JKAU: Comp. IT. Sci., **Vol. 12** No. 1, pp: 63– 73 (1444 A.H. / 2023 A.D.)
DOI: 10.4197/Comp.12-1.6

Arabic Extractive Summarization Using Pre-Trained Models

Yasmin Einieh, Amal AlMansour, Amani Jamal
Faculty of Computing and Information Technology
King Abdulaziz University
Jeddah, Saudi Arabia
ymohammadeinieh@stu.kau.edu.sa m aalmansour@kau.edu.sa, atjamal@kau.edu.sa

***Abstract*— Automatic Text Summarization (ATS) is a crucial area of study in Natural Language Processing (NLP) due to the vast amount of online information available. Extractive summarization, which involves selecting important sentences from the original document without altering their wording, is one approach to generating summaries. While many methods for Arabic text summarization exist, deep learning applications are still in their early stages, and there is a shortage of available datasets. Unlike English, there have been fewer experiments conducted on Arabic language summarization due to its unique characteristics. This study aims to fill this gap by experimenting with several models for summarizing Arabic text, including QARiB, AraELECTRA, and AraBERT-base models, all trained using the KALIMA dataset. The AraBERT model performed exceptionally well, achieving high scores of 0.44, 0.26, and 0.44 on the ROUGE-1, ROUGE-2, and ROUGE-L measures, respectively.**

***Keywords*— NLP, Extractive Summarization, ATS, Pre-Trained Models, Arabic text Summarization.**

I. Introduction

Through the internet and other platforms, new text is uploaded to repositories every day, leading to an overwhelming amount of information to sift through. As a result, the need for text summarization has become increasingly important. Automatic Text Summarization (ATS) aims to create a comprehensive summary of received data that highlights crucial details

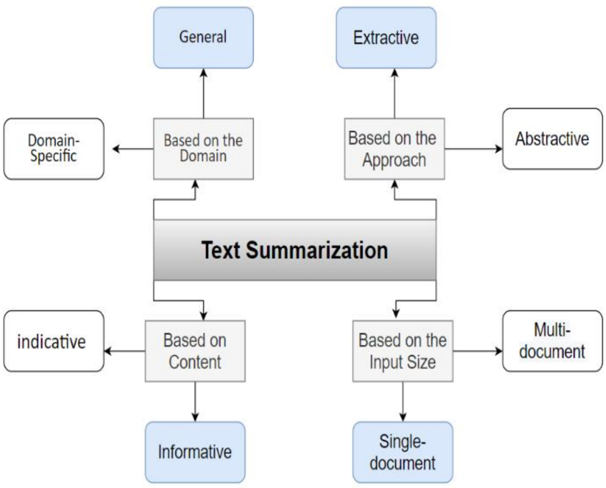

Fig. 1. Automatic Text Summarization Types

[1]. By utilizing ATS, users can quickly understand the key concepts of a document without Generating a summary of a text saves a considerable amount of time and effort by allowing users to understand the essential concepts without having to read the entire document. The summary can be classified based on several factors, such as the domain, input text, content, and method employed [2], as depicted in Figure 1. Summaries can be generated for a single document or multiple documents, depending on the input size. Single-document summarization uses only one input document, while multi-document summarization uses several. Text summarization can also be divided into informative and indicative categories based on content. An informative summary covers all topics and important information in the text, while an indicative summary provides a broad overview of the text's content. Field-based summaries can also be either generic or specific to a certain field, such as scientific papers or medical documents. The approach to text summarization, as described in [1], involves two methods: extractive and abstractive. Extractive summarization selects critical sentences from the input document(s) and combines them to create a summary. In contrast, abstractive summarization presents a preliminary representation of the input document(s) and generates a summary by incorporating new sentences that are not identical to those in the source text. Text summarization can be performed on both single

documents and multiple documents, depending on the input size. Single-document summarization involves summarizing a single document, while multi-document summarization uses a collection of input documents. There are two main types of text summarization based on content: indicative and informative. Indicative summaries provide an overview of the text and its scope, while informative summaries focus on important information in the text and cover all topics. Field-based summaries can be either generic or specific to a certain field, such as medical documents or scientific papers.

The Text Summarization Approach [1] categorizes text summarization into two techniques: extractive and abstractive. Extractive summarization involves selecting essential sentences from the input document(s) and concatenating them into the output summary. On the other hand, abstractive summarization involves representing the input document(s) in a preliminary manner and forming the output summary from this representation. Unlike extractive summaries, abstractive summaries may include sentences that are not identical to those in the source document(s) [2].

