



Department of Computer Science
COMP4200-Introduction to Graduation Project
Spring 2022/2023

Title of Project: Q&Arabic - An Automatic Question & Answering FAQ Generator

Group Members with IDs:

Obada Tahayna - 1191319

Kareem Halayqa - 1192087

Mo'men Dar Bazzar - 1192214

Supervisor: Dr. Radi Jarrar

Section – B

Title of Project: Q&Arabic - An Automatic Question & Answering FAQ Generator

Project No: 016

Supervisor: Dr. Radi Jarrar

Key Areas: Question Generation, Question Answering, Deep Learning, NLP (Natural Language Processing), Arabic NLP, Transformer Models, T5, BERT.

Section – C

Student Signature: _____

Date Submitted: 20/02/2023

First Supervisor Name: Dr. Radi Jarrar

First Supervisor Signature: _____

Date Approved: ___ / ___ / ___

Abstract

Q&Arabic is a natural language processing (NLP)[1] framework that automatically creates Arabic frequently asked questions (FAQs) and answers from a given material. By utilizing NLP techniques to find and extract the most pertinent information from a document and present it in a clear and succinct question and answer format, the project seeks to accelerate the process of creating FAQs.

The system is designed to generate questions from a given text and then provide answers to those questions. The system can be used in various applications such as creating FAQs for websites, articles or creating exams from Arabic documents.

Users can submit a text or upload a document and the system will automatically generate related questions and answers. Users are then able to view and export the generated questions and answers. The user history is stored in the application for enabling users to view their previous submission.

Our approach leverages the latest advances in NLP and Machine Learning to generate questions and provide accurate answers in the Arabic language. In particular, we will make use of Deep Learning frameworks, such as transformer models, to model complex patterns in the Arabic text. In particular, we will use Arabic T5 models for question generation and Arabic BERT models for question answering. These models have shown great success in the literature in various NLP tasks.

Q&Arabic is a unique tool that is specifically designed to support the Arabic language. The Arabic FAQ Generator is crafted for the complexity of the Arabic language, unlike other FAQ generators that are primarily created for English or other languages. This indicates that the system can correctly and efficiently process Arabic text, recognize important details and concepts, and provide questions and answers that are precise and concise.

Table of Contents

| | |
|---|----|
| Chapter 1: Introduction | 2 |
| 1.1 Motivation | 2 |
| 1.2 Aims and Objectives | 2 |
| 1.2.1 Aims | 2 |
| 1.2.2 Objectives..... | 3 |
| 1.3 Overview of the technical area | 3 |
| 1.3.1 Artificial intelligence (AI)..... | 3 |
| 1.3.2 Machine Learning (ML)..... | 4 |
| 1.3.3 Natural language processing (NLP) | 5 |
| 1.3.4 Question Generation..... | 6 |
| 1.3.5 Question Answering..... | 6 |
| 1.4 Overview of the report | 7 |
| Chapter 2: Background & Literature Review | 8 |
| 2.1 Background | 8 |
| 2.1.1 Language Transformer Models | 8 |
| 2.1.2 Datasets | 12 |
| 2.2 Literature Review | 14 |
| 2.2.1 NLP | 14 |
| 2.2.2 Question Generation..... | 15 |
| 2.2.3 Question Answering..... | 16 |
| 2.2.4 Q/A Combined | 18 |
| 2.3 Related Works | 18 |
| 2.3.1 LongShot | 18 |
| 2.3.2 Quillionz..... | 18 |
| 2.3.3 Questgen | 19 |
| 2.3.4 PrepAI | 21 |
| 2.3.5 ChatGPT..... | 25 |
| Chapter 3: System Analysis | 28 |
| 3.1 System Description | 28 |
| 3.1.1 System Objectives | 28 |
| 3.1.2 System Main Features | 28 |
| 3.1.3 Operating Environments | 29 |
| 3.1.4 Constrains..... | 29 |
| 3.1.5 Functional Requirements..... | 30 |
| 3.1.6 Non-Functional Requirements | 31 |

| | |
|--|----|
| 3.2 System Design | 31 |
| 3.2.1 System Architecture | 31 |
| 3.2.2 Class Diagram | 33 |
| 3.2.3 Activity Diagrams | 33 |
| 3.2.4 Use Case Diagram..... | 35 |
| 3.2.5 Sequence Diagrams | 36 |
| 3.2.6 Deployment Diagram | 37 |
| 3.3 Interface Demo | 37 |
| 3.3.1 Registration Page..... | 37 |
| 3.3.2 Login Page..... | 38 |
| 3.3.3 Main Page (Text Input and Q/A generation)..... | 39 |
| 3.3.4 History Page | 41 |
| Chapter 4: Conclusion and Future works | 42 |
| 4.1 Review of the project | 42 |
| 4.2 Future Works | 42 |
| References | 43 |

List of Figures

| | |
|--|----|
| Figure 1: Transformer Model Architecture | 8 |
| Figure 2: BERT pre-training and fine-tuning procedures | 9 |
| Figure 3: BERT input representation - token, segmentation and position embeddings..... | 9 |
| Figure 4: GPT-1 Transformer architecture and fine-tuning process | 10 |
| Figure 5: InstructGPT's three-step training process | 10 |
| Figure 6: T5 Model Tasks | 11 |
| Figure 7: AraT5 Model Tasks | 12 |
| Figure 8: SQuAD format of Q&A datasets | 13 |
| Figure 9: System Overall Architecture..... | 31 |
| Figure 10: Question Generation and Answering Model | 32 |
| Figure 11: Fine-Tuning Models for QG and QA | 32 |
| Figure 12: Class Diagram..... | 33 |
| Figure 133: Register Activity Diagram..... | 33 |
| Figure 14: QA Generation Activity Diagram..... | 34 |
| Figure 15: View History Activity Diagram..... | 35 |
| Figure 16: Use Case Diagram | 35 |
| Figure 178: Register Sequence Diagram..... | 36 |
| Figure 189: Login Sequence Diagram | 36 |
| Figure 19: QA Generation Sequence Diagram | 36 |
| Figure 20: Deployment Diagram..... | 37 |
| Figure 21: Main Page (File Upload) Demo | 39 |
| Figure 22: Main Page (Text Input) Demo | 39 |
| Figure 23: Main Page (With Options Pop-up Demo | 40 |
| Figure 24: Main Page (With Q/A Generated) Demo | 40 |
| Figure 25: History Page Demo | 41 |

Acknowledgements

We would like to pay gratitude towards everyone that supported us to come up with and create this idea. First and foremost, we would like to thank our supervisor Dr. Radi Jarrar, who has been beyond supportive throughout the whole journey, providing guidance while still giving us creative freedom, in order to come up with a project that actually represents what we want.

We would also like to express our gratitude to our friends, families, and anyone who helped us, whether it was providing support for the idea or suggestions that helped curate the idea.

Chapter 1: Introduction

Automatic question generation and answering is a rapidly growing area of research that has the potential to revolutionize the way companies and organizations provide customer support. The widespread use of websites and online platforms has made it increasingly important for companies to provide efficient and user-friendly support to their customers, and a common way of achieving this is through the creation of a Frequently Asked Questions (FAQ) section. However, manually creating and updating an FAQ can be time-consuming and resource-intensive, and it can be difficult to predict all the questions a user might ask.

1.1 Motivation

In the field of NLP for the Arabic language, there has been relatively limited work on the topic of automatic question generation and answering. In particular, there hasn't been a single research article that concentrates on both question generation and answering in the Arabic language. This makes our contribution even more valuable because we will be filling a crucial gap in Arabic NLP.

The widespread use of websites and online platforms has made it increasingly important for companies and organizations to provide efficient and user-friendly support to their customers. A common way of achieving this is through the creation of a Frequently Asked Questions (FAQ) section. However, manually creating and updating an FAQ can be time-consuming and resource-intensive, and it can be difficult to predict all the questions a user might ask.

We can provide an efficient solution that can conserve time and resources while enhancing user experience by automating the process of question generation and answering. Moreover, the abundance of Arabic-speaking internet users makes it even more crucial to be able to manage Arabic texts.

In conclusion, our project has the potential to be both a significant contribution in the area of NLP and a practical tool for businesses and organizations seeking to enhance their customer experience. By automating the process of question generation and answering, users can help to create more effective and user-friendly FAQs and exams for Arabic documents, ultimately improving the user experience.

1.2 Aims and Objectives

1.2.1 Aims

- To develop a machine learning-based system for automatic question generation and answering for Arabic documents.
- To create an effective solution that can save time and resources while improving the user experience by automating the process of question generation and answering.

- To integrate the question generation and answering system into a web-based platform that can be used by companies and organizations to create FAQs and exams for Arabic documents.
- To make a contribution to the fields of Machine Learning and NLP by focusing on the underrepresented language of Arabic.

1.2.2 Objectives

1. To research and analyze existing methods and algorithms for automatic question generation and answering for Arabic language.
2. To design and implement a machine learning model for automatic question generation that can identify key information from an Arabic document and generate appropriate questions based on that information.
3. To design and implement a machine learning model for automatic question answering that can accurately answer questions generated by the question generation model.
4. To evaluate the performance of the automatic question generation and answering system using appropriate evaluation metrics such as accuracy, recall, and precision.
5. To integrate the question generation and answering system into a web-based platform that can be used by companies, organizations, and individuals to create FAQs for Arabic documents.
6. To evaluate the usefulness and effectiveness of the system through user testing and feedback.

1.3 Overview of the technical area

Natural language processing (NLP) is a subfield of artificial intelligence[2] and linguistics that focuses on enabling computers to understand, interpret, and generate human language. It has a wide range of applications, including question generation and answering systems, language translation, and text classification. In recent years, advances in NLP have made it possible for computers to understand and generate human language with increasing accuracy, and it has become an important tool for a variety of applications in industry and research[3].

As a subfield of artificial intelligence and machine learning[4], this section will discuss these terminologies, as well as introducing the subfields of NLP that will be used in this paper.

1.3.1 Artificial intelligence (AI)

Artificial intelligence (AI) is the simulation of human intelligence in machines that are programmed to think and act like humans. It involves the use of algorithms and statistical models

to process and analyze data, and it has a wide range of applications in fields such as computer science, engineering, and business[2].

There are several types of AI, depending on the level of intelligence and autonomy that the system exhibits, but there are three general ones. Rule-based AI, which adheres to a set of predetermined rules to carry out a certain activity, is the most fundamental type of AI, which is also referred to as weak AI. Expert systems, which are created to address problems in a particular field of knowledge, frequently utilize this kind of AI[5].

The second form of AI is called artificial general intelligence (AGI), which involves the development of systems that can perform any intellectual task that a human being can. This type of AI is still in the early stages of development, and it raises important ethical and philosophical questions about the potential impact of AI on society[6].

