Fine-tuning and multilingual pre-training for abstractive summarization task for the Arabic language

Mram Kahla[a], Attila Novák[ab], Zijian Győző Yang[bc]

[a]Pázmány Péter Catholic University, Faculty of Information Technology and Bionics
{kahla.mram,novak.attila,yang.zijian.gyozo}@itk.ppke.hu

[b]MTA-PPKE Hungarian Language Technology Research Group

[c]Hungarian Research Centre for Linguistics
yang.zijian.gyozo@nytud.hu

Abstract. The main task of our research is to train various abstractive summarization models for the Arabic language. The work for abstractive Arabic text summarization has hardly begun so far due to the unavailability of the datasets needed for that. In our previous research, we created the first monolingual corpus in the Arabic language for abstractive text summarization. Based on this corpus, we fine-tuned various transformer models. We tested the PreSumm and multilingual BART models. We achieved a "state of the art" result in this area with the PreSumm method.

The present study continues the same series of research. We extended our corpus "AraSum" and managed to reach up to 50 thousand items, each consisting of an article and its corresponding lead. In addition, we pre-trained our own monolingual and trilingual BART models for the Arabic language and fine-tuned them in addition to the mT5 model for abstractive text summarization for the same language, using the AraSum corpus. While there is room for improvement in the resources and the infrastructure we possess, the results clearly demonstrate that most of our models surpassed the XL-Sum which is considered to be state of the art for abstractive Arabic text summarization so far. Our corpus "AraSum" will be released to facilitate future work on abstractive Arabic text summarization.

Keywords: Arabic, mT5, BART, AraSum, Abstractive Summarization

AMS Subject Classification: 68T07, 68T50

1. Introduction and motivation

Automatic text summarization means teaching the machine to subtract information from a text and provide a shorter overview of it. We distinguish between two methods of text summarization. The first one is the extractive [18], in which we select parts from the text that can function as a summary, this is practically a classification task. The other method is abstractive [19], where, like humans, the model independently generates a summary from a given text and sometimes uses terms that were not included in the original text. Recent advances in the field are usually utilizing abstractive models to get better summaries.

The focus of our research is the Arabic language, one out of the 6 languages which the U.N recognizes as an official language. Given Arabic is complicated, it raises a number of challenges in the present field of research.

Arabic is a morphologically and structurally diverse language. First of all, we should keep in mind that there is a massive difference both between the written Arabic and the spoken language and between the numerous dialects themselves. A matter often not addressed properly causes confusion in the subject. While the spoken form of the universal written Arabic – Fusha – is practically unused in the daily general parlor the dialects used in its stead are extremely diverse many times even within the same country. Grasping all these forms and the variables they present is usually beyond the capabilities of the native speakers themselves. Thus creating a lingual cacophony within the same linguistic realm.

Despite all these challenges the great benefit of Arabic, at least in its written form we see, is that in its complexity it is one and the same all over the Arabic world. So while addressing spoken Arabic in all its diversity would be an immense task the uniform nature of its written form makes text summarization not only possible but even reliable.

The main contributions presented in this paper include a) presenting the extended version of the first monolingual corpus 'AraSum' for abstractive Arabic text summarization, b) pre-training the monolingual BART model and the trilingual BART model including Arabic, English, and Hungarian for the Arabic language, and c) fine-tuning the mT5 model for abstractive Arabic text summarization.

The rest of the paper is structured as follows. Section 2 presents related work published on Arabic summarization and available corpora. Section 3 describes the AraSum corpus source and its characteristics. Section 4 describes the models used for training and fine-tuning, and section 4 describes the experiments we have done. Section 5 presents the results. The last section concludes the paper.

