Proceedings of the OSACT 2022 Workshop @LREC2022, pages 120–125
Marseille, 20 June 2022

LK2022 at Qur'an QA 2022: Simple Transformers Model for Finding Answers to Questions from Qur'an

Abdullah Alsaleh[1][2], Saud Althabiti[1][2], Ibtisam Alshammari[1][3], Sarah Alnefaie[1][2], Sanaa Alowaidi[1][2], Alaa Alsaqer[1][4], Eric Atwell[1], Abdulrahman Altahhan[1], Mohammad Ammar Alsalka[1]
[1]University of Leeds, [2]King Abdulaziz University, [3]University of Hafr Al-Batin, [4]King Faisal University
{scanaa, scssal, ml18ikfa, scsaln, ml20sara, scafa, e.s.atwell, a.altahhan, m.a.alsalka}@leeds.ac.uk

Abstract

Question answering is a specialized area in the field of NLP that aims to extract the answer to a user question from a given text. Most studies in this area focus on the English language, while other languages, such as Arabic, are still in their early stage. Recently, research tend to develop question answering systems for Arabic Islamic texts, which may impose challenges due to Classical Arabic. In this paper, we use Simple Transformers Question Answering model with three Arabic pre-trained language models (AraBERT, CAMeL-BERT, ArabicBERT) for Qur'an Question Answering task using Qur'anic Reading Comprehension Dataset. The model is set to return five answers ranking from the best to worst based on their probability scores according to the task details. Our experiments with development set shows that AraBERT V0.2 model outperformed the other Arabic pre-trained models. Therefore, AraBERT V0.2 was chosen for the the test set and it performed fair results with 0.45 pRR score, 0.16 EM score and 0.42 F1 score.

Keywords: NLP, Simple-Transformers, AraBERT, Question-Answering, Quran

1. Introduction

Natural Language Processing (NLP) is widely used for English language tasks; however, in case of Arabic language, it is still a challenging task especially for The Holy Qu'ran as it is considered a Classical Arabic and the Quranic terms have distinct meanings that differ from all Arabic variants, which make it more challenging for researchers (Altammami et al., 2020). Recent studies concentrate on Arabic language tasks such as Quranic Question Answering Systems, which plays significant role on NLP generally and Arabic language processing field specially. One of these recent studies is the 2022 Qur'an Question Answering shared task, where researchers can participate in teams to develop solutions to improve Question Answering Systems in terms of partial Reciprocal Rank (pRR) score (Malhas and Elsayed, 2020) (Malhas et al., 2022).

Recently, one of the most emerging NLP techniques is applying transformers-based models on Question Answering systems, which entails retrieving the required information from particular text according to a specific query or question (Rajpurkar et al., 2016) (Yang et al., 2015) (Mahdi, 2021). Therefore, we aim in this paper to utilize one of the transformers-based model, which is Simple Transformers model, on the Shared Task Question Answering over the Holy Qur'an. It mainly focuses on adapting Simple Transformers model to obtain the needed information from Qur'an passages and improve the accuracy of results using three Arabic pre-trained language models that are based on Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018).

The paper structure as follows: In section 2, the related work are discussed. In section 3, the Qur'anic Reading Comprehension Dataset (QRCD) is presented. Section 4 describes the methodology that includes Simple Transformers model, dataset formation and Arabic pre-trained language models. Followed by section 5 as the results gained over the experiments given and supported by the discussion and analysis in section 6.

2. Related Work

2.1. Quranic Questions and Answers Systems

Many studies have been conducted to retrieve answers from the Holy Quran for the user's questions using information retrieval techniques such as semantic similarity and pattern matching. Abdelnasser et al. (2014) developed Al-Bayan system to answer Arabic Quranic questions. Al-Bayan represents verses by concept vectors and then uses cosine semantic similarity to retrieve the related verses for the user query. The system achieved 65% in terms of accuracy.

Moreover, AbuShawar and Atwell (2016) used the ALICE platform to implement a simple Quranic chatbot model based on the pattern matching technique. The model allows the user to ask questions in English and answer them in Arabic and English Quran verses. The experiment demonstrated that 54% of the results were wrong Answers.