Recently, deep learning has gained increasing attention in automatic summarization, especially for the English language. Nonetheless, Arabic does not usually employ it. Mainly Pre-trained models, along with deep learning procedures [3], have improved summarization and other facets of the NLP. For that reason, this research concentrates on deep models for Arabic text summarization. Prior training in the linguistic model has resulted in the creation of state-of-the-art techniques for several NLP applications, including sentiment analysis and named entity recognition [4]. Studies in English employed multiple pre-training models for summarizing [28, 29], but studies in Arabic only tested multilingual BERT [25]. As a result, and to make use of these models' capabilities, we determined to test out many extractive summarization methods in this study. “Bidirectional Encoder Representations from Transformers (BERT)” is an innovative pre-training language representation method developed by Google AI Language researchers. A very large text data set was used to develop a language comprehension model. [5]. Such models have shown excellent efficacy in language understanding by producing precise results in many NLP missions [5]. A fresh Arabic language model depending on BERT named ARABERT was also developed by academics at the start of 2020, and it was evaluated in the disciplines of question-answering and sentiment analysis [6]. This research aims to expand the extractive Arabic text summarization research by introducing this Study.

The paper's structure is as follows: Section 2 offers a literature review, Section 3 presents the methodology, Section 4 shows the results, and Section 5 concludes with final remarks

II. LITERATURE REVIEW

This Prior Arabic research that concentrates on the comprehensive extractive summary of a single document is covered in this part. Many methods to summarize the extractive content are suggested in the literature. Machine learning, semantic, statistical, optimization-based, cluster-based, graph-based, discourse-based, and deep learning-based techniques can all be grouped according to the methodology they use. Moreover, it is feasible to mix different techniques.

A. *Machine-Learning*

Methods Summarizing is considered as a categorization task in machine learning, where sentences are classified as summary or non-summary based on their properties. Guessoum and Belkebir launched a machine learning approach based on AdaBoost. F-measure, recall, and precision were assessed while using their own dataset [9]. Ghanem and his collaborators introduced a machine learning-based method for summarizing Arabic literature. For assessment, they used the Essex Arabic Summaries Corpus (EASC) and ROUGE Corpus [10].

B. *Semantic-Based Methods*

To construct the conceptions of the text, semantic-based methods are concerned with the meaning of words and the connections between words, sentences, and phrases. Put it another way, they look at how the statements are related semantically. A.A. Mohamed summarized a single Arabic text using the Nonnegative Matrix

Factorization (NMF) technique employing a manually assembled dataset of 150 articles. He assessed the approach using Precision, Accuracy, and Recall [12]. Al-Sabahi et al. unveiled a novel "latent semantic analysis" (LSA)-based method for extractive summarization of Arabic text that can identify the latent semantic structure. EASC and Linguistic Data Consortium (LDC) datasets were used. Software operating ROUGE and human source judgment were applied for evaluation that system [13]. Bialy and his associates suggested using an approach based on NLP to condense one Arabic document. The extractive approach is used to identify the information that is the most valuable. 33 Arabic documents were used. Two experts examined the system, and F-measure and ROUGE were used to analyze it digitally [14].

C. *Statistical-Based Methods*

Significant phrases and words are extracted from the text using a statistical analysis of a list of parameters, such as the most prevalent words. In their study, Alami et al. examined the effects of three stemmers on summarizing Arabic text and developed an extraction approach using statistics. They used the cosine similarity measure to gauge how identical each couple of sentences was and then used that information to construct a graphic representation of the input text [11].

D. *An Optimization-Based Method*

These techniques strive to provide a great summary that includes cohesion, diversity, balance, and coverage, and they view summarizing as an improvement challenge. To extract an overview of key Arabic documents, EJaradat and Al-Taani combine the semantic and informative scoring methodologies in a hybrid-base depending on genetic algorithms [17]. For evaluation, they used ROUGE and the EASC Corpus. Additionally, Al-Radaideh and Bataineh employed a hybrid strategy. They combined domain knowledge and genetic algorithms [18]. KALIMAT and EASC Corpus were used, along with ROUGE. Furthermore, by a particle swarm optimization algorithm, Al-Abdallah and Al-Taani devised a method for condensing a single Arabic document. They used semantic and informative scores to enhance the accuracy of the summarizing operation. They employed ROUGE as well as the EASC Corpus [19].

E. *Cluster-Based Method*

Organize items (sentences) into groups based on their similarity. The multi-documents content in Arabic was condensed by Fejer and Omar via Clustering and Keyphrase Extraction. Similar papers (from a variety of sources) are gathered for the purpose of key phrase extraction, and the extracted key phrases are then used to identify the key phrases. They employed ROUGE and the DUC2002 corpus [15]. To identify the sentences for summary picking that are most closely associated towards the centroid, Abu Nada et al. use the BERT model for text word embedding and K-Means grouping [16].

