Journal of Theoretical & Applied Information Technology, 96(6).

Kanaan, G., Al-shalabi, R., & Sawalha, M. (2003). Full automatic Arabic text tagging system. In proceedings of the International Conference on Information Technology and Natural Sciences (pp. 258-267).

Kupiec, J., Pedersen, J., & Chen, F. (1995, July). A trainable document summarizer. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 68-73).

Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).

Lloret, E. (2008). Text summarization: an overview. Paper supported by the Spanish Government under the project TEXT-MESS (TIN2006-15265-C06-01).

Lokhande, M. M. P., Gawande, M. N., Koprade, M. S., & Bewoor, M. M. TEXT SUMMARIZATION USING HIERARCHICAL CLUSTERING ALGORITHM AND EXPECTATION MAXIMIZATION CLUSTERING ALGORITHM.

Luhn, H. P. (1958). A business intelligence system. IBM Journal of research and development, 2(4), 314-319.

McKeown, K., & Radev, D. R. (1995, July). Generating summaries of multiple news articles. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 74-82).

Meselhi, M. A., Bakr, H. M. A., Ziedan, I., & Shaalan, K. (2014, December). Hybrid named entity recognition-application to Arabic language. In 2014 9th International Conference on Computer Engineering & Systems (ICCES) (pp. 80-85). IEEE.

Mhamdi, C., Al-Emran, M., & Salloum, S. A. (2018). Text mining and analytics: A case study from news channels posts on Facebook. In Intelligent Natural Language Processing: Trends and Applications (pp. 399-415). Springer, Cham.

Mohamed, S. S., & Hariharan, S. (2018). A performance study of text summarization model using heterogeneous data sources. International Journal of Pure and Applied Mathematics, 119(16), 2001-2007.

Nenkova, A. (2005). Automatic text summarization of newswire: Lessons learned from the document understanding conference.

Nenkova, A., & Passonneau, R. J. (2004). Evaluating content selection in summarization: The pyramid method. In Proceedings of the human language technology conference of the north american chapter of the association for computational linguistics: Hlt-naacl 2004 (pp. 145-152).

PadmaPriya, G., & Duraiswamy, K. (2012). An approach for concept-based automatic multi-document summarization using machine learning. Int. J. Appl. Inf. Syst, 3, 49-55.

Paice, A. D. B., & Moore, J. B. (1990). On the Youla-Kucera parametrization for nonlinear systems. Systems & Control Letters, 14(2), 121-129.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

Radev, D. R., Hovy, E., & McKeown, K. (2002). Introduction to the special issue on summarization. Computational linguistics, 28(4), 399-408.

Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. Information Processing & Management, 40(6), 919-938.

Saleh, D. I. (2017). feature-based opinion summarization for arabic reviews. feature-based opinion summarization for arabic reviews.

Salton, G., Singhal, A., Mitra, M., & Buckley, C. (1997). Automatic text structuring and summarization. Information processing & management, 33(2), 193-207.

Waheeb, S. A., & Husni, H. (2014). Multi-Document Arabic Summarization Using Text Clustering to Reduce Redundancy. International Journal of Advances in Science and Technology (IJAST), 2(1), 194-199.

Waheeb, S. A., Khan, N. A., Chen, B., & Shang, X. (2020). Multidocument Arabic Text Summarization Based on Clustering and Word2Vec to Reduce Redundancy. Information, 11(2), 59.