In addition, Alqahtani (2019) built a model to answer the Arabic Quranic questions. This model enriched the user query with semantic features using the word2Vec algorithm. Then it extracted the most related concepts to the query from Quranic ontology using the cosine similarity. After that, it displayed the verses of the matched concepts to the user. The evaluation results

showed 41% in terms of recall.
These studies have contributed significantly to enriching the field of Quranic research. However, they can extract the required answer from the verse for only specific types of questions or answer the questions with the whole verse.

2.2. Arabic Questions and Answers Systems

The research on Arabic Question Answering systems has recently tended to apply deep learning transformers models such as the BERT model. According to Devlin et al. (2018) , it proved its effectiveness in several NLP tasks. Mozannar et al. (2019) trained the BERT model using an Arabic dataset. They constructed a specialized Arabic Reading Comprehension Dataset (ARCD) for the question and answering task consisting of 1,395 questions from Wikipedia. In addition, they created an Arabic version of the SQuAD 1.1 QA dataset by translating about 2,966 question pairs. The resulting datasets were used to train the BERT model. The experimental results achieved 61.3% in terms of F1 score. Additionally, Antoun et al. (2020) created an Arabic language model based on BERT named AraBERT by pre-trained the model using an extensive Arabic dataset. The dataset includes about 70 million sentences from available Arabic corpora and news websites. In addition, they tested the AraBERTv0.1 in question answering application using Arabic-SQuAD and ARCD datasets. This model showed better performance than the multilingual BERT (mBERT) by 1.4% and 3.0% improvement in F1 score and sentence match. The proposed model is available online for public use. Furthermore, Alsubhi et al. (2021) compared the performance of two transformer models, AraBERTv2-base and AraBERTv0.2-large. These models are trained on four Arabic QA datasets Arabic-SQuAD, ARCD, Arabic TyDiQA-GoldP, and AQAD, that generally are extracted from Wikipedia articles. The results showed that general AraBERTv0.2-large outperforms the other models, and the best results were achieved using the Arabic TyDiQA-GoldP dataset with 86.49% F1 score and 75.14% exact matches.
The current Arabic research train BERT models to answer questions in different domains. As far as we know, no conducted study train BERT models to the Quranic questions and answering systems.

3. Data

This section provides an overview of the dataset used in this paper. Quran QA shared task dataset is called QRCD [1] (Malhas and Elsayed, 2020), an Arabic Question Answering dataset. For each record, it includes a passage in plain text style that is derived from the Tanzil project [2], a question that is presented in Modern


```
{ "pq_id": "74:32-48_330",
"passage": "كلا والقمر. والليل إذ أدبر. والصبح إذا أسفر. إنها لإحدى الكبر."
نذيرا للبشر. لمن شاء منكم أن يتقدم أو يتأخر. كل نفس بما كسبت رهينة.
إلا أصحاب اليمين. في جنات يتساءلون. عن المجرمين. ما سلككم في سقر. قالوا لم نك من المصلين.
ولم نك نطعم المسكين. وكنا نخوض مع الخائضين. وكنا نكذب بيوم الدين.
حتى أتانا اليقين. فما تنفعهم شفاعة الشافعين.",
"surah": 74, "verses": "32-48",
"question": "ما هي الدلائل التي تشير بأن الانسان مخير؟",
"answers":
[ {
"text": "لمن شاء منكم أن يتقدم أو يتأخر", "start_char": 76 }
{ "text": "كل نفس بما كسبت رهينة", "start_char": 108 }
] }
```


Figure 1: QRCD Structure

Standard Arabic (MSA), and one or more answers that are extracted from the given passage. It also includes PQ_ID, Surah number and verse numbers of the given passage. Ultimately, the structure of the QRCD is a JSON Line, as shown in Figure 1.
Figure 2 illustrates the distribution of the provided dataset. As it can be seen, the dataset contains 1093 question-passage pairs with their extracted answers. The training and development (validation) sets divided as 710, and 109 respectively. Furthermore, the test set includes 274 pairs of questions and passages without answers. However, through our experiments, we noticed that 99 questions in the training set and 15 in the development set have more than one answer. For example, this question (من هو قارون؟) meaning (Who is Qarun?) in the training set, has five answers from the Sura Al-Qasas (سورة القصص) verses (76-81 أيات).
The same question was mentioned in other IDs and they have different answers from other passages.