2. Related work

The work for Arabic summarization is limited. Most existing systems use the extractive approach. Lakhas [4] is considered the first extractive Arabic summarization system that produces a 10-word summary and translates it to English and

then it is evaluated using the ROUGE measure [14]. Another Arabic text summarization approach based on fuzzy logic was proposed by [1]. SumSAT [12] adopts an extractive approach using a hybrid of three techniques: a) *contextual exploration*; b) identification of *indicative expressions*, and c) the *graph method.*

In terms of abstractive summarization, one research [2] proposed a four-phase abstractive summarizer for Arabic where the core of the system is an extractive summarizer. Another research proposed by [17] was trained to generate headlines based on the first paragraph of Arabic articles, a task that can be classified as a kind of abstractive summarization. Using the PreSumm method[15] and the multilingual BERT model [3], [5] fine-tuned both extractive and abstractive models.

Abstractive datasets for any language other than English are still scarce. So, the progress in abstractive summarization for the Arabic language has hardly ever been scratched. There are two extractive datasets available for the Arabic language. The first is the Essex Arabic Summaries Corpus (EASC) [8], which contains 153 Arabic articles and 765 human-generated extractive summaries of those articles created using Mechanical Turk. The second is the KALIMAT dataset [7], which contains 20,291 machine-generated article summaries output by the extractive Gen-Summ (=AQBTSS) algorithm [8]. For abstractive datasets, there is a headline generation dataset which was presented in [17] where they crawled an Arabic dataset consisting of approximately 300 thousand *article headline: introductory paragraph* pairs. This can be classified as a kind of abstractive dataset. In addition, there is the WikiLingua dataset [11] which is a multilingual abstractive summarization dataset in 18 languages including Arabic. It contains articles and their summaries from WikiHow[1]. A majority of the non-English articles are translated from the English versions to the target language. The Arabic part includes summaries for 29,229 articles. A large-scale dataset crawled from the BBC news site, XL-Sum [9] also includes Arabic news summarization data. A multilingual summarization model was created using the whole corpus.

Available corpora in the field of Arabic text summarization were either extractive or part of a multilingual abstractive dataset. There was no major monolingual corpus to work within the field of abstractive Arabic text summarization. This was the first thing we created in our previous research[10]. Using this corpus, we experimented with the PreSumm abstractive summarization method [15]. In addition, we fine-tuned the multilingual mBART-50 [22] model. We improved the performance of the system with cross-lingual fine-tuning: using fine-tuning on a summarization dataset in another language before further fine-tuning on Arabic. The evaluation was performed in terms of ROUGE, and for the sake of a more accurate assessment, we conducted a human evaluation of fluency and adequacy. Our results were the best compared with other models for Arabic, but compared to the results in other languages like English they are weak since we used a relatively small corpus.

---

[1]https://www.wikihow.com

3. AraSum corpus

Looking for a stable dataset, the most ideal source proved to be the press. Such is the case with the trend-setting dataset by CNN/Daily Mail dataset [21]. That is because most articles include a lead, a short summary of the given articles. The ideal lead summary, which is usually two or three sentences maximum, sums up the article not altering the general meaning. That, however, raised the problem of finding articles with high-quality abstractive leads, as these are not easy to find, especially in big quantities. What we found is that, in a majority of Arabic sources, the lead is only a direct copy of the article's first paragraph. Also many times the lead provides clickbait terms, or sentences, but has little in common with the general tone of the article. Therefore we cannot really rely on these, as these are far from being good abstractive leads.

The focus of our attention, therefore, turned to the evaluation of Arabic versions of global news channels. These included *CNN, BBC, France 24, DW*, and *Sky News*. Also, a number of popular Arabic-language news sites were considered. Amongst these, we looked into the sites of *al-Mayadeen, al-Ālam, al-Ahrām, al-Jazeera, al-Arabiyya*, and *Sada-elbalad*.

Eventually, we managed to identify two Arabic news sources ideal for the Arabic abstractive news summaries dataset. One of them is the Arabic version of the German *Deutsche Welle (DW)* news website[2]. DW is also a public state-owned international broadcaster, its satellite television service also includes a channel in Arabic. DW has the best abstractive Arabic-language summaries we could find so far.

