International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Performance Analysis for ROUGE And F-Measure in Arabic Text Summarization

Hawraa Ali Taher[1], Hasanain Ali Hussein[2], Sabah Khudhair Abbas[3]

Submitted: 11/11/2022 **Accepted:** 13/02/2023

Abstract*:* Summarizing a document has become a necessity as, because so much information is produced every day. Document summary makes it simpler to understand the text document than it would be to read through a collection of documents. A foundation for creating an condensed version of one or more text documents is provided by text summary. It is a crucial method for finding pertinent information on the Internet or in sizable text libraries. Additionally, it is essential to extract data in a way that the user would find the information interesting. Extractive summarization and abstract summarization are the two basic approaches used for text summarizing. In order to create the summary, the extractive summarization method chooses the sentences from a Word document and arranges them according to their weight. Abstractive summarizing is a technique that takes the key ideas from a document's content and expresses them abstractly in plain English. Numerous summary methods have been created on the foundation of these two approaches. There are numerous techniques that are language-specific exclusively. In this paper, we used extractive summarization methods and got good results.

Keywords: *Arabic Language, text summarization, ROUGE, F-measure , preprocessing .*

1. Introduction

Users find it harder and harder to find content that interests them, efficiently search for specific content, or receive an overview of influential, important, and relevant content when there is more and more data available. In today's IT industry, people frequently surf the web for relevant document, but they are seldom able to get all the pertinent information in a single document or web page. As a search result, they can learn how many web pages there are[1].

Different trends in the use of Arabic text summary and form design have emerged as a result of the increasing volume of documents and associated informational text kinds on the internet. The fundamental processing step that defines the summarizing activity is the ranking stage. Arabic can actually be summarized using the same fundamental design concepts as Latin. Due to the nature of the Arabic language and the abundance of functional word derivations, a deeper level of grammatical research is possible. As a result, similar conceptual phrases can be formalized by using words that are either similar or dissimilar [2].

Text document summarizing (TDS) systems frequently incorporate subtasks like text parsing, natural language understanding, reference accuracy, and smoothing accuracy that are taken from the field of natural language processing. The automatic text summarizing technology is at a mature stage and, when combined with conventional information search engines, may provide a solution to the issue of information overload to provide quick access to the most important documents found. This explains the competition to create several algorithmic models and the increasing significance of the field of automatic text summarization[3].

Researchers have created a variety of techniques for summarizing Arabic texts. For instance, Qaroush et al. (2019) [4] suggested an automatic, generic, and abstract summation approach from a single Arabic document with the intention of producing an informative summary. The suggested extraction method analyzes each sentence using a collection of statistical and semantic variables, and a novel formulation is utilized to account for the importance, coverage, and diversity of the sentences.

Additionally, Al Qassem, et al. (2019)[5] provided a novel method of name extraction and fuzzy logic-based summarization of the Arabic text. Using the EASC suite, the suggested summary was assessed and compared to the most recent widely used Arabic text summarization

[1] Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq
ORCID ID : 0000-0003-3375-3677
[2] University of Kufa-Information Technology Research and Development Center, Iraq
Email: hassanean.almayali@uokufa.edu.iq
[3] College of Imam Kadhim for Islamic Sciences University, Iraq

Corresponding Author Email: hawraaa.alshimirty@uokufa.edu.iq

systems. The findings show that our suggested fuzzy-logic method for name extraction is superior to current methods.

In addition, Al-Radaideh, & Bataineh, (2018) [6] developed a mixed approach to The term "single-document text summarization" (ASDKGA) is used. The method uses statistical features, genetic algorithms, and domain knowledge to extract key information from Arab political papers. Two "KALIMAT and Essex Arabic Summaries Corpus (EASC) groups" were used to evaluate the ASDKGA methodology. When summarizing Arab political materials, the (ASDKGA) method produced encouraging results, with an average F of 0.605 and a compression ratio of 40%.