The last type of AI is referred to as “Super AI”, which represents an AI that thinks and performs better than humans. It’s supposed to have the ability to think, plan, learn, make judgments and communicate on its own. This type of AI is still purely Hypothetical, and achieving it is considered a world-changing task[7].

In conclusion, AI is the simulation of human intelligence in machines, and it has a wide range of applications in fields such as computer science, engineering, and business. There are several types of AI, ranging from rule-based systems to artificial general intelligence and super artificial intelligence, and the development of AI raises important ethical and philosophical questions.

1.3.2 Machine Learning (ML)

Machine learning is a subfield of artificial intelligence that involves the use of algorithms and statistical models to enable computers to learn and improve their performance on a specific task without explicitly being programmed to do so.

In machine learning, a computer is fed a large amount of data (called the "training data") and a set of rules or algorithms (called the "model") that can be used to make predictions or decisions based on that data. The computer then uses the model to make predictions or decisions about new, previously unseen data[4].

There are several types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning[8].

In supervised learning, the computer is given both the training data and the correct labeled output for each set of data. Then, with the aim of minimizing the difference between the predicted output and the correct output, the model is trained to generate predictions based on the provided training data[9].

In unsupervised learning, the computer is only given the input data and is left to find patterns or relationships in the data on its own. In other words, the data provided doesn't have a labeled

outcome, and the computer should, with utilizing several methods, figure it out. These methods include dimensionality reduction, association and clustering[10].

Reinforcement learning is another form of machine learning, in which the computer is provided with a goal or reward system and is made to learn by trial and error to maximize its reward by taking certain actions in an environment[11].

Many real-life applications incorporate machine learning, including personalized recommendations, fraud detection, healthcare, transportation, and manufacturing[12], and it's being employed to enhance efficiency, accuracy, and decision-making in these fields.

Overall, machine learning is a powerful tool for automating tasks and decision-making processes, and is being used in a wide variety of industries, including finance, healthcare, and transportation.

1.3.3 Natural language processing (NLP)

Natural language processing (NLP) is a subfield of artificial intelligence which focuses on how computers and human languages interact. It involves processing and analyzing natural language data, such as text or speech, using algorithms and statistical models[1].

Language translation, text summarization, sentiment analysis, and question-answering are just a few of the many uses for NLP. It has the potential to have an impact on many sectors of society, including education, business, and health care. It is frequently employed in subjects including computer science, linguistics, and psychology[3].

NLP, for instance, can be used to automatically translate text or speech from one language to another. This can be particularly helpful for firms that work in various languages, or for communication between people who speak different languages[13].

NLP can also be used in text summarization to automatically generate a condensed version of a longer text[14]. This can be useful for information retrieval, as it can help to identify the most important or relevant points in a text[15].

NLP may be used in sentiment analysis to automatically determine whether a text is expressing a good, negative, or neutral sentiment or emotion[16]. This might be helpful for analyzing consumer feedback as it enables organizations to analyze the viewpoints of their clients. It can also be used to identify threats and danger inducing texts, which helps organizations to protect their clients from dealing with these abnormal texts[17].

As far as question answering is concerned, NLP can be used to automatically answer a given question based on a given input text[18]. This can be useful in various applications such as education as it can help students to practice and learn, as well to help teachers in the process of generating questions for given material. It can also be used for clarification, as it can help to identify important or confusing points in a text.

In conclusion, NLP is a subfield of AI that focuses on the interaction between computers and human languages, and it has a wide range of applications in fields such as computer science and linguistics. It uses a variety of techniques and algorithms, such as tokenization, stemming, and part-of-speech tagging, to process and analyze natural language data, and it has the potential to impact many aspects of society.

1.3.4 Question Generation

Question generation is an important application of NLP, which involves automatic question generation from a given input text. This can be useful for education, customer service, data analysis and language learning[19].

A question generation system can be implemented using varying approaches. The first approach is syntax based, where subject-verb-object (SVO) sentences are used to form the questions[20].

Another approach to question generation using NLP is template-based generation, in which a template for the questions is created, and questions get generated according to that template[21].

A challenge faced in NLP question generation is to generate high-quality questions that are relevant, coherent, and appropriate for the given input. Another challenge is to generate questions in multiple languages or across domains, as this requires the ability to handle language-specific nuances and domain-specific knowledge.

In conclusion, question generation using NLP is a useful application of natural language processing that can help to automatically generate practice questions or identify important or confusing points in a text. It can be implemented using either a rule-based system or a template-based model, with the ability to incorporate machine learning, depending on the specific requirements of the application.

1.3.5 Question Answering

Natural language processing (NLP) can be used for question answering, which involves automatically answering a given question based on a given input text[18]. This is a more challenging task than question generation, as it requires the system to not only identify the relevant information in the input text, but also to understand the meaning of the question and generate a coherent and accurate answer.[22]

Question answering systems are implemented with various approaches. The most popular approach to NLP question answering is to use a machine learning-based system that has been trained on a large dataset of question-answer pairs[18].

Question Answering can be integrated into various applications, such as customer service platforms, search engines, and educational resources, to help users quickly find the information they need. It can also improve the efficiency and effectiveness of these applications by enabling them to understand and respond to natural language queries.

In conclusion, question answering using NLP is a useful application of natural language processing that can help to automatically answer questions based on a given input text. It can be implemented using a machine learning model, depending on the specific requirements of the application. Question answering systems can have many types, such as closed-domain and open-domain systems.

1.4 Overview of the report

This report presents the development of a machine learning-based system for automatic question generation and answering for Arabic documents. The report is organized into the following chapters:

- **Chapter 1: Introduction:** this chapter provides an overview of the project, including the problem being addressed, the importance of the problem, and the key features and contributions of the project.
- **Chapter 2: Background & Literature Review:** this chapter provides a comprehensive overview of the existing research and methods related to automatic question generation and answering. It includes a discussion of the relevant theory and a review of the previous work in the field.
- **Chapter 3: System Analysis:** a detailed analysis of the system model and architecture is introduced, including a description of the main work, solution approach, theory.
- **Chapter 4: Conclusion and Future Works:** this chapter summarizes the key findings of the project, provides recommendations for future work, and outlines potential future directions for the development of the system.

Chapter 2: Background & Literature Review

This chapter reviews similar work that is related to our project and points the differences from our work. It also clarifies the aims and objectives of our work, the constraints that we have to tackle, and the general flow of our work.

2.1 Background

2.1.1 Language Transformer Models

Before the advent of Language Transformer Models, Natural Language Processing (NLP) models relied primarily on Feedforward Neural Networks[23] and Recurrent Neural Networks (RNNs)[24] for tasks such as text classification, language generation, and question answering[25]. However, these models suffered from limitations such as slow training times, difficulty capturing long-term dependencies, and limited capacity to process large amounts of data.

Language Transformer Models, based on the Transformer architecture, emerged as a solution to these limitations. They have proven to be effective in capturing contextual relationships between words in a sentence and have achieved state-of-the-art performance on many NLP tasks.

The Transformer architecture was introduced in 2017 by a team at Google Brain in the paper "Attention is All You Need". It uses self-attention mechanisms to process the entire input sequence at once, allowing for parallel computation and reduced training times. In each layer of the model, the attention mechanism determines the significance of each element in the input sequence relative to the others[26].

This architecture has proven to be effective in NLP tasks such as text classification, machine translation, and question answering, with several language-specific models being developed, including BERT, GPT, and T5. These models have been pre-trained on large language datasets and fine-tuned for specific tasks.

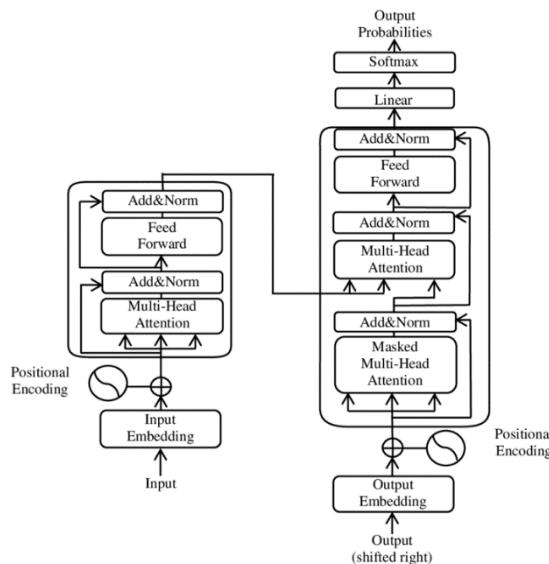


Figure 1: Transformer Model Architecture

2.1.1.1 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model developed by Google Research in 2018. It is based on the Transformer architecture and trained on a large corpus of text data, allowing it to capture contextual relationships between words in a sentence. BERT has achieved state-of-the-art performance on many NLP tasks and has become a popular base model for many downstream NLP applications. It is important to note that BERT is only the encoder part of the Transformer architecture, with 24 encoders and 16 bidirectional self-attention heads, totaling 340 million parameters[27].

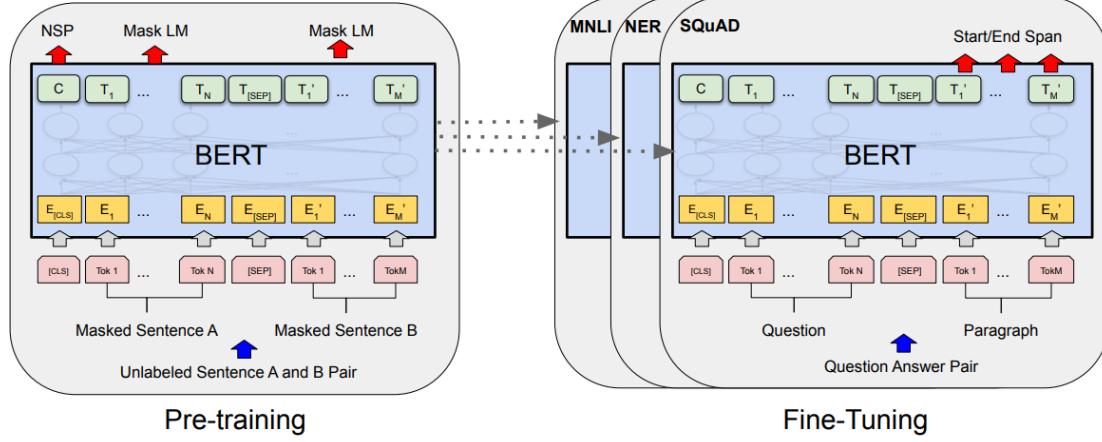


Figure 2: BERT pre-training and fine-tuning procedures

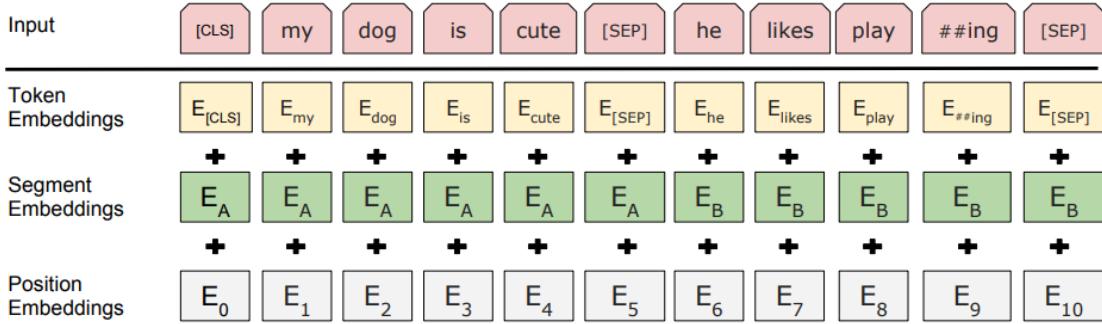


Figure 3: BERT input representation - token, segmentation and position embeddings