F. *Graph-Based Method*

These techniques treat the document like a graph. The edges linking the sentences in this connected graph indicate how identity two phrases are to one another. Sentences are illustrated by vertices in the graph. For the 2D graph, Alami et al. applied the PageRank classification algorithm. Semantic data was measured using WordNet. The statistical scale is based on how much the substance of two phrases overlaps. The EASC dataset was utilized. Precision, F-measure, and recall were employed for evaluation. [21]. Additionally, Elbarougy et al. suggested a method for employing the Modified PageRank algorithm to condense a single document in Arabic. The EASC dataset was utilized. F-measure, recall, and Precision were employed for evaluation. [22]. Al-Abdallah et al. introduced a graph-based method to use the Firefly algorithm to condense the text in Arabic of a single document. They employed ROUGE and the EASC dataset was used for evaluation. [23].

G. *Discourse-Based Method*

To maintain the discourse's coherence, the text is processed as discourse units rather than as a series of phrases and words. Ibrahim and Elghazaly created the RS-Tree and extracted the most significant sentence to serve as a summary using Rhetorical Structure Theory (RST). They used precision for their evaluation and the BBC's web Arabic news as a dataset [20].

H. *Deep Learning Based Method*

The Arabic content was condensed by Alami et al. applied an unsupervised deep-learning algorithm. The application of deep learning to the Arabic language began with this work. Using high-dimensional input data and the variable auto-encoder model (VAE), the feature area was defined. They made use of both their own data set and the EASC. They evaluated using ROUGE, and the technique scored 0.660 [24]. Also, Elmadani and his associates have used the pre-trained BERT model, particularly the multilingual BER, for both extractive and abstractive summarization [25].

Arabic literature has been demonstrated in numerous research. Some of these techniques, such as cluster-based and optimization-based techniques, are more applicable for handling multiple documents than others. The hybrid approach, which integrated several approaches, produced successful outcomes. This result can be because of the complex nature of the task of summarizing, which incorporates phrase and word scoring, text evaluation, tokenization, segmentation, etc. Given that Arabic is a distinct language and the deficit of existing data, there is a gap in summarization when compared to English. The majority of the ATS works now being produced in Arabic take a lot of effort and computation. The extraction summary's matching and accuracy are still being worked on by researchers. There has been a recent increase in the use of deep learning for automatic summarization, particularly in English, with promising outcomes [36]. The topic is still developing and needs more study, attention, and growth; hence it is still not widely used in Arabic. Considering pre-trained models are simple to use and have good results across a range of NLP applications, we focused on them in this research. Moreover, the M-BERT model was the only Arabic model that had its automatic summarization evaluated [25].

III. METHODOLOGY

In this section, we will outline our research design and methodology, which includes four subsections. The first subsection focuses on the models used in our study. The second subsection describes the dataset we utilized for our experiments. The third subsection details the proposed approach we employed for summarizing Arabic text. Finally, the fourth subsection discusses the performance evaluation metrics we used to measure the effectiveness of our approach.

A. *Models*

Pre-trained models refer to deep learning models that have undergone training on large data sets and are capable of performing specific natural language processing (NLP) tasks. These models can acquire universal language representations when trained on a vast corpus, which can be beneficial for downstream NLP tasks and eliminate the need to start training a new model from scratch. Consequently, pre-trained models are reusable and can be utilized by developers to quickly build an NLP application [3].

- BERT [5]

The BERT (Bidirectional Encoder Representations from Transformers) deep learning model is built on the Transformer architecture, and it connects each input element to each output element by dynamically determining the weights between them based on their connection. BERT employs a stack of encoder layers that are stacked on top of each other, while the "BERT base" and "BERT big" models differ in the number of transformer layers, attention layers, and parameters. The transformer layer combines encoder and decoder layers, as well as any intermediate connections. Unlike previous language models that could only interpret text input in one direction, BERT is unique in that it can read in both directions simultaneously. This bidirectionality was made possible by the development of Transformers.

B. *Dataset*

In this subsection, we will focus on the datasets used for extractive summarization in our study, specifically KALIMAT and EASC [13], which contain Arabic text.

- KALIMAT: a Multipurpose Arabic Corpus [26]

An Arabic corpus called KALIMAT is utilized for extractive summarization. Arabic natural language processing is a key challenge in ANLP due to the shortage of Arabic resources. KALIMAT is considered a Multipurpose Arabic Corpus Dataset which comprises twenty-two thousand two hundred ninety-one papers from the

Omani daily Alwatan [26]. The data sets are divided into six categories: local news, global news, the economy, sports, culture, and religion as displayed in Table I. Each topic of data set in the raw has its text document with articles per month.

TABLE I. KALIMAT DATASET

Subject	Articles count	Summaries count
Economy	3.468	3.468
International-News	2.035	2.035
Culture	2.782	2.782
Local News	3.596	3.596
Religion	3.860	3.860
Sports	4.550	4.550
The Overall	20.291	20.291

- The Essex Arabic Summaries Corpus (EASC) [27].