The other resource we found promising was the *Files* section of the website named *Sada-elbalad*[3]. However, Sada-elbalad later proved to be problematic due to the fact that many 'files' contained several widely diverse topics. In these cases, the articles address a much wider range of matters, than is mentioned in the summary, i.e. the summary lacks key information present in the article. Therefore we decided to omit Sada-elbalad when creating the news summary database.

We chose to download Arabic Deutsche Welle resources from Common Crawl[4], as this solution does not interfere with the site. A very positive aspect of this resource is that the items downloaded from the DW news website address a wide range of topics, not only dealing with politics, sports, or art. This allows the summary database to cover a wide range of topics, making more realistic testing of the capabilities of the summarization models as well as the creation of more robust and less domain-dependent models possible.

We performed data processing steps on the collected articles to be ready for the abstractive summarization task. We needed to perform text tokenization to use our corpus with some language models (e.g. for training PreSumm-based models), for that, we used the NLTK platform.

---

[2]https://www.dw.com/ar
[3]https://www.elbalad.news/category/2065
[4]https://commoncrawl.org/

We presented the first monolingual corpus of human-written abstractive news summaries in Arabic "AraSum". In our previous research, we compiled the first version of the dataset consisting of more than 21000 items. The version we present in this paper contains 50525 articles and their corresponding leads, which We randomly split into (a) train, validation and test sets (90/2/8 split).

4. Models used

In this research, we trained a monolingual and a trilingual BART model, and we fine-tuned these and the mT5 model for abstractive text summarization for the Arabic language.

4.1. The BART model

The BART model [13] is a transformer model with an encoder-decoder architecture developed by Fairseq (Facebook AI Research Sequence-to-Sequence Toolkit)[5] (Figure 1). There are two types of BART models that have been published:

- BART-base: 6 encoder layers and 6 decoder layers; 12 attention heads; word embedding dimension: 768; input sequence length: 512; 140 million parameters

- BART-large: 12 encoder layers and 12 decoder layers; 16 attention heads; word embedding dimension: 1024; input sequence length: 1024; 400 million parameters.
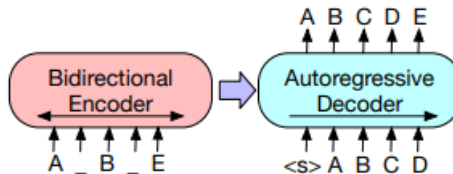

Figure 1. BART model architecture [13].

4.2. The multilingual mBART model

mBART [16] is a multilingual BART model trained by applying the BART model training algorithm to a large-scale monolingual corpus covering many languages. For pre-training, the first version mBART, the CC25 corpus [23] was used, which covers data in 25 languages extracted from Common Crawl. Later mBART was extended to mBART-50 covering 50 languages [22].

---

[5]https://github.com/pytorch/fairseq/tree/master/examples/bart

4.3. The mT5 model

mT5 [24] is an extended version of the T5 model (Text-To-Text Transfer Transformer) [20], which converts all text-based language problems (also ones originally formulated as classification problems) into a text-to-text format (Figure 2), and uses these "translation" tasks as a multitask training regime to create a unified generative language model. The T5 model allows knowledge transfer from high-resource tasks to low-resource tasks without the need for changes in model architecture. Unlike contextual language models such as BERT [3], which contain only the encoder part of a transformer, the T5 model is based on a full encoder-decoder architecture that can be used both for natural language understanding and language generation tasks. The mT5 model is a multilingual variant of T5 that was pre-trained on the Multilingual Colossal Clean Crawled Corpus (mC4) which covers 101 languages including Arabic.
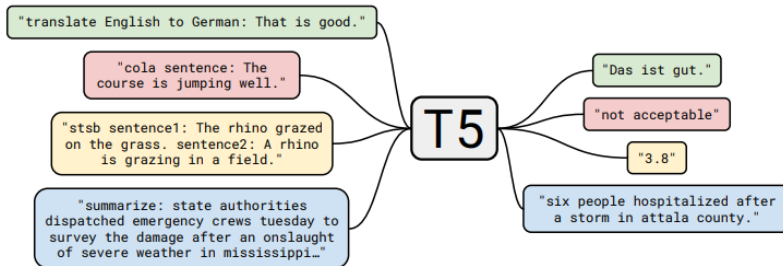

Figure 2. T5 model architecture [20].