There are many Arabic versions of the BERT model, including AraBERT(2020)[28], Arabic-BERT(2020)¹, Arabic-ALBERT(2020)[29], GigaBERT(2020)[30], MarBERT(2020)[31], QARiB(2020)[32], CAMeLBERT(2021)[33], and JABER and SABER(2022)[34]. These models were developed by researchers in the field of NLP to improve performance on Arabic language tasks. Each model is pre-trained on a large corpus of Arabic text data, allowing them to understand the context and relationships between words in the Arabic language.

¹ <https://github.com/KUIS-AI/Arabic-ALBERT>

2.1.1.2 GPT

GPT (Generative Pretrained Transformer) is a family of language models developed by OpenAI, starting with GPT-1 in 2018[35], followed by GPT-2 (with 1.5 billion parameters) in November 2019[36] and finally with GPT-3 (with a 2048-token-long context and 175 billion parameters) in 2020[37], InstructGPT (1.3B, 6B, and 175B parameters) in 2022[38], and GPT-3.5² in 2022[39]. They are based on the Transformer architecture and are trained on a large corpus of text data, allowing them to generate coherent and contextually appropriate text. It is worth noting that GPT models are the decoder part of the Transformer architecture.

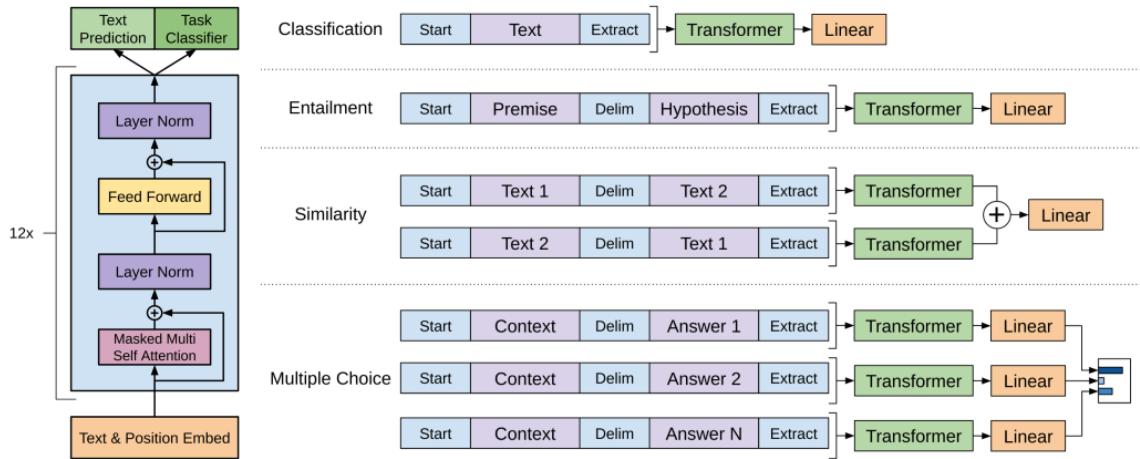


Figure 4: GPT-1 Transformer architecture and fine-tuning process

In November 30th, 2022, a GPT-3 based chatbot called ChatGPT was released to the public, causing waves in the world, and creating a market gap that businesses were racing to fill.³

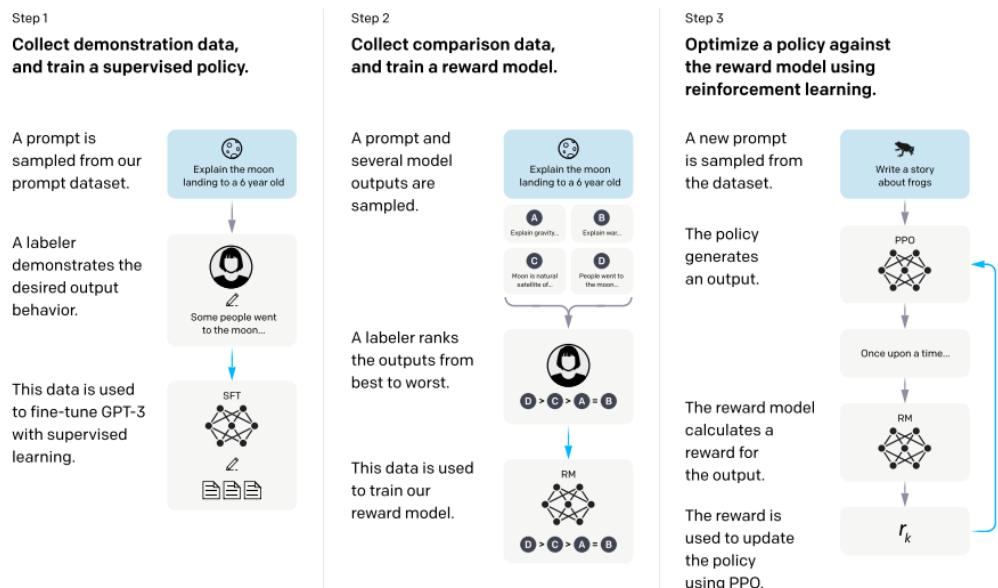


Figure 5: InstructGPT's three-step training process

² <https://platform.openai.com/docs/model-index-for-researchers>

³ <https://openai.com/blog/chatgpt/>

For the Arabic language, several models have been developed, one of these models is AraGPT2, developed by researchers at the American University of Beirut in 2020. With 1.46 billion parameters and a context size of 1024 tokens, it is pre-trained on a large corpus of Arabic text data, allowing it to understand the context and relationships between words in the Arabic language[40]. Another model is JASMINE: Arabic GPT Models for Few-Shot Learning 2022[41].

In addition to these models, two works have attempted to create an Arabic transformer causal language model. Khooli (2020)⁴ finetuned the OpenAI GPT-2 base model on Arabic Wikipedia dataset⁵, which was mainly trained on English text. Doiron (2020)⁶ also continued training on a collection of dialectal Arabic datasets, in order to create a dialectal Arabic GPT-2.

2.1.1.3 T5

T5 (Text-to-Text Transfer Transformer) adopts a standard encoder-decoder structure in contrast to other models like BERT and GPT. T5 can execute a variety of NLP tasks by fine-tuning the pre-trained model on a smaller task-specific dataset as it was trained on a huge corpus of text data[42].

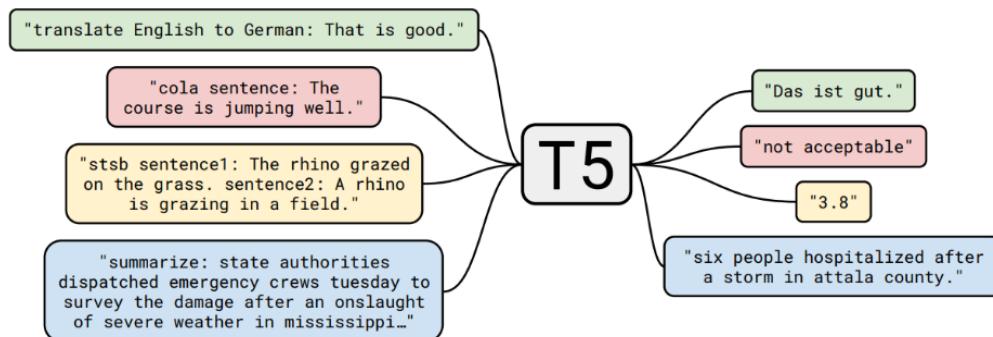


Figure 6: T5 Model Tasks

AraT5 is an Arabic version of T5, developed by researchers in 2022, that aims to evaluate the performance of the multilingual T5 (mT5) model on non-English tasks involving diverse data, specifically Arabic. The researchers pre-trained three powerful Arabic T5-style models and evaluated them on a novel benchmark called ARabic language GENeration (ARGEN), which covers seven important tasks. The results showed that the new models, pre-trained with 49% less data, performed significantly better than mT5 on all ARGEN tasks and set several new state-of-the-art results. Furthermore, the new models established new state-of-the-art results on the large Arabic language understanding evaluation benchmark ARLUE[43].

⁴ <https://huggingface.co/akhooli/gpt2-small-arabic>

⁵ <https://www.kaggle.com/datasets/abedkhooli/arabic-wiki-data-dump-2018>

⁶ <https://mapmeld.medium.com/making-a-mini-gpt-2-with-dialect-prompts-605a0edbc1e4>

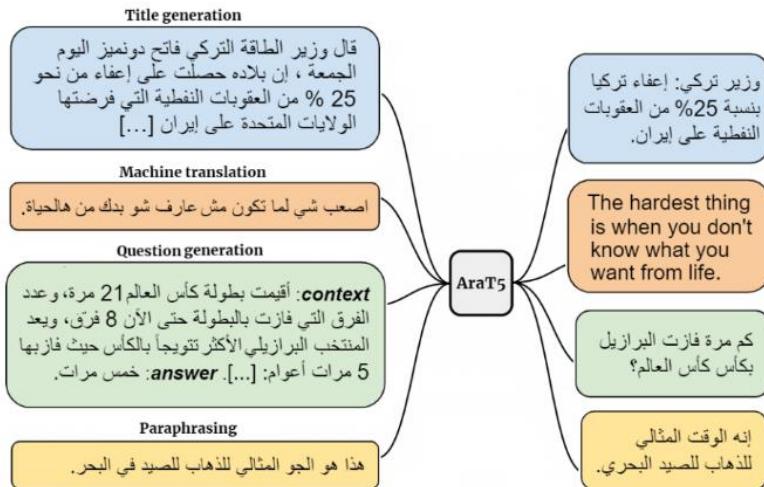


Figure 7: AraT5 Model Tasks

2.1.1.4 Models Comparison for QG and QA

The T5 model is a versatile language model that performs at jobs requiring question generating. T5 is a better alternative for question generation compared to BERT and GPT since it has been pre-trained for a variety of NLP tasks, including question generation, and has the capacity to produce text in response to a prompt.