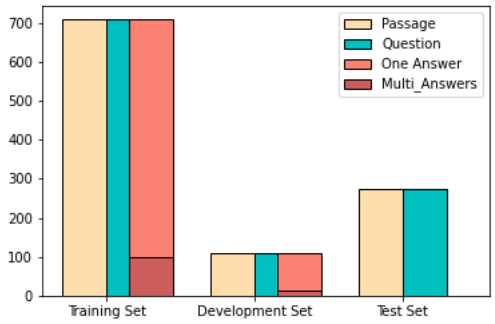

Figure 2: Distribution of QRCD

4. Method

This section outlines the methodology that has been utilised for this shared task. The model that was used

[1]The dataset can be accessed via this GitLab page: https://gitlab.com/bigirqu/quranqa/-/tree/main/datasets

[2]Tnazil Project contains a number of text styles including the plain text: https://tanzil.net/download/

for this task is Simple Transformers Question Answering model. The Question Answering model requires specific format that will be entailed in this section. Since the model is compatible with BERT, we applied three Arabic pre-trained language models with their weights and different sizes (base or large). Finally, the experiments were run on Google Colab with cuda for faster processing.

4.1. Simple Transformers

Simple Transformers model is a library that is built on Transformers architecture to solve downstream tasks such as binary or multi-class text classification. The library has since been developed to include question answering model, named-entity recognition, and language generations. The Question Answering model can be trained, evaluated and tested using different parameters that suit specific tasks and may improve performances. During training, the parameters that we focused on for this task are: batch size, learning rate and number of epochs. As for the prediction, we set the model to return five answers for each question. Finally, the output of the model is two lists that contain the answers and their probability scores.

4.2. Dataset Formation

Simple Transformers model requires specific data format prior feeding it to the model. The model expects a dictionary with two attributes `context` and `qas`, where "context" in this case is the verse. As for the "qas", it is a list that contains the ID, question and its answers. So, we have modified the existing scripts, given by the organisers, to convert the current QRCD format to the structure that Simple Transformers requires. The end format is shown in figure 3. Finally, there has been no pre-processing or pre-treatment on the dataset.

4.3. Arabic Pre-trained Language Models

There are three Arabic pre-trained language models that have been implemented in our experiments. They

```
{'context': ' كلا والقمر. والليل إذ أدبر. والصبح إذا أسفر. إنها لإحدى الكبر. نذيرا للبشر. لمن
شاء منكم أن يتقدم أو يتأخر. كل نفس بما كسبت رهينة. إلا أصحاب اليمين. في جنات يتساءلون. عن
المجرمين. ما سلككم في سقر. قالوا لم نك من المصلين. ولم نك نطعم المسكين. وكنا نخوض مع
الخائضين. وكنا نكذب بيوم الدين. حتى أتانا اليقين. فما تنفعهم شفاعة الشافعين.',
'qas': [
        {'id': '74:32-48_330',
        'surah': 74,
        'verses': '32-48',
        'question': 'ما هي الدلائل التي تشير بأن الانسان مخير؟',|
        'answers': [
                {'text': 'لمن شاء منكم أن يتقدم أو يتأخر',
                'answer_start': 76},
                {'text': 'كل نفس بما كسبت رهينة',
                'answer_start': 108}
                ]
        }
]
}
```

Figure 3: Dataset Formation

are AraBERT (Antoun et al., 2020), CAMeL-BERT (Inoue et al., 2021) and ArabicBERT (Safaya et al., 2020).

4.3.1. AraBERT

AraBERT is a BERT based language model with pre-trained corpus that includes 1.5 billion words from Arabic corpora (El-Khair, 2016) and Open-Sourced International Arabic News Corpus (Zeroual et al., 2019). AraBERT has two models which are AraBERT V2 and AraBERT V0.2 and the only difference is the use of Farasa Segmenter on V2. So for the model, AraBERT V0.2 was chosen since it performed better on recent Quran semantic similarity research (Alsaleh et al., 2021). Also, AraBERT provides base and large variants, and we opted with the latter since it performed better in the initial experiments on the development set.

4.3.2. CAMeL-BERT

CAMeL-BERT is a deep learning Arabic language model that is based on BERT architecture. The model provides more than 8 variants that are specific for Classical Arabic, Modern Standard Arabic and dialects. The pre-trained corpus for Classical Arabic is OpenITI (Nigst et al., 2020), which is an Arabic corpus that pertain to pre-modern Islamic texts. For our experiments, we have opted for the Classical Arabic base variant.