The extractive summarization of Arabic literature was the purpose behind the creation of the EASC dataset [27]. A total of 765 summaries have been collected from its 153 Arabic papers, each of which has five distinct summaries. Articles on subjects such as art, politics, teaching, religion, the environment, health, investing, sports, technology, and knowledge are available. See TABLE II.

Since the KALIMAT dataset is more extensive than EASC and suitable for model training, we started using the KALIMAT dataset. The results were the same when EASC and KALIMAT were merged.

TABLE II. EASC DATASET

Articles	summaries
153	765

C. *The Proposed Approach*

In this section, we introduce a proposed summarization approach for the Arabic language. Our approach builds upon and extends the solution presented in [29] to accommodate different models and work with Arabic datasets. Specifically, we fine-tune pre-trained transformer-based models QARiB[8], AraBERT [6], and AraELECTRA [7] on KALIMAT and EASC datasets. The method involves several basic stages as described in [28][29][31] and illustrated in Fig. 2.

However, summarizing with BERT is not straightforward as its output vectors are token-based rather than sentence-based due to its masked-language model training. While BERT's segmentation embeddings can represent different sentences, they only apply to sentence pair inputs, making it challenging to encode and summarize multi-sentence inputs.

The Stanford CoreNLP library was utilized to pre-process the input, including tokenization and sentence splitting [32]. We have two options that indicate whether a specific sentence will be chosen or not for a set of sentences (sentence 1, sentence 2, sentence n). We can suppose that extractive summarization entails the work of binary classification, in which each sentence is given a label indicating whether it should be included in the summary or not. The BERT encoder and the summarizing layer are the two components of the BERT summarizer. Tokens [CLS] and [SEP] must be added before and after each sentence, respectively. The last hidden layer of these [CLS] tokens will be used to represent our sentences after the encoder has completed a forward pass. After we have each sentence's vector representation, we can use a straightforward feed-forward layer as our classifier to assign each sentence a score. In this study, a 3-layer, compact Transformer model was used. The sigmoid classifier is the last output layer. Hence, we learn the interactions between our document's tokens in the encoder and its sentences in the summarization layer. See Fig.3.

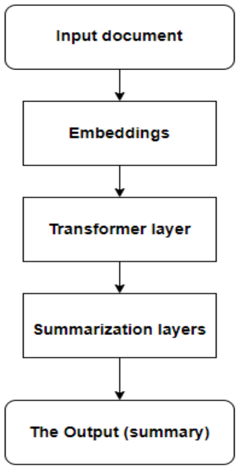

Fig.2. Basic Stages of Proposed Approach

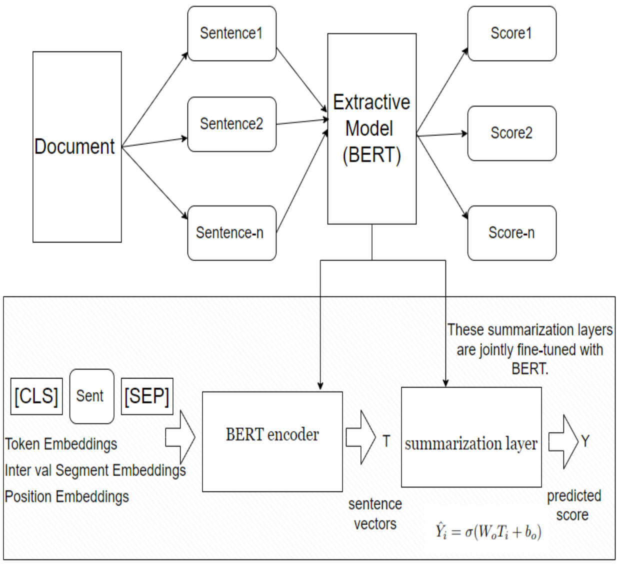

Fig. 3. The Proposed Approach

D. Performance evaluation metric [33]

The evaluation metric that will be used to rate the suggested strategy will be introduced in this section. ROUGE is a set of standards and instruments used in NLP to gauge machine translation and automated summarization programs. ROUGE's primary goal is to contrast various reference summaries written by people with computer-generated summaries. The most widely used technique for analyzing automatically generated summaries is the ROUGE Metric, which counts the number of overlapping units, such as overlapped n-grams, between the system-generated summary and benchmark summaries. ROUGE measurements are employed in various instances, such as using unigrams for the contrasting nominee and benchmark summaries in ROUGE-1 (R1) or comparing nominee and benchmark summaries using an n-gram recall mechanism in ROUGE-N. The largest trailing joint between the reference and candidate reference abstract is used in the ROUGE-L (R-L) technique. Since its introduction, ROUGE has become a standard for determining summarizing model accuracy. Its disadvantage is that it only compares strings between summaries, not specific word or phrase meanings [33].