It is possible to fine-tune BERT and GPT for question generation, but they might not perform as well as T5, which has been trained specifically for text-to-text transfer tasks. T5 may generate questions effectively without the need for fine-tuning, but in some circumstances when the job necessitates domain-specific knowledge that is not represented by the pre-training data, fine-tuning can still result in increased performance. The exact needs of the job and the quality of the pre-training data will determine whether or not T5 should be adjusted for question generating.

BERT and GPT can both be used for question answering tasks after fine-tuning[44]. **BERT** is specifically designed for this task, making it the best model for question answering[45][46]. GPT can also be used and is a strong performer in this task[47], but is not as well-suited as BERT. Fine-tuning allows each model to be adapted to the specific requirements of the question answering task and improves their performance, but BERT remains the best choice due to its specific design for this task.

2.1.2 Datasets

2.1.2.1 Arabic Q/A Datasets

To train our model, we have aggregated a dataset of 89,228 pairs of Arabic questions and answers derived from 35,021 paragraphs across 19,038 articles. The data was sourced from five different

sources and follows the format of SQuAD (Stanford Question Answering Dataset)[48], as described below.

```

file.json
├── "data"
└── [i]
    ├── "paragraphs"
    │   └── [j]
    │       ├── "context": "paragraph text"
    │       └── "qas"
    │           └── [k]
    │               ├── "answers"
    │               │   └── [l]
    │               │       ├── "answer_start": N
    │               │       └── "text": "answer"
    │               └── "id": "<uuid>"
    │                   └── "question": "paragraph question?"
    └── "title": "document id"
└── "version": 1.1

```

Figure 8: SQuAD format of Q&A datasets

1. The **Arabic-SQuAD dataset** is a translation of the English SQuAD dataset, which was translated for the purpose of the Neural Arabic Question Answering project[49]. It comprises 48,344 questions based on 10,364 paragraphs.
2. The **ARCD dataset** consists of 1,395 question-answer pairs derived from 465 paragraphs across 155 articles[49].
3. **TyDiQA** from Google comprises 15,726 paragraphs along with corresponding 15,726 questions and 16,425 answers[50].
4. The **MLQA (MultiLingual Question Answering) dataset** consists of 5,852 question-answer pairs derived from 5,085 paragraphs across 2,627 articles[51].
5. The **AAQAD dataset** comprises 17,911 question-answer pairs derived from 3,381 paragraphs across 299 articles[52].

2.1.2.2 English Q/A Datasets

In addition to the Arabic datasets, we have also collected a number of English datasets. These datasets will be translated using the Google Translate API if we find that the Arabic datasets are not sufficient. The English question-answer pairs will be evaluated after translation based on certain criteria; the pairs that pass the evaluation will be selected for further use.

- **MSMARCO (Microsoft Machine Reading Comprehension)** is a collection of datasets focused on deep learning in search developed by Microsoft. The first dataset, MSMARCO, featured 100,000 real Bing questions and corresponding human-generated answers. Since

then, the collection has expanded to include additional datasets, such as a natural language generation dataset, a conversational search dataset, and others[53].⁷

- **TriviaQA** is a reading comprehension dataset containing over 650K question-answer-evidence triples. It was authored by trivia enthusiasts and independently gathered evidence documents, providing high quality distant supervision for answering questions[54].⁸
- **HotpotQA** is a dataset for diverse, explainable multi-hop question answering.⁹
- **CoQA** is a large-scale dataset for building conversational question answering systems. It contains over 127,000 questions with answers collected from over 8000 conversations and includes unique features, such as conversational questions, free-form answers, and evidence subsequences highlighted in the passages[55].¹⁰
- **NewsQA** from Microsoft Research Montreal is a crowd-sourced machine reading comprehension dataset of 120K Q&A pairs, leveraging CNN articles from the DeepMind Q&A Dataset. The questions are written by human users and the answers may be multiword passages of the source text. The NewsQA dataset is collected through a three-stage, siloed process and is designed to challenge the research community to build algorithms capable of answering questions requiring human-level comprehension and reasoning skills.¹¹
- **SimpleQuestions** is a large-scale factoid question answering dataset consisting of 108,442 natural language questions paired with corresponding facts from the Freebase knowledge base[56].¹²
- There are many other English Q/A datasets available, including those found in a [GitHub repository](#) that collects information on various large Q/A.¹³

2.2 Literature Review

2.2.1 NLP

NLP utilizes several techniques and algorithms, depending on the specific application and the type of data that is being processed. To pre-process and evaluate the structure of natural language data, methods like tokenization, stemming, lemmatization and part-of-speech tagging are frequently utilized[1].

⁷ <https://github.com/microsoft/MSMARCO-Question-Answering>

⁸ <https://github.com/mandarjoshi90/triviaqa>

⁹ <https://github.com/hotpotqa/hotpot>

¹⁰ <https://stanfordnlp.github.io/coqa/>

¹¹ <https://www.microsoft.com/en-us/research/project/newsqa-dataset/>

¹² <https://github.com/davidgolub/SimpleQA/tree/master/datasets/SimpleQuestions>

¹³ <https://github.com/ad-freiburg/large-qa-datasets>

Tokenization involves splitting the input text into smaller units, called tokens, such as words or phrases. This is necessary for most NLP tasks, as instead of dealing with the data as a whole, it allows the system to process the data at the level of individual tokens[57].

Stemming involves reducing inflected or derived words to their word stem, which can be the base of the word. This is useful for text analysis, as it enables the system to group similar words, without taking different derivations or inflections into account[58].

Part-of-speech tagging involves identifying the part of speech (e.g., noun, verb, adjective) of each token in the input text. This allows the system to understand the structure of the data, and the role of each token in it[59].

Other techniques, such as syntactic parsing[60] and semantic analysis[61], are used to understand the meaning and context of the data.

2.2.2 Question Generation

There are many ways to implement a question generation system. The first approach is syntax based, in which a system can search for phrases that are in the form of a subject-verb-object (SVO) sentence, as these are often good candidates for questions. The system can replace a part of the sentence with a question form, e.g, replacing the subject with (“who”) or replacing the verb with (“what”). Moreover, this is the easiest way to create (“Yes”/”No”) questions. This is the approach that Intel uses in their smart question generation model[20].

Another approach to question generation using NLP is template-based generation, where a template for questions is predefined, and information is filled inside those placeholders. As templates have limits to how generic they are, this approach is usually used for question generation in specific fields[21].

Both of these approaches have their advantages and disadvantages. The rule-based approach is relatively simple to implement and can be tailored to the specific needs of the application. However, it is limited by the quality of the rules that are used, and it may not be able to generate high-quality questions for all input texts.

The template-based approach is really easy to implement, and can be specifically tailored to a specific subject. However, it can result in lack of uniqueness in questions, and inability to generate relevant questions if not implemented properly.

One potential enhancement suggested is to combine the previous approaches with machine learning algorithms, or use machine learning alone with a large amount of training data, in order to produce higher quality results.

Early question generation works used rule-based approaches, which rely on human-designed rules to extract features from the input text and capture linguistic relationships between words to generate questions[62]. However, these methods require extensive data annotation, which can be time-consuming and require domain-specific knowledge and experience. In recent years, there has

been a growth in the use of deep learning and neural networks for NLP tasks, including QG. These approaches, such as transformer models[26], have shown success in generating questions with high accuracy and performance on small datasets. These models can be pre-trained and fine-tuned on source tasks as a starting point, allowing them to save time and effort when applied to the target task. Some recent works have focused on generating questions from text only, without the use of answers as a guide, by relying on rule-based models or extracting meaningful phrases from the text using message-passing modules. These approaches have shown improvement in performance for QG tasks.

In a recent paper, Alhashedi, Mohd Suaib, and Bakri (2022) presented a model for generating questions automatically from long Arabic texts using a transformer BERT-base model[63]. The model is divided into two main categories: the question generation model, which is based on a BERT-base encoder-decoder architecture and has been adapted for the Arabic language, and a model for generating an N number of questions from a long text by extracting key sentences as input for the core model. The authors re-implemented the shared BERT-base encoder-decoder architecture from Rothe et al.[64] as the base model, and initialized all weights with an AraBERT checkpoint to support the Arabic language[28]. The model was trained on 70 million sentences containing 3 billion words from 23GB of Arabic text. To generate N number of questions from a single, unlimited-length document, the authors employed the TextRank algorithm[65] to extract the important sentences and used these as input for the question generation model.

In a paper published in 2020[66] an automatic question generation approach for the Arabic language called AQG was presented. The approach combines semantic role labeling using PropBank[67] and the flexibility of question models based on a semantic representation. AQG was designed to generate questions for use on Quizzito, an online platform for children's quizzes that currently contains more than 254.5k quizzes. The system was implemented using patterns and templates based on regular expressions and achieved an f-measure of 86%. The design steps for AQG begin with the analysis of a text in Arabic using the MADAMIRA[68] morphological analyzer and the STAR (Arabic Texts Slicer) tool[69] for sentence segmentation. The system then proceeds to the generation and assignment of roles to the actors in the sentence using PropBank SRLs adapted for the Arabic language. Finally, patterns and templates are used to generate the questions based on a semantic model. Preprocessing and annotation of the texts is carried out using MADAMIRA, which produces a rich set of features for each word and has an accuracy of 96% for lemma and stem and 99% for word segmentation.

2.2.3 Question Answering

There are several approaches to question answering using NLP. The most popular approach to NLP question answering is to use a deep learning-based system that has been trained on a large dataset of question-answer pairs. The Stanford Question Answering Dataset (SQuAD)[47] is a popular dataset for training and evaluating such systems. It consists of a large number of articles and corresponding questions, along with the answers to those questions.

Types of question answering systems include closed-domain systems, where a system can only answer questions about a specified field, with no ability to generate results beyond that domain[70].

Another type for NLP question answering is an open-domain system, where the system uses a pre-defined knowledge base or database to answer the question. The knowledge base can contain a large amount of information about the world, including facts, definitions, and relationships between entities[71]. The system should be able to deal with any kind of question, and is currently under development, with its latest systems being GPT-3 and BERT.