4.3.3. ArabicBERT

ArabicBERT is an Arabic language model that is based on BERT architecture with pre-trained corpus that includes Open Super-large Crawled Aalanch coRpus (OSCAR) (Ortiz Suárez et al., 2020), which includes Modern Standard Arabic, dialect texts and latest Arabic Wikipedia dump. The model provides different sizes including base, large and mega. We opted for the large model since we could not run our experiments with mega due to hardware limitations.

5. Results

5.1. Validation

In this competition, we conducted various experiments using the development set on multiple transformer-based models, namely ArabicBERT, CAMeL-BERT, and AraBERT. On each model, we further investigated different versions of these models large or based. In addition, we fine-tuned our models on three parameters batch size, learning rate, and number of epochs. We chose these hyper parameters to minimise losses, avoid overfitting, and try to reach the local optima. After training our models over 25 epochs, we concluded that five to seven epochs are sufficient and could provide promising results. The obtained results from these various experiments with the different selected hyperparameters indicate that the AraBERT model usually outperforms other transformers-based models and could provide promising results, as illustrated in Figure 4.
Table 1 shows the highest scores acquired from each model. Within each model, we demonstrated the

Model	Batch Size	Learning Rate	Epochs	pRR	EM	F1@1
CAMeL-BERT	50	1e-4	15	0.53	0.31	0.49
	10	2e-5	20	0.52	0.30	0.47
	25	1e-4	15	0.51	0.28	0.47
AraBERT V0.2	**15**	**1e-4**	**5**	**0.59**	**0.34**	**0.55**
	15	1e-4	5	0.56	0.36	0.53
	15	4e-5	5	0.55	0.31	0.52
ArabicBERT	20	2e-4	30	0.51	0.33	0.47
	15	1e-4	5	0.49	0.30	0.46
	20	1e-4	20	0.48	0.28	0.43

Table 1: Summary of development set results, which includes the models with their arguments and evaluation scores

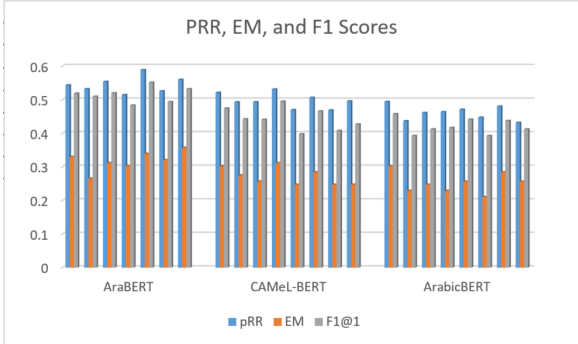

Figure 4: Overview of the experiments

used values of the three parameters manipulated in the employed transformer-based models (CAMeL-BERT, AraBERT, and ArabicBERT) to obtain the highest pRR scores (0.53, 0.59, and 0.48), respectively. We also attempt to adopt the best combination of parameters used with AraBERT (batch size of 15, learning rate of 1e-4, and five epochs) to the other two models. However, the comparison still indicates that this combination provides higher scores with AraBERT.

5.2. Testing

Accordingly, we employed these evaluated parameters in our final model's performance on the test dataset, and we got a fair result, with a 0.45 pRR score, 0.16 EM score and a 0.42 F1 score compared to the average scores of all participated teams with a 0.41 pRR score, 0.12 EM score and a 0.37 F1 score.

6. Discussion and Analysis

This section will discuss and analyse the development set as the true answers were not provided in the test set when publishing this paper.

The best result of the development set were using AraBERT V0.2 language model with parameters shown in table 1. When we analysed the results, we found that the model did not always return 5 answers. Also, there were 9 empty answers for the following IDs (9:1-6_400, 7:19-25_257, 22:30-37_313, 29:61-69_313, 20:95-98_163, 39:11-20_373, 31:12-19_132, 4:12-14_415, 33:36-40_415). The reason is that the model could not work out an answer for these questions based on given passage [3]. To avoid the warnings set in the official evaluation script on empty answers, we created a function to remove any empty answers except if the empty answer is the only answer that the model predicted.