IV. Results

In this section, we describe our findings and experiences related to our implementations. Initially, we utilized the BERTSum model implementation from a previous study [28], which was based on the OpenNMT framework [34]. However, we made some modifications to the original implementation to better suit our objectives of utilizing multiple pre-trained models. To achieve this, we employed the transformers library provided by Hugging Face [35], which includes standard PyTorch implementations of various Transformer-based models. We conducted several tests to compare and evaluate the performance of different pre-trained models for extractive text summarization. Specifically, we investigated three trained models:

1. AraELECTRA [7]
2. QARiB [8]
3. AraBERT [6]

We experimented with different hyperparameters such as learning rates (1e-4, 2e-5, 2e-3, 5e-3) and the number of training steps (40000, 50000, 60000) for fine-tuning. Based on the results obtained from the training set, we selected the optimal hyperparameters that produced the best results. Although there was only a slight difference in results based on the number of steps used for training, we chose 50,000 epochs as suggested in [28]. Additionally, our experiments showed that using a learning rate of 2e-3 produced the best results compared to other values tested. For fine-tuning purposes, we employed Adam with $\beta1 = 0.9$ and $\beta2 = 0.999$ as recommended in

[28].

Table III. and Fig.4. displayed the tests and findings from using several models. The results are considered good, and we observe that they are rather like one another. The QARiB model, which had the poorest results, made a significant difference, and its training on tweets may have contributed to this. It is important to note that the AraBERT model performed better than the other models and produced better results because it was specifically trained using Arabic data. For AraBERT versions, Arabertv0.2- Basic yields the best performance.

TABLE III. RRESULTS OF THE DIFFERENT MODELS ON KALIMAT DATASET

Model	R1	R2	RL
AraELECTRA	0.421	0.243	0.421
QARiB	0.345	0.189	0.345
AraBERTv1-base	0.421	0.237	0.421
AraBERTv2-base	0.444	0.258	0.444
AraBERTv0.2 base	0.440	0.262	0.440
AraBERTv0.2 Twitter-base	0.439	0.253	0.439

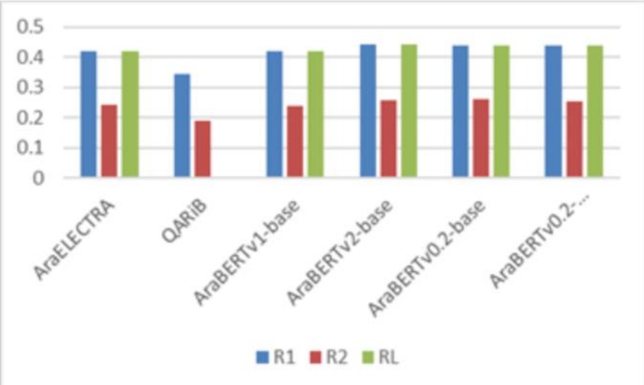

Fig. 4. Results of the different models on KALIMAT dataset

Furthermore, we combined the KALIMAT and EASC datasets and conducted fine-tuning to the selected models. The utilization of solely the KALIMAT and EASC datasets did not yield any noteworthy enhancements in the previous outcomes. The findings indicate a convergence between the Arabert model and the AraELECTRA model, but the AraBERT model outperforms the AraELECTRA model. See Table IV. and Fig.5.

TABLE IV. RRESULTS OF THE DIFFERENT MODELS ON KALIMAT AND EASC DATASETS

Model	R1	R2	RL
AraELECTRA	0.420	0.244	0.420
QARiB	0.329	0.175	0.328
AraBERTv2-base	0.428	0.230	0.427
AraBERTv0.2-base	0.440	0.254	0.439
AraBERTv0.2-Twitter-base	0.433	0.252	0.433

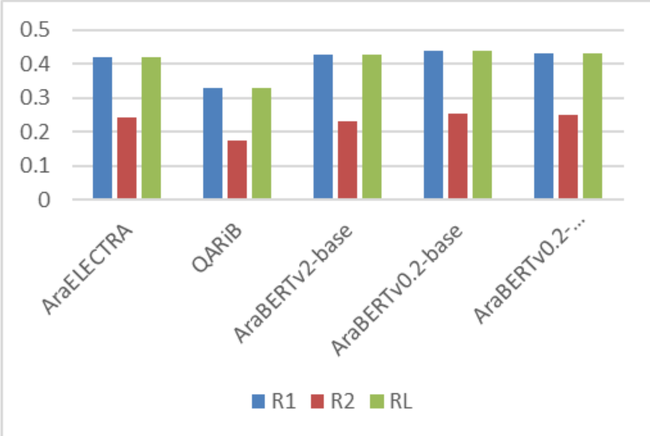

Fig.5. Results of the different models on KALIMAT and EASC datasets

We selected this research [25] to compare our work with the alternative model because it is the only study to our knowledge that used a pre-trained model and applied the same KALIMAT dataset that we used. The study used the M-BERT model and a non-trained Transformer. Table 5. and Fig.6. present our results compared to those achieved by the MBERT model in the previous study [25]. The results showed that the AraBERT model is superior to the M-BERT model in all measures.