Benajiba developed an Arabic Question Answering (ArabiQA) system[72], which includes a Passage Retrieval module, an Answer Extraction module, and an Answers Validation module. In order to build the system, Benajiba conducted a study of the Arabic language from a Natural Language Processing perspective and created an Arabic corpus for testing the text passage retrieval system. They also adapted the passage retrieval system to retrieve passages from Arabic text and were able to significantly improve its performance. The team also developed the annotated ANERcorp corpus and lexical resources for training and testing a Named Entity Recognition system, called ANERsys, which is based on the Maximum Entropy approach. They later enhanced ANERsys by using a two-step approach, which helped to increase its performance by nearly 10 points. In addition, the team developed an Arabic corpus of questions, relevant passages, and correct answers to test an Answer Extraction module for factoid questions. These efforts contribute to the field of question answering by addressing the specific challenges of processing Arabic language and text.

QARAB[73] is a question answering system that takes natural language questions expressed in the Arabic language and attempts to provide short answers. It uses traditional information retrieval techniques coupled with a sophisticated natural language processing approach to identify the answer. Keyword matching is used along with matching simple structures extracted from both the question and the candidate documents selected by the IR system. An analysis of Arabic question forms is also conducted to better understand the kinds of answers that users find satisfactory. QARAB is evaluated over a range of question types provided by Arabic users and the answers produced by the system are assessed for satisfaction. The IR system used in QARAB is based on Salton's vector space model and an inverted file system is constructed from the text collection of the Al-Raya newspaper. The NLP system in QARAB includes a dictionary and a tagger to identify proper names and other crucial lexical items, as well as a parser to process the input and candidate documents.

A knowledge-based Arabic question answering system (AQAS)[74] was developed to accept Arabic sentences in the form of declarative statements or questions and generate appropriate output for the user. AQAS is an Arabic implementation of a general natural language processing system and includes four main components: a parser, an interpreter, a generator, and a knowledge base. The parser converts the input sentence into an internal meaning representation (IMR) for processing by the interpreter. The interpreter processes the IMR using a knowledge base represented using frames, and the generator produces the final output. AQAS is designed to accept Arabic queries that follow specific rules, including argument interrogative sentences, imperative sentences, propositional interrogative sentences, and declarative sentences. The parser includes a

dictionary grammar and a morphological algorithm to identify the morphemes of a word and locate it in the dictionary. The interpreter processes the IMR using case frames and generates a plan, which is executed by the generator to produce the final output. The knowledge base is implemented as a set of case frames, each containing a set of slots representing the relevant information for a specific topic.

2.2.4 Q/A Combined

"Improving Q/A Using Arabic WordNet" by Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso[75] presents an approach for improving question and answering systems through a query expansion[76] process using an ontology built with Arabic WordNet[77]. The authors conduct preliminary experiments to demonstrate the effectiveness of their approach in improving the accuracy of expected answers. The ontology is built using the Amine open-source platform and Arabic WordNet, with the aim of combining the richness and accuracy of the lexical resource with the characteristics of the platform. The procedure for building the Amine AWN ontology involves exporting data from Arabic WordNet, creating a type hierarchy based on mappings to the Suggested Upper Merged Ontology, and adding Arabic synonyms to the types. The authors also discuss the use of the ontology in a QE process, including expansion using semantic relations and morphological variants.

2.3 Related Works

2.3.1 LongShot¹⁴

LongShot is an AI writing assistant that helps you and your team create helpful blogs that rank on Google. LongShot AI's 40+ dedicated features allow you to create helpful SEO content that gives you visibility on Google.

The main purpose of LongShot is not FAQ generating, LongShot provides some of the most comprehensive list of features, which offer lots of use-cases, such as Content Rephraser, Headline Generator, Blog Ideas Generator, Text Extender, FAQs Generator, and many other features.

The LongShot FAQ generator is only able to generate questions, not answers. Additionally, LongShot supports eight languages, but Arabic is not included.

2.3.2 Quillionz¹⁵

Quillionz is an online question generator that uses artificial intelligence (AI) to create questions from texts. starting by pasting or uploading from 300 to 3,000 words of text, a title, and a domain into the platform. Then following step-by-step prompts to select or deselect keywords, optimize the text, and generate questions for review. Quillionz offers various types of questions for generation, including true/false, fill-in-the-blank, and "Wh" questions (e.g. beginning with "what"

¹⁴ <https://www.longshot.ai/features/faqs-generator>

¹⁵ <https://www.quillionz.com/>

or "who"). Quillionz generates questions along with their answers and only supports the English language.

2.3.3 Questgen¹⁶ ¹⁷

Questgen is a website that offers a writing tool for creating various types of assessments such as “Multiple-Choice Questions (MCQs), True/False Questions, Higher-Order Questions (HOQs), Frequently Asked Questions (FAQs)”, and so on in a single click. Before exporting the produced worksheet as simple text or a PDF, you can modify and make changes to it.

Questgen AI is an open source NLP library that focuses on creating question generation algorithms that are simple to use. It is attempting to create the world's most advanced question generation AI by utilizing cutting-edge transformer models such as T5, BERT, and OpenAI GPT-2, among others.

Questgen only supports the English language for generating various types of questions and their corresponding answers.

Demo:

Article:

“Over a century, what began as a small girls’ school in Birzeit town has become the most prestigious Palestinian university, transforming Palestinian higher education through its impact on community awareness, culture and resistance. Birzeit University has been a thorn in the side of the occupation, insisting on playing its role of enlightenment and creating a multicultural Palestinian society on the campus grounds.

Birzeit University began a school for girls in the town of Birzeit that was founded by the educator Nabiha Nasir (1891-1951). The main objective of the school was to provide girls from Birzeit town and the neighboring villages with the opportunity for primary education, at a time when there were no schools in that region. The educator Rateebah Shouqair (1881-1957) was the school’s first principal.

In 1932, Ms. Shouqair moved to Beit Jala to establish the Good Shepherd School, now known as the Good Shepherd Swedish School, and became principal of Birzeit School.

In the same year, a boys school was established, headed by Mr. Wadi’ Tarazi. This became Birzeit Secondary School, which included both boys and girls. The school was headed by Musa Nasir (1895-1971).

Birzeit College teachers were divided between boys’ teachers and girls’ teachers. The girls’ school was located in the house of the pastor Hanna Nasir, which comprised classes and a dormitory, as well as administration offices. The boys’ school was located in rented houses and buildings in the neighborhood. One of these buildings was used for classes and teachers’ accommodation, and another building was used

¹⁶ <https://www.questgen.ai/>

¹⁷ <https://github.com/ramsrigouthamg/Questgen.ai>

as residence for boys and a dining hall for teachers and students. Both schools used a third building, which was owned by the Nasir family and was used for public lectures, plays, debates and other activities.

In June 1947, the first issue of the College Voice Magazine was published. It was a monthly magazine published in cooperation with the college alumni. It included high level scientific research and articles on social issues. The aim was to offer community service and create an opportunity to train students in publishing their views. After the Nakba in 1948, Birzeit College continued to play its role by hosting Palestinians taking refuge in the West Bank and Gaza Strip.

Musa Nasir took over the chairmanship of Birzeit University (1859-1971 in 1953 and the freshman class in Arts & Sciences was initiated, followed in 1961 by the sophomore class.

Since Birzeit College was the only institution in the country at that time, it was decided to focus solely on university education. Thus, beginning in 1961, the primary, intermediate and secondary grade classes were cancelled by the end of the academic year 1966/1967 and lessons were restricted to freshman and sophomore classes.”

Generated FAQ:

Question: “What is Birzeit University?”

Answer: “Birzeit University is a Palestinian university located in Birzeit town in the central West Bank. It was founded in 1937 as a girls' school and became a university in 1969. Birzeit is known for its emphasis on community awareness, culture and resistance.”

Question: “Who established the first school in Birzeit town and the surrounding villages?”

Answer: “Rateebah Shouqair established the first school in Birzeit town and the surrounding villages in 1932.”

Question: “What was the main objective of the school?”

Answer: “the main objective of the school was to provide girls from Birzeit town and the neighboring villages with the opportunity for primary education, at a time when there were no schools in that region.”

Question: “What was the name of the school headed by Musa Nasir?”

Answer: “Birzeit College teachers were divided between boys' teachers and girls' teachers. The girls' school was located in the house of the pastor Hanna Nasir, which comprised classes and a dormitory, as well as administration offices. The boys' school was located in rented houses and buildings in the neighborhood. One of these buildings was used for classes and teachers' accommodation, and another building was used as residence for boys and a dining hall for teachers and students.”

Question: “Who was Musa Nasir?”

Answer: “Musa Nasir was a Palestinian educational figure who served as the first president of Birzeit College.”

Question: “Who owned the third building?”

Answer: “the Nasir family owned the third building.”

Question: “What was the first issue of the College Voice Magazine?”

Answer: “the first issue of the College Voice Magazine was published in June 1947.”

Question: “What was the Nakba?”

Answer: “the Nakba, or ‘Catastrophe,’ refers to the displacement of more than 700,000 Palestinians during the 1948 war.”

Question: “What is the Nakba?”

Answer: “the Nakba, or Catastrophe, refers to the Palestinian loss of independence and homeland due to the establishment of the Israeli state in 1948.”

2.3.4 PrepAI¹⁸

PrepAI is a smart, user-friendly, and intuitive English question generation platform driven by cutting-edge AI/ML technology. It allows users to generate text and media rich-question papers in a variety of formats, as well as make tests less of a hassle for any class or course.

PrepAI takes a variety of input types. The tool is intended to make it simple for instructors and educators to generate questions. PrepAI reads data from the following sources:

- Type directly into the input box or Copy/ Paste Text.
- PDF files.
- MS Word Docx.
- YouTube URL (with or without captions).
- Video files.
- Wikipedia Search Results from Wikipedia search by entering the topic.

¹⁸ <https://www.prepai.in/>

PrepAI provides a variety of question categories by automatically generating questions and classifying them. PrepAI offers the following kinds of questions:

- MCQs (Multiple Choice Questions).
- True/False.
- Descriptive questions.
- Fillups.
- Statement Based Questions.

PrepAI was developed with educators in mind. The ML and NLP algorithms function similarly to how a teacher's thought would function when producing an exam paper. Before users complete the question paper and send it in the format of their choosing, users can select one of three challenge levels: "Keep it easy, Medium/ moderately easy or Hard."

Demo:

Article:

"Over a century, what began as a small girls' school in Birzeit town has become the most prestigious Palestinian university, transforming Palestinian higher education through its impact on community awareness, culture and resistance. Birzeit University has been a thorn in the side of the occupation, insisting on playing its role of enlightenment and creating a multicultural Palestinian society on the campus grounds.

Birzeit University began a school for girls in the town of Birzeit that was founded by the educator Nabiha Nasir (1891-1951). The main objective of the school was to provide girls from Birzeit town and the neighboring villages with the opportunity for primary education, at a time when there were no schools in that region. The educator Rateebah Shouqair (1881-1957) was the school's first principal.