According to the development set results, the model can predict the answer when there are matched words and/or synonyms between the questions and the passage; otherwise, it may face some difficulties. For example, in figure 5 the correct answer for the question "من الذي صنع عجلا من الحلي لبني إسرائيل؟" "Who made a calf out of jewelry for Israelites?" is "Samaritan" (السامري). In the first passage, there is a matching word "calf" (عجلا) , and "ornaments" (زينة) which is an Arabic synonyms for the word (الحلي) in the question. Therefore, the model successfully answered it when retrieving the answer from the passage.

In contrast, in the second passage in figure 6, the model could not predict the correct answer and produced empty answer. We notice that there are no matched words, although there are phrases that have related meanings, such as "إلهك الذي ظلت عليه عاكفا لنحرقنه" meaning (your 'god' to which you remained devoted. We will surely burn it), which points to the "calf", and "فقبضت قبضة من أثر الرسول فنبذتها" meaning (so I took a handful [of dust] from the track of the messenger and threw it), which is referred to how the "calf" was built according to Ibn-Kathir explication.

Moreover, the model predicts correct answers for some questions, but it was not mentioned in the gold answers. For example, in question 2:190-194_400 "متى يحل الإسلام دم الشخص؟" meaning "When does Islam allow the blood of a person?", the gold answer is "قاتلوا في سبيل الله الذين يقاتلونكم" which translates "Fight in the way of Allah those who fight you", while the model predicted this answer

[3]Please refer to https://simpletransformers.ai/docs/qa-specifics/

"فمن اعتدى عليكم فاعتدوا عليه بمثل ما اعتدى عليكم" which translates to "So whoever has assaulted you, then assault him in the same way that he has assaulted you". According to the scholar Al-Tabari (1954), the interpretation of the predicted answer "فقاتلوهم فيه كما قاتلوكم" meaning "fight them in it as they fought you" which has a similar meaning to the gold answer. Therefore, there maybe other correct answers that could potentially be added as gold answers in the datasets.

7. Conclusion

This paper presented Simple Transformers model to retrieve the best answer of particular questions related to the Holy Qur'an Shared Task competition. The experiments have been conducted using three Arabic language models AraBERT, CAMeL-BERT, and ArabicBERT. As the results shown that the AraBERT V0.2 model outperforms the other transformers-based models in this task with a 0.59 pRR score, 0.34 EM score and a 0.55 F1 score for the development set. As a consequence, in the test set shown fair results with 0.45 pRR score, 0.16 EM score and 0.42 F1 score.
Moreover, our findings shown that our developed model not only retrieves matching words as correct answers, but also predicts other additional answers that could be considered as accurate answers and potentially be added as gold answers to the datasets.

PQ_ID: 20:86-94_163

Passage:
فرجع موسى إلى قومه غضبان أسفا قال يا قوم ألم يعدكم ربكم وعدا حسنا أفطال عليكم العهد أم أردتم أن يحل عليكم غضب من ربكم فأخلفتم موعدي. قالوا ما أخلفنا موعدك بملكنا ولكنا حملنا أوزارا من زينة القوم فقذفناها فكذلك ألقى السامري. فأخرج لهم عجلا جسدا له خوار فقالوا هذا إلهكم وإله موسى فنسي. أفلا يرون ألا يرجع إليهم قولا ولا يملك لهم ضرا ولا نفعا. ولقد قال لهم هارون من قبل يا قوم إنما فتنتم به وإن ربكم الرحمن فاتبعوني وأطيعوا أمري. قالوا لن نبرح عليه عاكفين حتى يرجع إلينا موسى. قال يا هارون ما منعك إذ رأيتهم ضلوا. ألا تتبعن أفعصيت أمري. قال يا ابن أم لا تأخذ بلحيتي ولا برأسي إني خشيت أن تقول فرقت بين بني إسرائيل ولم ترقب قولي.