However, Belkebir & Guessoum, (2015)[7] suggested an AdaBoost-based machine learning strategy for summarizing Arabic text. This method is used to determine if a new sentence will probably be included in the abstract or not. I used a collection of Arabic publications to assess the strategy. This method has been compared to other machine learning methods, and the outcomes demonstrate that the method we suggest utilizing AdaBoost is superior than other methods already in use.

2. System Architecture

In the proposed system, we will apply extractive summarization, which selects words and sentences from the original text to be used as a summary of the content. The system includes two stages as following fig(1):

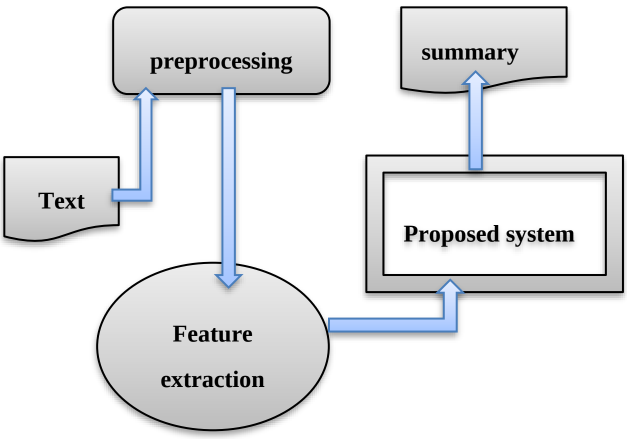

Fig 1. Generalized structure of text summarization

A:preprocessing

a) Tokenization :is a necessary step in natural language processing. In the case of Arabic, where a single word can have up to four independent symbols. Arabic tokenization, is an important step in many researches and has been implemented in many solutions because it is an initial stage that is required for further processing[8].

b) Normalize: is the process of converting a token into its basic form and standardizing the characters. In the process of normalization, the inflectional form is removed to obtain the basic form.

c) stop words removal : It is the process of deleting repeated words such as prepositions, which deleting it does not affect the meaning of the sentence. In our research, the stop words were removed from all sites.

d) Stemming :It is the process of reducing inflectional words and returning them to their origin, base or root form of the written word in general to get rid of the problem of vocabulary mismatch[9].

B: 1-proposed system steps

First, we calculate the number of breaks in the text to determine the number of sentences selected in the summary

Second, we calculate the frequency of each word in the document for all the sentences

Third, we take the highest frequency of a word in the document

Fourth, we divide the frequency of each word by the highest frequency according to the equation:

$$word\_freq[word] = \frac{word\_freq[word]}{\max of\ freq} \quad \dots\dots (1)$$

Fifthly, we divide the text into sentences and we will test each a word in it if it is present in the dictionary of word frequencies and add the sentence that contains the largest number of word frequencies, according to the following law:

$$weight(s) = \frac{1}{|S|}\sum_{w\in s}^{n} \text{weight(w)} \quad \dots\dots (2)$$

Represent s is sentence , S is number of words in sentence, w is word in sentence.

<u>Algorithm A:</u>

Input :T//Text

Output: Text Summarization

Step1:preprocessing**:**

- *Tokenization*
- *Remove stopsword or with stop words*
- *Stemming(Get all word roots)*
- *normalize*

Step2: input the text to proposed system

Step4: return the text summarization

Fig 2. Algorithm of the Text Summarization

B: 2-cosine similarity

Cosine Similarity is used to calculate the similarity between documents, as it determines The angle between the vectors is what defines how similar two papers are, and the document with the smallest angles between its vectors is considered to be the most similar [10][11].

$$\text{CosSim}\,(\vec{A}\,,\vec{B}) = \frac{\vec{A}\cdot\vec{B}}{|\vec{A}|\cdot|\vec{B}|} \ldots\ldots\ldots\ldots (3)$$

$$\text{CosSim}(\vec{A}\,,\vec{B})= \frac{\sum_{i=1}^{n} A*B}{\sqrt{\sum_{i=1}^{n} A^2} \ * \sqrt{\sum_{i=1}^{n} B^2}} \ldots\ldots\ldots\ldots (4)$$