In 1932, Ms. Shouqair moved to Beit Jala to establish the Good Shepherd School, now known as the Good Shepherd Swedish School, and became principal of Birzeit School.

In the same year, a boys school was established, headed by Mr. Wadi' Tarazi. This became Birzeit Secondary School, which included both boys and girls. The school was headed by Musa Nasir (1895-1971).

Birzeit College teachers were divided between boys' teachers and girls' teachers. The girls' school was located in the house of the pastor Hanna Nasir, which comprised classes and a dormitory, as well as administration offices. The boys' school was located in rented houses and buildings in the neighborhood. One of these buildings was used for classes and teachers' accommodation, and another building was used as residence for boys and a dining hall for teachers and students. Both schools used a third building, which was owned by the Nasir family and was used for public lectures, plays, debates and other activities.

In June 1947, the first issue of the College Voice Magazine was published. It was a monthly magazine published in cooperation with the college alumni. It included high level scientific research and articles on social issues. The aim was to offer community service and create an opportunity to train students in publishing their views. After the Nakba in 1948, Birzeit College continued to play its role by hosting Palestinians taking refuge in the West Bank and Gaza Strip.

Musa Nasir took over the chairmanship of Birzeit University (1859-1971 in 1953 and the freshman class in Arts & Sciences was initiated, followed in 1961 by the sophomore class.

Since Birzeit College was the only institution in the country at that time, it was decided to focus solely on university education. Thus, beginning in 1961, the primary, intermediate and secondary grade classes were cancelled by the end of the academic year 1966/1967 and lessons were restricted to freshman and sophomore classes.”

Generated questions and answers:

Total Questions: 76

• MCQ. EASY LEVEL (16 Questions)

Q1. What town began as a small girls school?

- I. Jerusalem
- II. Ramallah
- III. Hebron
- IV. Birzeit

Q2. Who founded birzeit university?

- I. Khalil Al-Sakakini
- II. Nabiha Nasir
- III. Edward Said
- IV. Mahmoud Darwish

Q3. What was the main objective of birzeit university?

- I. Research
- II. Primary Education
- III. Business
- IV. Tourism

• MCQ. MEDIUM LEVEL (27 Questions)

Q17. What town began as a small girls school?

- I. Ramallah
- II. Hebron
- III. Jerusalem
- IV. Birzeit

Which of the options given above is/are correct:

- I. II and I only.
- II. IV only.
- III. IV, II and I only.
- IV. III only.

Q18. Who founded birzeit university?

- I. Mahmoud Darwish
- II. Khalil Al-Sakakini
- III. Nabiha Nasir
- IV. Edward Said

Which of the options given above is/are correct:

- I. II and I only.
- II. III only.
- III. III, II and I only.
- IV. IV only.

- MCQ. HARD LEVEL (3 Questions)

Q44. Read the following statements carefully:

Statement I: Over a century, what began as a small girls school in birzeit town has become the most prestigious palestinian university, transforming palestinian higher education through its impact on community awareness, culture and resistance.

Statement II: The main objective of the school was not to provide girls from birzeit town and the neighboring villages with the opportunity for primary education, at a time when there were no schools in that region.

Statement III: In 1932, ms. shouqair moved to beit jala to establish the good shepherd school, now known as the good shepherd swedish school, and became principal of birzeit school.

Which of the above Statement(s) is/are correct:

I. I and III.

II. II and III only.

III. II only.

IV. All Statements are Correct.

- True & False (18 Questions)

Q47. Primary education was the main objective of birzeit university.

I. True

II. False

Q48. Read the following statements carefully:

Statement I: research was the main objective of birzeit university.

Statement II: Primary education was the main objective of birzeit university.

Which of the following option is correct:

I. Both Statements are False.

II. Statement I is True and Statement II is False.

III. Statement II is True.

IV. Both Statements are True.

Q49. Rateebah shouqair was the first principal of birzeit university.

I. True

II. False

- MCQ. Fillups (9 Questions)

Q68. _____ was the main objective of birzeit university.

I. Tourism

II. Business

III. Research

IV. Primary Education

Q69. _____ was the first principal of birzeit university.

I. Yasser Arafat

II. Rateebah Shouqair

III. Saeb Erekat

IV. Mahmoud Abbas

Q70. _____ was the name of the boys school that included both boys and girls.

I. Al-Quds University

II. Al-Quds Open University

III. Birzeit Secondary School

IV. Al-Najah University

- Descriptive (3 Questions)

Q65. Where was the boys school located?

Answer. Rented houses and buildings in the neighborhood

Q66. What was the other building used as for teachers and students?

Answer. Residence for boys and a dining hall

Q67. What was included in the first issue of the college voice magazine?

Answer. High level scientific research and articles on social issues

2.3.5 ChatGPT¹⁹

ChatGPT is an innovative model designed to interact with users in a conversational manner. The model's unique dialogue format allows it to answer follow-up questions, acknowledge its mistakes, challenge false assumptions, and reject inappropriate requests. ChatGPT is closely related to InstructGPT, a model that is specially trained to provide detailed responses to prompts.

The development of ChatGPT was a result of Reinforcement Learning from Human Feedback (RLHF), using methods similar to InstructGPT[38]. However, there were slight differences in the data collection setup. Initially, a supervised fine-tuning process was used to train the model. The AI trainers played both sides of a conversation, the user and the AI assistant, and had access to model-generated suggestions to help with their responses.

To create a reward model for reinforcement learning, comparison data was collected by taking conversations between the AI trainers and the chatbot. A model-generated message was randomly selected, and several alternative responses were sampled. The AI trainers then ranked these alternatives based on their quality. By utilizing these reward models, the model was fine-tuned using Proximal Policy Optimization, which was repeated several times.

ChatGPT can generate Frequently Asked Questions and their answers if you ask it to, but it was not designed for this purpose, so it may not provide the correct and necessary output. While ChatGPT can process and generate FAQs in Arabic, it is not optimized for this language and may not always produce accurate results. Additionally, if the input text is long, ChatGPT may not be able to process it properly, and sometimes you need to ask the bot multiple times using different ways in order to accomplish the task.

Demo:

Given Article:

“وَقَعَتْ جَامِعَةُ بَيْرُزِيتْ مُمَثَّلَةً بِرَئِيسِهَا دُ. بِشَارَةِ دُوْمَانِي، وَمُجْمِعَةِ النَّبَالِيِّ وَالْفَارِسِ الْقَابِضَةِ مُمَثَّلَةً بِرَئِيسِهَا التَّنْفِيذِيِّ مُ. خَالِدَ فَارِسِ، الْيَوْمِ الْخَمِيسِ 15 كَانُونِ الْأَوَّلِ 2022، مَذَكُورَةٌ تَفَاهُمٌ مُشَتَّرٌ بَيْنِ الْطَّرْفَيْنِ، وَإِطْلَاقُ مُنْحَةٍ بِاسْمِ ‘مَنْحَةِ الْمُهَندِسِ خَالِدِ الْفَارِسِ’ لِطَلَابِيَّةِ الْهُوَدِيَّةِ وَالتَّكْنُولُوْجِيَّةِ فِي جَامِعَةِ بَيْرُزِيتْ، حِيثُ يَأْتِي هَذَا التَّعَاوُنُ تَابِيًّا لِلْأَهَدَافِ الْإِسْتَراتِيجِيَّةِ الْمُشَتَّرَكَةِ، وَخَدْمَةً لِلْمُجَمَّعِ الْفَلَسْطِينِيِّ.

¹⁹ <https://openai.com/blog/chatgpt/>

وأكـدـ مـ الفـارـسـ عـلـىـ أـهـمـيـةـ هـذـاـ التـعـاـونـ فـيـ تـبـادـلـ الـخـبـرـاتـ وـالـكـفـاءـاتـ بـيـنـ الـطـرـفـيـنـ،ـ مـشـيـرـاـ إـلـىـ أـنـ توـقـيعـ مـذـكـرـاتـ التـفـاهـمـ مـعـ الـمـؤـسـسـاتـ الـتـعـلـيمـيـةـ يـهـدـيـ إـلـىـ تـعـزـيزـ رـسـالـةـ الـمـجـمـوعـةـ بـتـحـقـيقـ الـمـصـلـحةـ الـعـامـةـ وـخـدـمـةـ كـافـةـ شـرـائـحـ الـمـجـتمـعـ،ـ وـأـضـافـ:ـ أـنـ هـذـهـ الـمـنـحـةـ تـأتـيـ فـيـ إـطـارـ الـمـسـؤـلـيـةـ الـمـجـتمـعـيـةـ مـنـ أـجـلـ إـفـادـةـ الـطـلـبـةـ وـمـسـاعـدـتـهـمـ خـلـالـ رـحـلـتـهـمـ جـامـعـيـةـ،ـ وـتـحـفيـزـهـمـ عـلـىـ السـيرـ ضـمـنـ أـهـدـافـ وـاـضـحـةـ تـسـاعـدـهـمـ عـلـىـ تـنـمـيـةـ خـبـرـاتـهـمـ وـقـدـرـاتـهـمـ الـتـعـلـيمـيـةـ لـتـأـهـيلـهـمـ لـسـوقـ الـعـلـمـ.“

Generated FAQ:

السؤال: “ما هي الجامعة الممثلة من قبل رئيسها؟”

الإجابة: “الجامعة الممثلة من قبل رئيسها هي جامعة بيرزيت.”

السؤال: “ما هي المجموعة التي تمثلها مـ الفـارـسـ؟”

الإجابة: “المجموعة التي تمثلها مـ الفـارـسـ هي مـجمـوعـةـ النـبـالـيـ وـالفـارـسـ الـقـابـضـةـ.”

السؤال: “ما هي الهدف من إطلاق المنحة؟”

الإجابة: “الهدف من إطلاق المنحة هو تعزيز رسالة المجموعة بتحقيق المصلحة العامة وخدمة كافة شرائح المجتمع، وتعزيز رسالة المجموعة بتحقيق المصلحة العامة وخدمة كافة شرائح المجتمع.”

السؤال: “ما هي مـسـؤـلـيـةـ الـمـجـمـوعـةـ الـذـكـورـةـ فـيـ النـصـ؟ـ”

الإجابة: “مسـؤـلـيـةـ الـمـجـمـوعـةـ الـذـكـورـةـ فـيـ النـصـ هيـ الـمـسـؤـلـيـةـ الـمـجـتمـعـيـةـ.”

السؤال: “ما هي التفاهم المشترك الذي وقعت عليه الجامعة والمجموعة؟”

الإجابة: “التفاهم المشترك الذي وقعت عليه الجامعة والمجموعة هو منكرة تفاهم مشتركة.”

السؤال: “متى وقعت التفاهم؟”

الإجابة: “التفاهم وقع في الخميس 15 كانون الأول 2022.”