TABLE V. RESULTS OF THE DIFFERENT MODELS ON KALIMAT AND EASC DATASETS

Model	R1	R2	RL
M-BERT[25]	0.42	0.24	0.41
TRANSFORMER EX[25]	0.28	0.14	0.28
AraBERTv0.2-base(ours)	0.44	0.26	0.44

Article

تتعدد أنواع الحواسيب من حيث طريقة عملها وحجمها بالإضافة إلى سرعتها، فأوائل الحواسيب الإلكترونية كانت بحجم غرفة كبيره وتستهلك طاقة مماثلة لما يستهلكه بضعة مئات من الحواسيب الشخصيّة اليوم. كما أن السنوات الأخيره شهدت انخفاضاً في تكاليف صناعة البنية الصلبة إلى الحد الذي أصبحت معه الحواسيب الشخصية سلعة منتشره بشكل كبير. كما تنوعت تطبيقات الحواسيب وتوسعت لتشمل مختلف المجالات والأجهزه في وقتنا الحالي، فصنعت الساعة الذكية، وطبقت أنظمة الملاحة الإلكترونية بشكل واسع عن طريق نظام التموضع العالمي وأصبحت برامجه وأجهزته في متناول الجميع، كما أن كثيرًا من رجال الأعمال يهتمون بتطبيقها في أعمالهم التجارية لتقليل الأيدي العاملة وتخفيض تكلفة الإنتاج. وينظر المجتمع إلى الحاسوب الشخصي)(الحاسوب المكتبي) ونظيره المتنقل (الحاسوب المحمول) على أنهما رمزي عصر المعلومات؛ فهما أول ما يتبادر إلى ذهن أي شخص تقريبا عند الحديث عن الحاسوب. ومع هذا فأكثر أشكال الحاسوب استخدامًا اليوم هي الحواسيب المضمّنة وهي الحواسيب المضمنة في أجهزه صغيره وبسيطة تستخدم عاده للتحكم في أجهزه أخرى، فعلى سبيل المثال يمكنك أن تجدها في آلات شتى مثل الطائرات المقاتلة، والآليين، وآلات التصوير الرقمية إلى لعب الأطفال، وأجهزه التحكم.

لا يمكن القول بأن الحاسوب هو اختراع بحد ذاته، لأنه كان نتاج الكثير من الابتكارات العلمية والتطبيقات الرياضية. الحواسيب متنوعة في الواقع، وطبقًا لفرض تشرش في آلة تورنغ فإن حاسوبًا له قدره ذات حد منخفض يكون قادرًا على إنجاز المهام الخاصة بأي حاسوب آخر، بدءاً من المساعد الرقمي الشخصي إلى الحاسوب الفائق، طالما أن الوقت وسعة الذاكره ليست في الاعتبار. لذلك فإن التصميمات المتماثلة من الحاسوب من الممكن أن تضبط من أجل القيام بمهام بسيطة كمعالجة حسابات موظفي الشركات أو معقده كالتحكم في المركبات الفضائية دون طيار. وبسبب التطور التقني فإن الحواسيب الحديثة تكون بشكل جبري أكثر قدره من الحواسيب السابقة، وهي ظاهره موصوفة ومشروحة جزئيا في قانون مور.

Summary

و ينظر المجتمع الى الحاسوب الشخصي)(الحاسوب المكتبي) و نظيره المتنقل (الحاسوب المحمول) على أن هما رمزي عصر المعلومات ; فهما اول ما يتبادر الى ذهن اي شخص تقريبا عند الحديث عن الحاسوب . الحواسيب متنوعة في الواقع , و طبقا لفرض تشرش في الة تورنغ فان حاسوبا له قدره ذات حد منخفض يكون قادرا على انجاز المهام الخاصة باي حاسوب اخر , بدءا من المساعد الرقمي الشخصي الى الحاسوب الفائق , طالما ان الوقت و سعة الذاكره ليست في الاعتبار .و بسبب التطور التقني فان الحواسيب الحديثة تكون بشكل جبري اكثر قدره من الحواسيب السابقة , و هي ظاهره موصوفة و مشروحة جزئيا في قانون مور .

Fig. 7. An example of AraBERTv0.2 results

Fig.7. displays a summary that the AraBERT model generated. For using examples to assess the model's performance, we used a random assortment of Wikipedia articles. The suggested method results in an effective summary by emphasizing the important portions of the article and producing a concise summary.