5. Experiments

5.1. Training

We trained a monolingual and a trilingual (English, Hungarian, Arabic) BART base model. Unfortunately, Facebook did not publish the pre-training implementation, so we used the pre-training functions provided by the Huggingface transformers[6] libraries. The BartForCausalLM[7] model was used to train the BART models we present here.

For the monolingual Arabic BART model, we used content from the Arabic version of Wikipedia. We used about 250,000 paragraphs for training. Hyperparameters for the training were the following: vocab: 30000, batch size: 6/GPU on 8 RTX/GTX 11 GB GPU's, lr: 5e-6, warmup step: 500. We used the checkpoint at step 42000 (epoch 2.7) for further fine-tuning.

---

[6] https://huggingface.co/transformers/model_doc/bart.html
[7] https://huggingface.co/transformers/model_doc/bart.html#bartforcausallm

Next, we trained a trilingual model using a similar amount of additional Wikipedia content in English and Hungarian. We used the same hyperparameters, except the vocabulary size: 50000.

- Arabic BART: Monolingual Arabic BART base model, trained on 244,885 paragraphs of Arabic Wikipedia text.

- Arabic 3BART: a trilingual BART base model, trained on Arabic, English, and Hungarian Wikipedia content, about 250,000 paragraphs for each language.

Table 1 shows the properties of the corpus used for pre-training.

Table 1. Properties of the corpus used for pre-training the Arabic BART and 3BART model.

	Arabic	**English**	**Hungarian**
Segments	244,885	250,000	250,000
Tokens	10,391,179	34,098,745	13,838,277
Token types	415,628	365,998	1,018,315
Avg. sent. #	3.78	5.08	2.91
Avg. token #	42.43	136.39	55.35

5.2. Fine-tuning

In the fine-tuning experiments, we trained three models:

- **Arabic BART**: Arabic BART fine-tuned on the AraSum corpus.

- **Arabic 3BART**: Following the cross-lingual approach we used in our previous research[10], the 3BART model was first fine-tuned on a multilingual summarization corpus containing a mixture of English and Hungarian segments, and then further fine-tuned on the AraSum corpus. The English segments were taken from the CNN / Daily Mail corpus [18], while the Hungarian segments were taken from the H+I corpus [25]. Hyperparameters: batch: 4/GPU, 8 GTX/RTX 11 GB GPU's, warmup: 5000, 80 epochs, max. source: 512, max. target: 256, lr: 5e-5.

- **mT5**: We fine-tuned the mT5-small model for abstractive Arabic summarization using the AraSum corpus only. The hyperparameters were: prefix = sum, batch: 2/GPU, 8 GTX/RTX 11 GB GPU's, lr: 2e-5, warmup: 5000, 80 epochs, max. source: 512, max. target: 128.

Table 2 shows properties of the corpora used for fine-tuning the models.

Table 2. Properties of the corpora used for fine-tuning the BART
and 3BART models.

	Arabic (train / test)		English		Hungarian	
	Article	Lead	Article	Lead	Article	Lead
Segments	45,504 / 4,026		45,000		45,000	
Tokens	19,328,851 / 1,701,039	1,633,170 / 144,424	35,502,390	2,371,380	12,052,818	1,350,827
Types	466,387 / 129,987	111,689 / 29,792	253,113	83,672	656,060	166,092
Avg. sent. #	15.82 / 15.70	1.51 / 1.54	28.69	1	11.28	1,55
Avg. token #	424.77 / 422.51	35.89 / 35.87	788.94	52,69	267.84	30.01

6. Results

We evaluated system outputs using stemmed ROUGE-N and ROUGE-L metrics.[8] ROUGE-1 and ROUGE-2 measure the overlap of word unigrams and bigrams, respectively. ROUGE-L measures the overlap of the longest common sub-sequence between two texts. Stemmed ROUGE scoring is silently used in most recent publications on summarization, because it yields much nicer numbers than unstemmed ROUGE, especially for morphologically rich languages. While it may account better for content overlap, it ignores affixation disfluencies. It was used e.g. for evaluating the XL-Sum model [9], a multilingual summarization model fine-tuned from mT5 using the multilingual XL-Sum corpus for abstractive text summarization consisting of content crawled from the BBC news site. Training data for the XL-Sum model included about the same amount of Arabic data as in the current version of our corpus. XL-Sum can be considered a state-of-the-art model for abstractive Arabic text summarization so far.