السؤال: “ما هي الجامعة التي تستفيد من المنحة؟”

الإجابة: “الجامعة التي تستفيد من المنحة هي جامعة بيرزيت.”

السؤال: “ما هي كلمة الرئيس التنفيذي للمجموعة عن التعاون مع الجامعة؟”

الإجابة: “كلمة الرئيس التنفيذي للمجموعة عن التعاون مع الجامعة هي أنه يهدف إلى تعزيز رسالة المجموعة بتحقيق المصلحة العامة وخدمة كافة شرائح المجتمع، وأن هذا التعاون يهدف إلى تبادل الخبرات والكفاءات بين الطرفين.”

السؤال: “ما هي الهدف الرئيسي من المنحة؟”

الإجابة: “الهدف الرئيسي من المنحة هو مساعدة الطلبة وتحفيزهم على السير ضمن أهداف واضحة تساعدهم على تنمية خبراتهم وقدراتهم التعليمية”

Chapter 3: System Analysis

3.1 System Description

The product being developed is a deep learning-based system for automatic question generation and answering for Arabic documents, with a user-friendly interface that will be built using ReactJS and a backend built using Python Django. The system utilizes a MySQL database to store and manage the data.

The system leverages advanced machine learning techniques, such as natural language processing (NLP) and deep learning, to analyze and understand the text, identify important information, and generate relevant questions. The system also uses these techniques to provide accurate answers to the generated questions.

The system is designed to be highly scalable and efficient. The user-friendly interface, built using ReactJS, allows users to input the text they would like to generate questions from, and view the generated questions and answers. The backend, built using Python Django, manages the data and processes the requests, utilizing the power and efficiency of the Python programming language. Each user will have a history of their usage stored in the MySQL database, allowing for tracking and analysis of their usage patterns.

3.1.1 System Objectives

The objectives of the system are as follows:

1. To generate relevant and high-quality questions from Arabic text documents.
2. To provide accurate answers to the generated questions.
3. To offer a user-friendly interface for inputting text and viewing generated questions and answers.
4. To store and manage data, in a scalable and efficient manner using a MySQL database.

3.1.2 System Main Features

1. Automatic Question Generation: the system automatically generates relevant and high-quality questions from Arabic text documents.
2. Accurate Question Answering: the system provides accurate answers to the generated questions, leveraging advanced machine learning algorithms and NLP techniques.

3. User-Friendly Interface: the system offers a user-friendly interface for inputting text and viewing generated questions and answers, allowing for a seamless and intuitive user experience.
4. Data Management: the system stores and manages data, including each user's history of usage, in a scalable and efficient manner using a MySQL database.

3.1.3 Operating Environments

The system is designed to operate in the following environments:

1. **Frontend:** the system's frontend is built using ReactJS, a JavaScript library for building user interfaces.
2. **Backend:** the system's backend is built using Python Django, a high-level Python web framework for rapid development.
3. **Database:** the system uses a MySQL database for data storage and management, providing a scalable and efficient solution for handling large amounts of data.
4. **Computing Environment:** the backend system requires a computer with sufficient processing power and memory to run the machine learning algorithms and NLP techniques used in question generation and answering. This can be achieved through a cloud computing service like AWS.
5. **User Environment:** To use the system, users will need a web browser and an internet connection.

3.1.4 Constraints

Our project team has taken into consideration a number of constraints that may impact the development and implementation of the question generation and answering system. Some of these constraints include:

1. Data Availability: the availability and quality of the training data used for question generation and answering is a key constraint that we have taken into account.
2. Language Processing: the accurate processing and understanding of the Arabic language, particularly with regards to idiomatic expressions and cultural nuances, is a challenge that we have to tackle.
3. Algorithm Performance: the performance of the question generation and answering algorithms may be limited by the complexity and variability of the input data, and we have had to consider this when designing the system.

4. Technical Skills: Our development team must have specialized technical skills in order to implement and optimize the system, including proficiency in Python, Django, ReactJS, and database management.

By taking these constraints into account, we aim to minimize their impact and ensure the success of the project.

3.1.5 Functional Requirements

User Requirements:

1. Question Generation: the system must be able to automatically generate questions based on input text or documents
2. Question Answering: the system must be able to provide accurate and relevant answers to the generated questions.
3. Arabic Language Support: the system must support the Arabic language, allowing users to input Arabic text and receive questions and answers in Arabic.
4. User History: the system must store a history of each user's submissions, allowing them to review previous questions and answers along with the original text.

System Requirements:

1. Question Generation Model: the system must include an algorithm for automatically generating questions based on input text.
2. Question Answering Model: the system must utilize a machine learning model for answering the generated questions.
3. Database Management: the system must store all generated questions, answers, and user information in a database, using MySQL.
4. Front-End Interface: the system must include a user-friendly front-end interface, built using a technology such as ReactJS, to allow users to interact with the system and input text.
5. Back-End Support: the system must include a back-end support system, built using Django, to manage the algorithms and database interactions.
6. Cloud Services: the system must be able to run on a cloud-based service, such as AWS, to allow for scalability and improved performance.
7. Security Measures: the system must include appropriate security measures to protect user information and ensure the confidentiality of the data stored in the database.
8. Data Privacy: the system must comply with data privacy regulations and protect user information from unauthorized access.

3.1.6 Non-Functional Requirements

The non-functional requirements for the question generation and answering system include:

1. Security: the system must ensure the security of user data and be protected against unauthorized access, modification, or theft.
2. Reliability: the system must be reliable, with minimal downtime and errors, and must provide consistent performance even under heavy loads.
3. Usability: the system must be easy to use, with a clear and intuitive interface, and must provide a positive user experience.
4. Maintainability: the system must be easy to maintain and upgrade over time, with clear and well-documented code and a modular architecture.
5. Compatibility: the system must be compatible with a wide range of operating systems, browsers, and hardware platforms, and must be accessible to users with varying levels of technical skill and experience.
6. Performance: the system should be able to respond to user requests in a timely manner, with minimal latency and downtime.
7. Scalability: the system should be able to handle an increasing number of users and requests without significant degradation in performance.

3.2 System Design

3.2.1 System Architecture

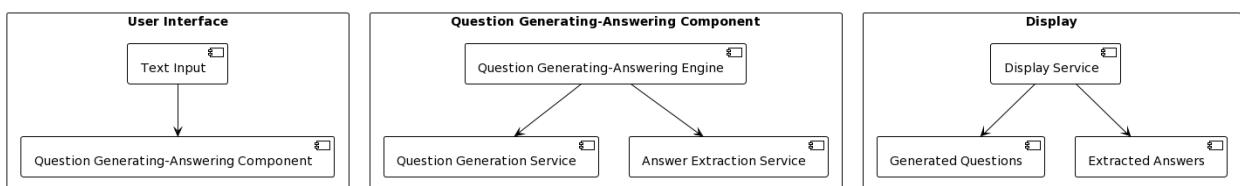


Figure 9: System Overall Architecture

Our framework mainly utilizes two models, a question generation model which is T5 based, and a question answering model which is BERT based. First, the input text is passed into the question generation model. After generating the questions, these questions are passed into the question answering model along with the original text. Then, the question answering model uses both the questions and the text to output the desired answers. Lastly, both the generated questions and answers are outputted for our system to display and store.