Question: من الذي صنع عجلا من الحلي لبني إسرائيل؟

Gold Answer:
1- السامري

Predicted Answer:
1- السامري. فأخرج لهم عجلا جسدا له خوار فقالوا هذا إلهكم وإله موسى فنسي
2- السامري
3- السامري.
4- السامري. فأخرج لهم عجلا جسدا له خوار
5- فأخرج لهم عجلا جسدا له خوار فقالوا هذا إلهكم وإله موسى فنسي

Figure 5: Example of predicting correct answers when the question has same words or synonyms in the passage

PQ_ID: 20:95-98_163

Passage:
قال فما خطبك يا سامري. قال بصرت بما لم يبصروا به فقبضت قبضة من أثر الرسول فنبذتها وكذلك سولت لي نفسي. قال فاذهب فإن لك في الحياة أن تقول لا مساس وإن لك موعدا لن تخلفه وانظر إلى إلهك الذي ظلت عليه عاكفا لنحرقنه ثم لننسفنه في اليم نسفا. إنما إلهكم الله الذي لا إله إلا هو وسع كل شيء علما.

Question: من الذي صنع عجلا من الحلي لبني إسرائيل؟

Gold Answer:
1- سامري

Predicted Answer:
1- "empty"

Figure 6: Example of incorrect prediction when there is no matching words or phrases in the question and in the passage

8. Acknowledgements

Authors, would like to express their deepest gratitude to King Abdulaziz University, University of Hafr Al-Batin, and King Faisal University for the support. We thank the reviewers for their constructive comments.

9. Bibliographical References

Abdelnasser, H., Ragab, M., Mohamed, R., Mohamed, A., Farouk, B., El-Makky, N. M., and Torki, M. (2014). Al-bayan: an arabic question answering system for the holy quran. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, pages 57–64.

AbuShawar, B. and Atwell, E. (2016). Usefulness, localizability, humanness, and language-benefit: additional evaluation criteria for natural language dialogue systems. *International Journal of Speech Technology*, 19(2):373–383.

Al-Tabari, M. B. J. (1954). جامع البيان عن تاويل اي القرآن. Dar Al-Fikr.

Alqahtani, M. M. A. (2019). *Quranic Arabic Semantic Search Model Based on Ontology of Concepts*. Ph.D. thesis, University of Leeds.

Alsaleh, A., Atwell, E., and Altahhan, A. (2021). Quranic verses semantic relatedness using AraBERT. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 185–190, Kyiv, Ukraine (Virtual), April. Association for Computational Linguistics.

Alsubhi, K., Jamal, A., and Alhothali, A. (2021). Pretrained transformer-based approach for arabic question answering: A comparative study. *arXiv preprint arXiv:2111.05671*.

Altammami, S., Atwell, E., and Alsalka, A. (2020). Towards a joint ontology of quran and hadith. *International Journal on Islamic Applications in Computer Science And Technology*.

Antoun, W., Baly, F., and Hajj, H. (2020). AraBERT: Transformer-based model for Arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France, May. European Language Resource Association.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

El-Khair, I. A. (2016). 1.5 billion words arabic corpus. *ArXiv*, abs/1611.04033.

Inoue, G., Alhafni, B., Baimukan, N., Bouamor, H., and Habash, N. (2021). The interplay of variant, size, and task type in Arabic pre-trained language models. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, Kyiv, Ukraine (Online), April. Association for Computational Linguistics.

Mahdi, A. F. (2021). Survey: using bert model for arabic question answering system. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(13):723–729.

Malhas, R. and Elsayed, T. (2020). Ayatec: building a reusable verse-based test collection for arabic question answering on the holy qur'an. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(6):1–21.

Malhas, R., Mansour, W., and Elsayed, T. (2022). Qur'an QA 2022: Overview of the first shared task on question answering over the holy qur'an. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Mozannar, H., Hajal, K. E., Maamary, E., and Hajj, H. (2019). Neural arabic question answering. *arXiv preprint arXiv:1906.05394*.

Nigst, L., Romanov, M., Savant, S. B., Seydi, M., and Verkinderen, P. (2020). Openiti: a machine-readable corpus of islamicate texts, Jun.

Ortiz Suárez, P. J., Romary, L., and Sagot, B. (2020). A monolingual approach to contextualized word embeddings for mid-resource languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.

Safaya, A., Abdullatif, M., and Yuret, D. (2020). KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2054–2059, Barcelona (online), December. International Committee for Computational Linguistics.

Yang, Y., Yih, W.-t., and Meek, C. (2015). Wikiqa: A challenge dataset for open-domain question answering. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2013–2018.

Zeroual, I., Goldhahn, D., Eckart, T., and Lakhouaja, A. (2019). Osian: Open source international arabic news corpus - preparation and integration into the clarin-infrastructure. In *WANLP@ACL 2019*.