Table 3 illustrates the experimental results of the different models trained or fine-tuned using the corpus, compared to the performance of the XL-Sum model. The evaluated models are:

- XL-Sum: tested on our test corpus.

- mBART-50: mBART-50 fine-tuned for Arabic summarization using the previous version of our corpus (about half the size of the current version).

- mBART-50-rus: mBART-50 first fine-tuned for Russian using the Gazeta corpus [6], then further fine-tuned for Arabic using the previous version of our corpus.

- PreSumm: mBERT first fine-tuned for English using the CNN/Daily Mail corpus[21], then further fine-tuned for Arabic using the previous version of our corpus.

- Arabic BART: the monolingual BART model pre-trained for Arabic as described in Section 5.1, then fine-tuned using the AraSum corpus.

---

[8]https://github.com/csebuetnlp/xl-sum/tree/master/multilingual_rouge_scoring

- Arabic 3BART: the trilingual BART model pre-trained for Arabic, English and Hungarian then fine-tuned using English and Hungarian the AraSum corpus as described in Section 5.1 .

- mT5: mT5-small model fine-tuned using the AraSum corpus.

- mT5++: the previous mT5-small model further fine-tuned on the union of the AraSum and XL-Sum Arabic training sets.

Table 3. ROUGE scores on the AraSum (top) and the XL-Sum Arabic (bottom) test sets.

Model	ROUGE-1	ROUGE-2	ROUGE-L
AraSum test set			
XL-Sum	30.026	12.874	23.836
mBART-50	32.648	14.617	24.878
mBART-50-rus	33.842	16.049	26.531
PreSumm	27.142	9.049	19.681
Arabic BART	27.019	7.657	18.960
Arabic 3BART	27.105	7.735	19.089
mT5	**32.859**	**13.843**	**24.571**
mT5++	**33.172**	**13.914**	**24.782**
XL-Sum Arabic test set			
XL-Sum	34.911	14.794	29.162
mBART-50	23.079	6.115	16.397
mBART-50-rus	23.777	4.589	15.114
PreSumm	18.880	4.389	13.553
Arabic BART	21.148	4.666	15.371
Arabic 3BART	20.892	4.589	15.114
mT5	22.120	5.570	15.908
mT5++	29.128	11.049	24.070

The models based on our homemade BART and 3BART pre-training yielded the weakest results. This is not surprising, as our computational resources (we used NVIDIA-GTX/RTX cards with 11GB memory) were too limited to be able to create competitive language models from scratch. Fine-tuning the mT5 small model on the same resources using the same hardware, however, resulted in a model that performs better in our home field than the SOTA multilingual XL-Sum model, which is based on a much stronger mT5 base model and was trained on much more data. Unfortunately, we did not have the chance to beat the XL-Sum model at home, as can be seen in the bottom half of Table 3. where XL-Sum is a multilingual summarization model trained on the whole multilingual XL-Sum dataset crawled from BBC. mT5++ is the mT5 model further finetuned on the union of the AraSum and the Arabic part of the XL-Sum dataset. The mBART-based models and PerSumm were fine-tuned on an earlier version of AraSum

Unfortunately, our limited hardware did not allow us to improve results by further finetuning the XL-Sum model on our corpus. However, we managed to improve our results by further finetuning our mT5 small model for 60 epochs on the union of the XL-Sum Arabic and the AraSum training set. The results of this mT5++ model are better on both test corpora.