3. Evaluation Stage:

A. *Using ROUGE(Recall Oriented Understudy for Gisting Evaluation):* Recall Oriented Understudy for Gisting Evaluation, or ROUGE, as an example: By comparing an abstract to comparable perfect human-written summaries, ROUGE provides measures to automatically assess the quality of the summary. ROUGE measures the number of overlapping units between the ideal human-produced text summary and the machine generated text summary, such as N-grams, word sequences, and word pairings. The N-gram language model forecasts the likelihood that a specific N-gram will appear in any set of words. The value of p (w | h), or the next word in the phrase, can be predicted by a competent N-gram model. Additionally, an N-gram can be defined as a continuous sequence of n items taken from a particular sample of text or speech. Based on the application, the elements may be letters, words, or base pairs. Typically, N-grams are extracted from a text or data source (especially one with a lot of content)[12]. On single document summarization tasks, ROUGE-N "ie for N-gram, N = 2", ROUGE L, ROUGE-W, and ROUGES performed well. The N-gram, the common longest, and the common longest between the perfect (human summary) and the automatically generated summary are the foundation of the ROUGE evaluation technique[13][14].

$$ROUGE-2 = \frac{\sum_{s\in(Ref\ summarization)j} \sum_{bigrams\ i\in S} (\min(\text{count}(\text{i}\ ,\text{X}), \text{count}(\text{i}, \text{S}))}{\sum_{s\in(Ref\ summarization)j} \sum_{bigrams\ i\in S} (\text{count}(\text{i}, \text{S})} \ \ldots (5)$$

for evaluation, We compare algorithm summarization and individuals summarization in a specific way, where the number of bigrams shared between summarizing the algorithm and summarizing the first individual is calculated, and combined with the number of bigrams shared with the second individual, in the same way for the rest of the individuals, then divide all this by the total number bigrams of all individuals.

B. *Using F-measure:* metrics for Information retrieval, (precision, recall, and F-measure)for equations (6,7,8) are used to evaluate Arabic summaries versus human summaries[13]

$$Recal = \frac{system - human\ choice\ overlap}{\text{sentences chosen by human}} \ \ldots\ldots (6)$$

$$Precision = \frac{system - human\ choice\ overlap}{\text{sentences chosen by system}} \ \ldots\ldots (7)$$

$$F\_measure = 2 * \frac{Recall * Precision}{Recall + Precision} \ \ldots\ldots (8)$$

4. The Result

After checking the similarity-based approach and the suitability of the documents and applying the proposed method of taking the sentences with the highest word frequency, the relevant sentences are extracted and the related sentences are combined into one and thus, after merging the data, a final summary is generated. Also, for comparison of different algorithms and techniques on text summarization in related work (introduction part) with we work, as show in table (1):

Table 1. Comparison of different algorithms and techniques on Text Summarization

Author	Algorithm	Language	Different with Similarity Technique
Qaroush, Farha,, Ghanem, Washaha, & Maali[4]	proposed extractive method, evaluates each sentence based on a set of statistical and semantic features	Arabic	Used preprocessing , Not used cosine similarity
Al-Radaideh,& Bataineh,[5]	fuzzy logic	Arabic	Used preprocessing , and used TF-IDF
Al-Radaideh, & Bataineh,[6]	genetic algorithms	Arabic	Used preprocessing

Belkebir, & Guessoum, A.[7]	AdaBoost algorithm	Arabic	Used preprocessing, and not removal stop words

The performance analysis will be evaluated to prove the effectiveness of the proposed methodology by comparing the summary with summarizing three people once, and again two people and calculating the accuracy for both.