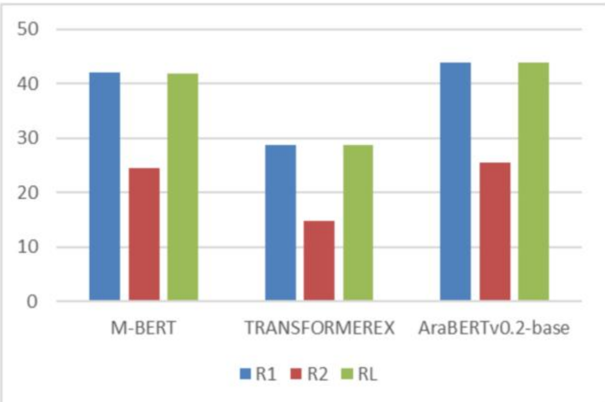

Fig.6. AraBERT model results vs other models

In conclusion of prior studies on extractive summarization of Arabic text using pre-trained models, we can concisely state the following:

- The results of the AraBERT model were better than those of the M-BERT model, validating the earlier discoveries made by Antoun that the model trained just on the Arabic language is superior to the multilingual one [17].
- AraBERT achieves better outcomes than AraELECTRA in Arabic text extractive summarization tasks.
- Arabertv0.2 is the most effective version of AraBERT. A richer vocabulary, greater training, and additional data are all included in this edition.

V. CONCLUSIONS

While many studies have been conducted on artificial summarization of the English language to identify the most advanced approaches and achieve optimal results, progress in artificial Arabic content summarization has been slow due to the unique features of the Arabic language and the lack of significant source datasets. Recently, several pre-trained language models have shown great accomplishment on several NLP tasks. To summarize the Arabic content that was gathered for this study, we trained the AraBERT model and other models using the KALIMAT dataset. The models performed very well when we used ROUGE for evaluation. To leverage the abilities of linguistic models, it is crucial to provide a dataset for abstractive summarization in Arabic text. Moreover, efforts should be made to include additional parts of the summary, such as consistency and coherence, in the summary review. Even in Arabic, there is still much opportunity for innovation and development. We only focused on extractive summarization; future research may additionally investigate abstractive summarization methods. Investigating techniques that combine extractive and abstractive methods would be interesting.

VI. References

[1] Al Qassem, L. M., Wang, D., Al Mahmoud, Z., Barada, H., Al-Rubaie, A., Almoosa, N. I. (2017). Automatic Arabic summarization: a survey of methodologies and systems. Procedia Computer Science, 117, 10-18.

[2] El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2020). Automatic text summarization: A comprehensive survey. Expert Systems with Applications, 113679.

[3] Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 63(10), 1872-1897.

[4] Magdum, P. G., & Rathi, S. (2021). A Survey on Deep Learning-Based Automatic Text Summarization Models. In Advances in Artificial Intelligence and Data Engineering (pp. 377-392). Springer, Singapore.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova Google, and A I Language. BERT: Pre-training of Deep Bidirectional Transformers fo Language Understanding. Technical report

[6] Wissam Antoun, Fady Baly, and Hazem Hajj. AraBERT: Transformer-based Model for Arabic LanguageUnderstanding, 2020.

[7] Wissam Antoun, Fady Baly, and Hazem Hajj. AraELECTRA: Pre-Training Text Discriminators for ArabicLanguage Understanding. 12 2020.

[8] Abdelali, A., Hassan, S., Mubarak, H., Darwish, K., & Samih, Y. (2021). Pre-training Bert on Arabic tweets: Practical considerations. arXiv preprint arXiv:2102.10684.

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

[10] Qaroush, A., Farha, I. A., Ghanem, W., Washaha, M., & Maali, E. (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.

[11] Alami, N., Meknassi, M., Ouatik, S. A., & Ennahnahi, N. (2016, October). Impact of stemming on Arabic text summarization. In 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt) (pp. 338-343). IEEE.

[12] Mohamed, A. A. (2016, December). Automatic summarization of the Arabic documents using NMF: A preliminary study. In 2016 11th International Conference on Computer Engineering & Systems (ICCES) (pp. 235-240). IEEE.

[13] Al-Sabahi, K., Zhang, Z., Long, J., & Alwesabi, K. (2018). An enhanced latent semantic analysis approach for arabic document summarization. Arabian Journal for Science and Engineering, 43(12), 8079-8094.

[14] Bialy, A. A., Gaheen, M. A., ElEraky, R. M., ElGamal, A. F., & Ewees, A. A. (2020). Single arabic document summarization using natural language processing technique. In Recent Advances in NLP: The Case of Arabic Language (pp. 17-37). Springer, Cham.

[15] Fejer, H. N., & Omar, N. (2014, November). Automatic Arabic text summarization using clustering and keyphrase extraction. In Proceedings of the 6th International Conference on Information Technology and Multimedia (pp. 293-298). IEEE.

[16] 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.

[17] 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.