7. Conclusion

In this paper, we present the extended version of the first monolingual human written corpus in the Arabic language for abstractive text summarization "AraSum". The corpus contains more than 50K Arabic articles and their corresponding leads. We pre-trained and fine-tuned a monolingual and a trilingual BART model for Arabic and one of today's most popular multilingual models, mT5.

With the resources and infrastructure we used, the results showed that the models trained on AraSum perform well, even surpassing the state-of-the-art XL-Sum model on the test set of our corpus. We release the corpus "AraSum" at our GitHub[9] in the hope that this will foster future work on abstractive Arabic text summarization.

References

[1] L. AL QASSEM, D. WANG, H. BARADA, A. AL-RUBAIE, N. ALMOOSA: *Automatic Arabic Text Summarization Based on Fuzzy Logic*, in: Proceedings of the 3rd International Conference on Natural Language and Speech Processing, 2019, pp. 42–48.

[2] A. M. AZMI, N. I. ALTMAMI: *An abstractive Arabic text summarizer with user controlled granularity*, Information Processing and Management 54.6 (2018), pp. 903–921, ISSN: 0306-4573, DOI: https://doi.org/10.1016/j.ipm.2018.06.002, URL: https://www.sciencedirect.com/science/article/pii/S030645731730417X.

[3] J. DEVLIN, M.-W. CHANG, K. LEE, K. TOUTANOVA: *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186, DOI: https://doi.org/10.18653/v1/N19-1423, URL: https://aclanthology.org/N19-1423.

[4] F. S. DOUZIDIA, G. LAPALME: *Lakhas, an Arabic summarization system*, Proceedings of DUC2004 (2004).

[5] K. N. ELMADANI, M. ELGEZOULI, A. SHOWK: *BERT Fine-tuning For Arabic Text Summarization*, ArXiv abs/2004.14135 (2020).

[6] I. GUSEV: *Dataset for Automatic Summarization of Russian News*, AINL 2020. Communications in Computer and Information Science, vol 1292. Springer, Cham (2020) (2020), DOI: https://doi.org/10.1007/978-3-030-59082-6_9, eprint: arXiv:2006.11063.

[7] M. EL-HAJ, R. KOULALI: *KALIMAT a multipurpose Arabic corpus*, in: Second Workshop on Arabic Corpus Linguistics (WACL-2), 2013, pp. 22–25.

---

[9]https://github.com/ppke-nlpg/AraSum

[8] M. EL-HAJ, U. KRUSCHWITZ, C. FOX: *Using Mechanical Turk to Create a Corpus of Arabic Summaries*, in: Language Resources (LRs) and Human Language Technologies (HLT) for Semitic Languages workshop in conjunction with the 7th International Language Resources and Evaluation Conference (LREC 2010), Jan. 2010.

[9] T. HASAN, A. BHATTACHARJEE, M. S. ISLAM, K. MUBASSIR, Y.-F. LI, Y.-B. KANG, M. S. RAHMAN, R. SHAHRIYAR: *XL-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages*, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online: Association for Computational Linguistics, Aug. 2021, pp. 4693–4703, DOI: https://doi.org/10.18653/v1/2021.findings-acl.413, URL: https://aclanthology.org/2021.findings-acl.413.

[10] M. KAHLA, Z. G. YANG, A. NOVÁK: *Cross-lingual Fine-tuning for Abstractive Arabic Text Summarization*, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Held Online: INCOMA Ltd., Sept. 2021, pp. 655–663, URL: https://aclanthology.org/2021.ranlp-main.74.

[11] F. LADHAK, E. DURMUS, C. CARDIE, K. MCKEOWN: *WikiLingua: A New Benchmark Dataset for Cross-Lingual Abstractive Summarization*, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Online: Association for Computational Linguistics, Nov. 2020, pp. 4034–4048, DOI: https://doi.org/10.18653/v1/2020.findings-emnlp.360, URL: https://aclanthology.org/2020.findings-emnlp.360.