Table (2)and(3), shows the number of N-grams shared between the individual summary for each user and the system summary

Table 2. Number of the Shared N-gram (2-Users)

Number of User	Algorithm	Total of N-gram		Shared N-gram	
		U1	U2	U1	U2
2	Cosine _similarity	21	23	19	15
2	Proposed System	20	23	19	15

Table 3. Number of the Shared N-gram (3-Users)

Number of User	Algorithm	Total of N-gram			Shared N-gram		
		U1	U2	U3	U1	U2	U3
3	Cosine _similarity	21	23	5	19	15	5
3	Proposed System	20	23	5	19	15	5

Table (4) and (5), and fig(3) and(4), show the accuracy results of the proposed method and cosine similarity by scale ROUGE.

Table 4. Accuracy for Proposed System using ROUGE

Number of User	Evaluation by ROUGE
2	0.79
3	0.81

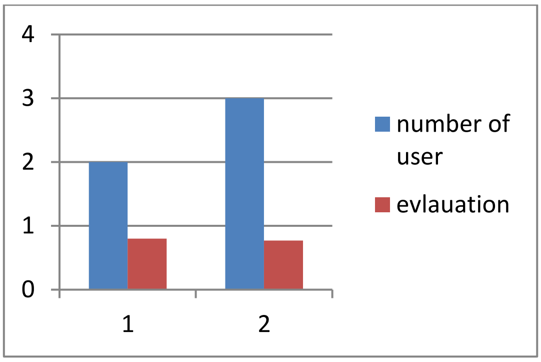

Fig 3. Accuracy with number of users using proposed system

Table 5. Accuracy for Cosine Similarity Using ROUGE

Number of user	Evaluation by ROUGE
2	0.77
3	0.80

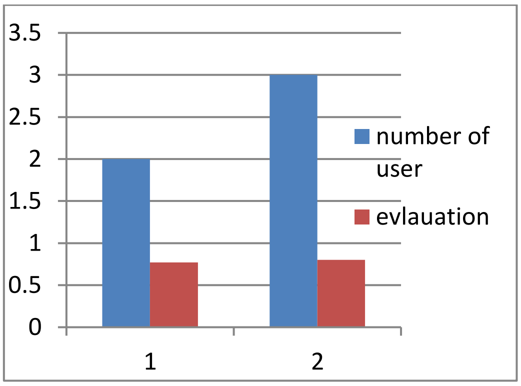

Fig 4. Accuracy with number of users using cosine similarity

Table (6) and(7), and fig(5),and (6) show the accuracy results of the proposed method and cosine similarity by scale recall, precision, F-measure

Table 6. Result for Proposed System using Recall, Precision, F-measure

User	Recall	Precision	F-measure
U1	0.95	0.86	0.90
U2	0.65	0.68	0.66
U3	1.0	0.23	0.37

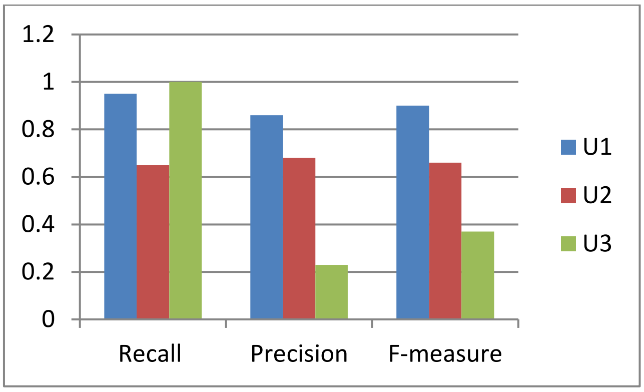

Fig 5. Accuracy for Proposed System using Recall, Precision, F-measure

Table 7. Result for Cosine Similarity using Recall, Precision, F-measure

User	Recall	Precision	F-measure
U1	0.90	0.95	0.93
U2	0.65	0.75	0.70
U3	1.0	0.25	0.4