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

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

[20] Ibrahim, A., & Elghazaly, T. (2013, November). Improve the automatic summarization of Arabic text depending on Rhetorical Structure Theory. In 2013 12th Mexican International Conference on Artificial Intelligence (pp. 223-227).IEEE

[21] Alami, N., El Adlouni, Y., En-nahnahi, N., & Meknassi, M. (2017, March). Using statistical and semantic analysis for Arabic text summarization. In International Conference on Information Technology and Communication Systems (pp. 35-50). Springer, Cham

[22] Elbarougy, R., Behery, G., & El Khatib, A. (2020). Extractive arabic text summarization using modified PageRank algorithm. Egyptian Informatics Journal, 21(2), 73-81.
[23] 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.
[24] Alami, N., En-nahnahi, N., Ouatik, S. A., & Meknassi, M. (2018). Using unsupervised deep learning for automatic summarization of Arabic documents. Arabian Journal for Science and Engineering, 43(12), 7803-7815.
[25] Elmadani, K. N., Elgezouli, M., & Showk, A. (2020). BERT Fine-tuning For Arabic Text Summarization. arXiv preprint arXiv:2004.14135
[26] El-Haj, M. and Koulali, R. [2013], Kalimat a multipurpose arabic corpus, in'Second workshop on Arabic corpus linguistics (WACL-2)', pp. 22–25.
[27] El-Haj, M., Kruschwitz, U. and Fox, C. [2010], 'Using mechanical turk to create a corpus of arabic summaries
[28] Liu, Y. (1903). Fine-tune BERT for extractive summarization. 2019. arXiv preprint arXiv:1903.10318.
[29] Liu, Y., & Lapata, M. (2019). Text summarization with pre-trained encoders. arXiv preprint arXiv:1908.08345. pre-trained transformer-based language models.
[30] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. ELECTRA: Pre-training TextEncoders as Discriminators Rather Than Generators. arXiv, 3 2020.
[31] Vashisht, Ashutosh. "BERT for Text Summarization." OpenGenus IQ: Learn Computer Science, OpenGenus IQ: Learn Computer Science, 27 July 2020, iq.opengenus.org/bert-for-text-summarization/.
[32] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations (pp. 55-60).
[33] Lin, C. (2004). Rouge: A package for automatic evaluation of summaries. In: Marie-Francine Moens SS (ed) Text Summarization Branches Out: Proceedings of the ACL-04 Workshop.,74-81.
[34] Klein, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A. M. (2017). Opennmt: Open-source toolkit for neural machine translation. arXiv preprint arXiv:1701.02810.
[35] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2019). Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771.
[36] Suleiman, D.,& Awajan, A. A. (2019, October). Deep learning based extractive text summarization: approaches, datasets and evaluation measures. In 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 204-210). IEEE
[37] Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual BERT?. arXiv preprint arXiv:1906.01502

نهج قائم على المحولات المدربة مسبقاً للتلخيص الاستخراجي للنص العربي

ياسمين عينيه [١]، أمل المنصور [٢] ، أماني جمال [٣]

[١] قسم علوم الحاسبات ، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة ، المملكة العربية السعودية

[2] قسم علوم الحاسبات ، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة ، المملكة العربية السعودية

[3] قسم علوم الحاسبات ، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة ، المملكة العربية السعودية

المستخلص. التلخيص التلقائي للنص (ATS) هو موضوع بحث بارز في معالجة اللغة الطبيعية (NLP) بسبب تنوع وانتشار مصادر المعلومات على الإنترنت. يحدد التلخيص الاستخراجي الجمل الأساسية من المستند لتشكيل ملخص دون تغيير كلمات الجمل. تم استخدام العديد من الأساليب في تلخيص النص العربي ، ولا يزال استخدام التعلم العميق في مهده. النماذج المدربة مسبقًا مثل BERT و ELECTRA هي نماذج التعلم العميق التي تم تطويرها مؤخرًا ، مما أدى إلى تحسينات كبيرة في العديد من مهام البرمجة اللغوية العصبية ، خاصة في اللغة الإنجليزية. إلا أن التجارب باللغة العربية لا تزال أقل من اللغة الإنجليزية بسبب طبيعة اللغة العربية وقلة مجموعات البيانات المتاحة. تهدف هذه الدراسة إلى تجربة نماذج مختلفة لتلخيص النص العربي. قمنا بضبط أداء نموذج AraBERT الأساسي ، و نموذج QARiB ، ونموذج AraELECTRA ومقارنتهم ببعض. تم تدريب هذه النماذج باستخدام مجموعة البيانات العربية KALIMAT في سياق تلخيص النص. كانت أفضل النتائج من نموذج AraBERT ، باستخدام مقاييس Rouge-1 و Rouge-2 و Rouge-L التي حصلت على ٠,٤٤،٠,٢٦،٠,٤٤ في مجموعة بيانات KALIMAT على التوالي.
الكلمات المفتاحية- معالجة اللغة الطبيعية ، التلخيص التلقائي للنص ، التلخيص الاستخراجي ، تلخيص النص العربي.