[12] S. M. LAKHDAR, M. A. CHÉRAGUI: *Building an Extractive Arabic Text Summarization Using a Hybrid Approach*, in: International Conference on Arabic Language Processing, Springer, 2019, pp. 135–148.

[13] M. LEWIS, Y. LIU, N. GOYAL, M. GHAZVININEJAD, A. MOHAMED, O. LEVY, V. STOYANOV, L. ZETTLEMOYER: *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension*, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online: Association for Computational Linguistics, July 2020, pp. 7871–7880, DOI: https://doi.org/10.18653/v1/2020.acl-main.703, URL: https://aclanthology.org/2020.acl-main.703.

[14] C.-Y. LIN: *ROUGE: A Package for Automatic Evaluation of Summaries*, in: Text Summarization Branches Out, Barcelona, Spain: Association for Computational Linguistics, July 2004, pp. 74–81, URL: https://www.aclweb.org/anthology/W04-1013.

[15] Y. LIU, M. LAPATA: *Text Summarization with Pretrained Encoders*, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3730–3740, DOI: https://doi.org/10.18653/v1/D19-1387, URL: https://aclanthology.org/D19-1387.

[16] Y. LIU, J. GU, N. GOYAL, X. LI, S. EDUNOV, M. GHAZVININEJAD, M. LEWIS, L. ZETTLEMOYER: *Multilingual Denoising Pre-training for Neural Machine Translation*, Transactions of the Association for Computational Linguistics 8 (2020), pp. 726–742, DOI: https://doi.org/10.1162/tacl_a_00343, URL: https://aclanthology.org/2020.tacl-1.47.

[17] M. AL-MALEH, S. DESOUKI: *Arabic text summarization using deep learning approach*, Journal of Big Data 7 (2020), pp. 1–17.

[18] R. NALLAPATI, B. ZHOU, M. MA: *Classify or select: Neural architectures for extractive document summarization*, arXiv preprint arXiv:1611.04244 (2016).

[19] R. PAULUS, C. XIONG, R. SOCHER: *A Deep Reinforced Model for Abstractive Summarization*, CoRR abs/1705.04304 (2017), arXiv: 1705.04304, URL: http://arxiv.org/abs/1705.04304.

[20] C. RAFFEL, N. SHAZEER, A. ROBERTS, K. LEE, S. NARANG, M. MATENA, Y. ZHOU, W. LI, P. J. LIU: *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, Journal of Machine Learning Research 21.140 (2020), pp. 1–67, URL: http://jmlr.org/papers/v21/20-074.html.

[21] A. SEE, P. J. LIU, C. D. MANNING: *Get To The Point: Summarization with Pointer-Generator Networks*, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada: Association for Computational Linguistics, July 2017, pp. 1073–1083, DOI: https://doi.org/10.18653/v1/P17-1099, URL: https://www.aclweb.org/anthology/P17-1099.

[22] Y. TANG, C. TRAN, X. LI, P.-J. CHEN, N. GOYAL, V. CHAUDHARY, J. GU, A. FAN: *Multilingual Translation with Extensible Multilingual Pretraining and Finetuning*, 2020, arXiv: 2008.00401 [cs.CL].

[23] G. WENZEK, M.-A. LACHAUX, A. CONNEAU, V. CHAUDHARY, F. GUZMÁN, A. JOULIN, E. GRAVE: *CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data*, in: Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France: European Language Resources Association, May 2020, pp. 4003–4012, URL: https://aclanthology.org/2020.lrec-1.494.

[24] L. XUE, N. CONSTANT, A. ROBERTS, M. KALE, R. AL-RFOU, A. SIDDHANT, A. BARUA, C. RAFFEL: *mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer*, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online: Association for Computational Linguistics, June 2021, pp. 483–498, DOI: https://doi.org/10.18653/v1/2021.naacl-main.41, URL: https://aclanthology.org/2021.naacl-main.41.

[25] Z. G. YANG, Á. AGÓCS, G. KUSPER, T. VÁRADI: *Abstractive text summarization for Hungarian*, Annales Mathematicae et Informaticae 53 (2021), pp. 299–316.