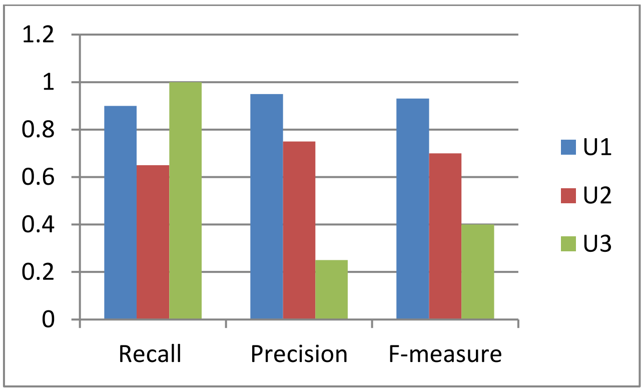

Fig 6. Accuracy for Cosine Similarity using Recall, Precision, F-measure

Acknowledgement

The lab and required documents were provided by the Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq, for which the authors are grateful.

5. Conclusion

The development and increasing growth of data in an organized or unstructured form makes us need to summarize that data and return the relevant data to the specific topic, but in short, and we want a summary of that data in less time. The aim of this paper is to formulate the texts in short so that they are sufficient information for the user without using too many sentences, text summarization saves reading time, facilitates document searches, improves indexing efficiency, and serves question-answering systems. And we applied the extractive summarization, which depends on the texts of the original text, and we take the most frequent sentences for their words by applied system proposed and cosine similarity, and it gave good results by compute scale RUOGE and F-measure.

References

[1]. S. Pawar, & S. Rathod, (2020). Text summarization using cosine similarity and clustering approach (No. 2927). EasyChair.

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

[3]. N. J. Kadhim, & Saleh, H. H. (2018). Improving Extractive Multi-Document Text Summarization Through Multi-Objective Optimization. Iraqi Journal of Science, 2135-2149.

[4]. A. Qaroush , I. A. Farha, W. Ghanem, M. Washaha &E. Maali (2019). An efficient single document Arabic text summarization using a combination of statistical and semantic features. Journal of King Saud University-Computer and Information Sciences.

[5]. L. Al Qassem, D. Wang, H. Barada, A. Al-Rubaie, & Almoosa, N. (2019, September). Automatic Arabic text summarization based on fuzzy logic. In Proceedings of the 3rd international conference on natural language and speech processing (pp. 42-48).

[6]. Q. A. Al-Radaideh, & D. Q. Bataineh (2018). A hybrid approach for arabic text summarization using domain knowledge and genetic algorithms. Cognitive Computation, 10(4), 651-669.

[7]. R. Belkebir & A. Guessoum, (2015). A supervised approach to arabic text summarization using adaboost. In New contributions in information systems and technologies (pp. 227-236). Springer, Cham.

[8]. M. Attia, (2007, June). Arabic tokenization system. In Proceedings of the 2007 workshop on computational approaches to semitic languages: Common issues and resources (pp. 65-72).

[9]. L. S. Larkey, L. Ballesteros & M. E. Connell (2007). Light stemming for Arabic information retrieval. In Arabic computational morphology (pp. 221-243). Springer, Dordrecht.

[10]. F. Rahutomo, T. Kitasuka, & M. Aritsugi, (2012). Semantic Cosine Similarity. In The 7th International Student Conference on Advanced Science and Technology ICAST.

[11]. D. Wali,, &N. Modhe, (2015). Word Sense Disambiguation Algorithms in Hindi.

[12]. V. Siivola,, & M. L. Pellom,. (2005). Growing an n-gram language model. In Ninth European conference on speech communication and technology.

[13]. F. T .AL-Khawaldeh, & V. W. Samawi, (2015). Lexical cohesion and entailment based segmentation for arabic text summarization (lceas). World of Computer Science & Information Technology Journal, 5(3).
[14]. W. Kryscinski, R. Paulus, C. Xiong, &R. Socher, (2018). Improving abstraction in text summarization. arXiv preprint arXiv:1808.07913.