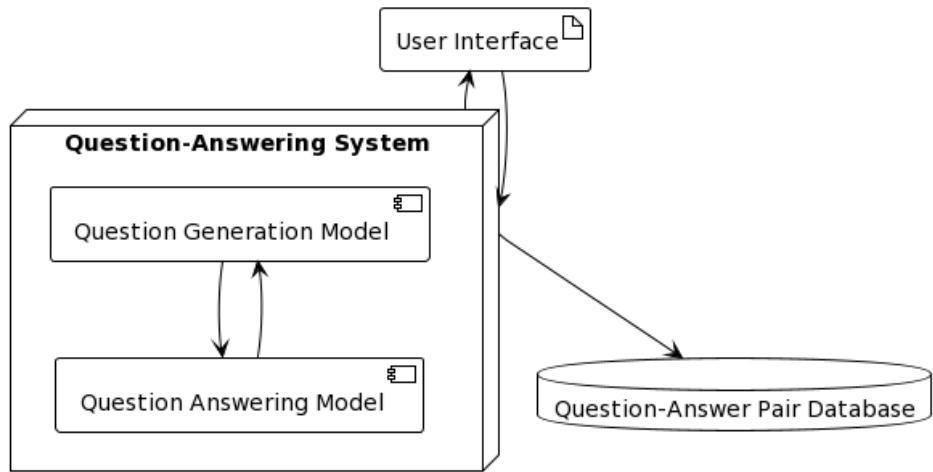


Figure 10: Question Generation and Answering Model

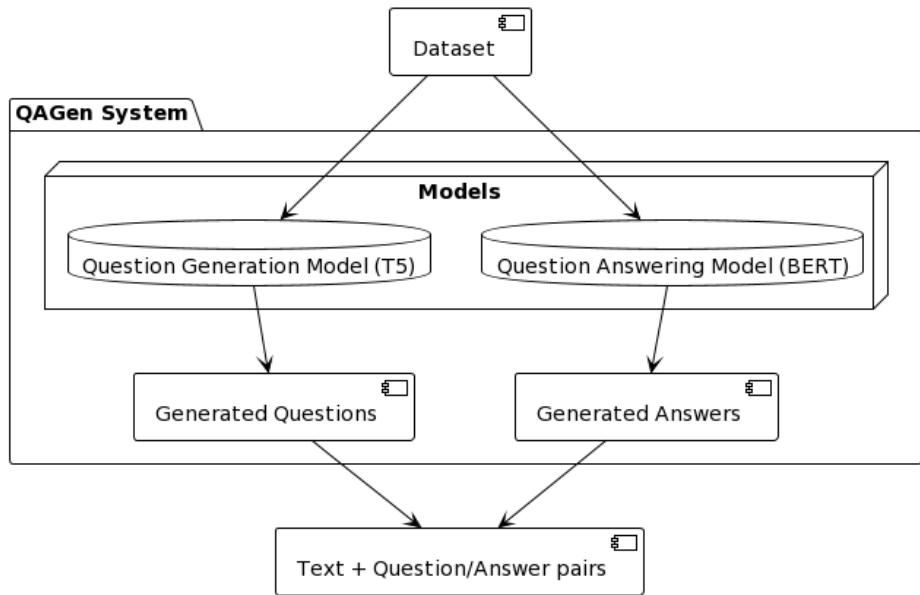


Figure 11: Fine-Tuning Models for QG and QA

3.2.2 Class Diagram

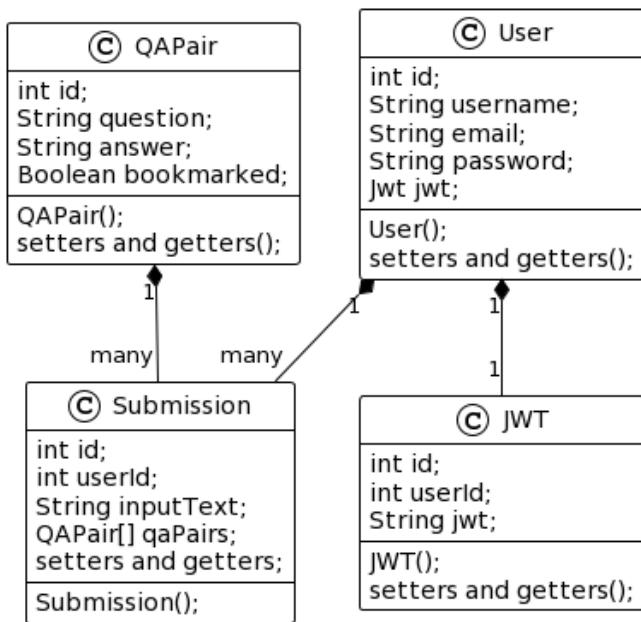


Figure 12: Class Diagram

3.2.3 Activity Diagrams

3.2.3.1 Register Activity Diagram

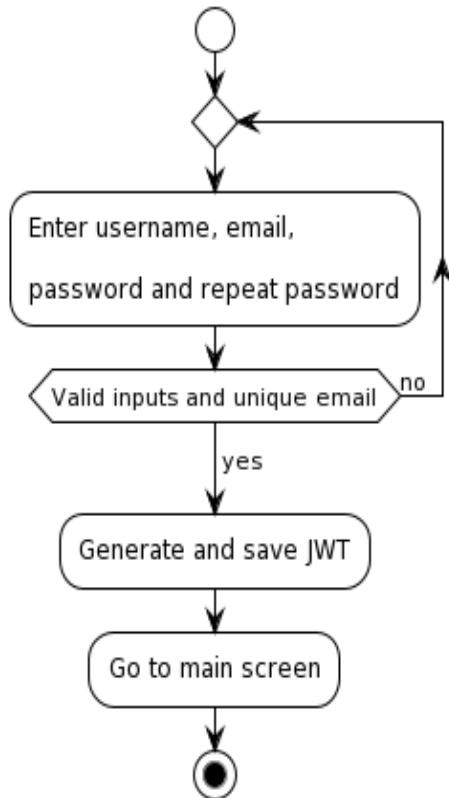


Figure 133: Register Activity Diagram

3.2.3.2 Login Activity Diagram

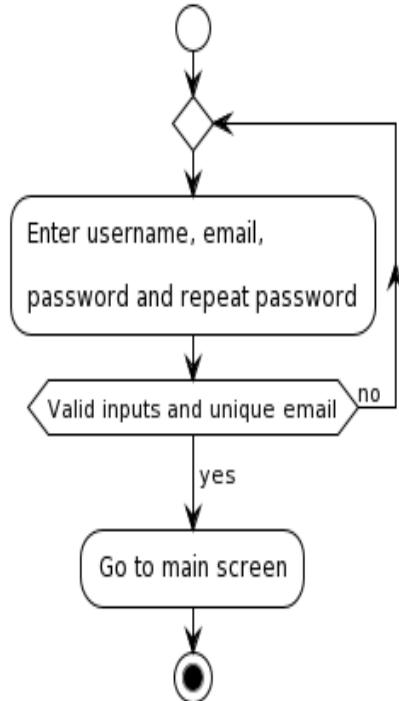


Figure 14: Login Activity Diagram

3.2.3.3 QA Generation Activity Diagram

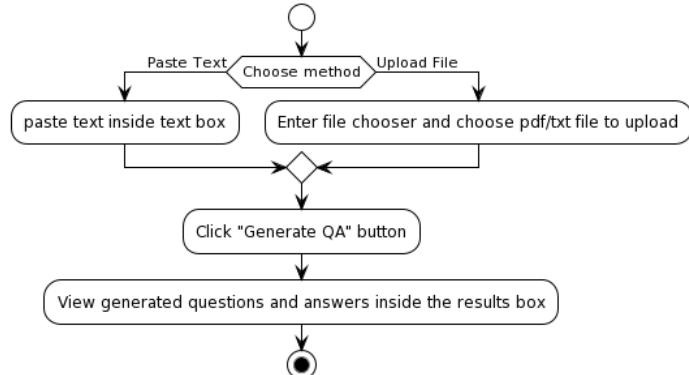


Figure 14: QA Generation Activity Diagram

3.2.3.4 View History Activity Diagram

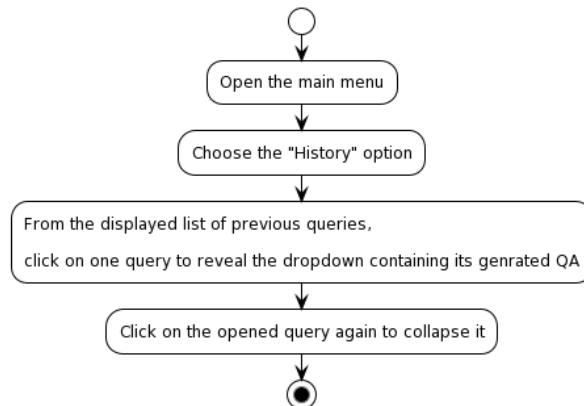


Figure 15: View History Activity Diagram

3.2.4 Use Case Diagram

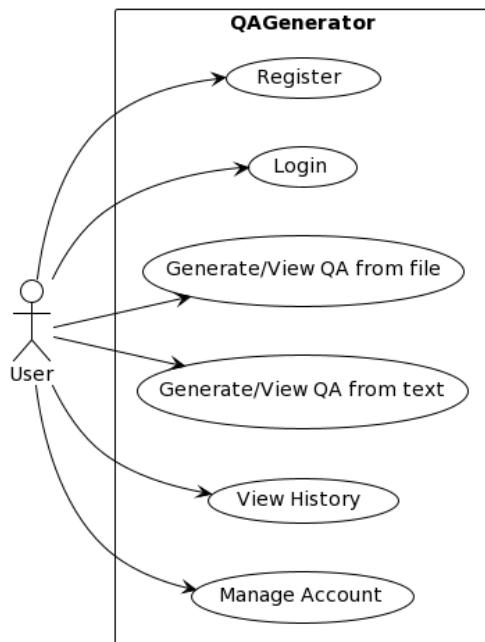


Figure 16: Use Case Diagram

- **Register:** the user registers a new account in order to access all functionalities.
- **Login:** the user logs into their existing account in order to access all functionalities.
- **Generate/View QA from file:** the user uploads a text file (.txt/.pdf) and clicks on the generate button to generate questions and answers for the given document.
- **Generate/View QA from text:** the user enters their input text in the text box and clicks on the generate button to generate questions and answers for the given document.
- **View history:** the user can view their history of submission, with each entry containing the original inputted text, along with the generated questions and answers.
- **Manage Account:** the user can manage their account, including viewing their information, as well as changing their name and password.

3.2.5 Sequence Diagrams

3.2.5.1 Register Sequence Diagram

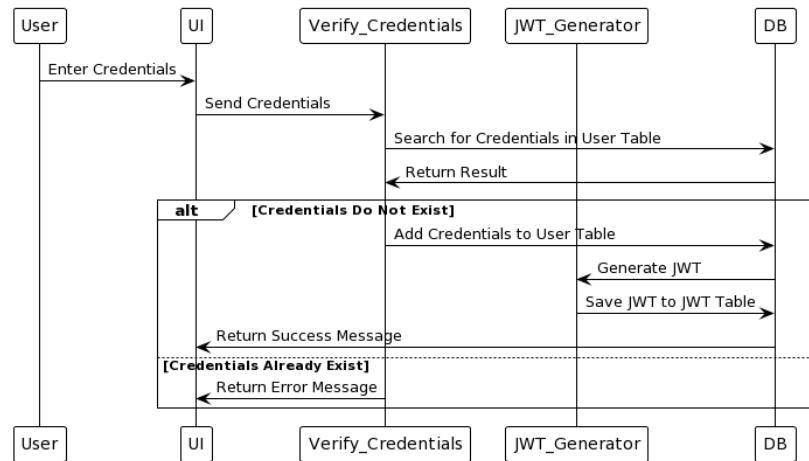


Figure 178: Register Sequence Diagram

3.2.5.2 Login Sequence Diagram

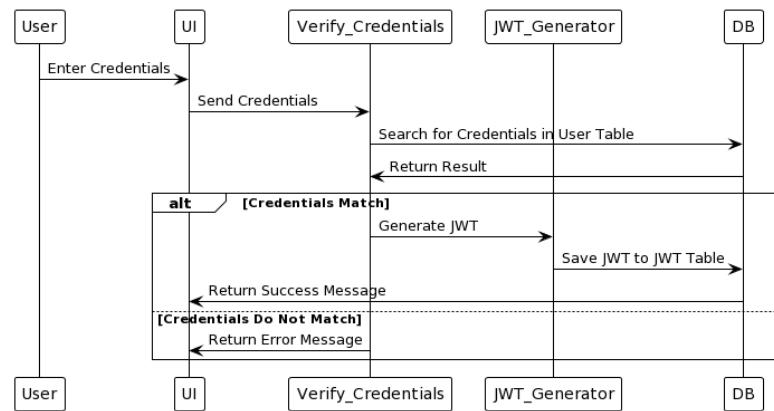


Figure 189: Login Sequence Diagram

3.2.5.3 QA Generation Sequence Diagram

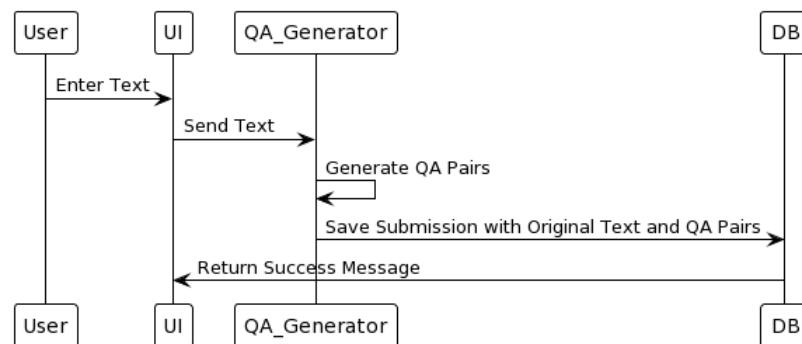


Figure 19: QA Generation Sequence Diagram

3.2.6 Deployment Diagram

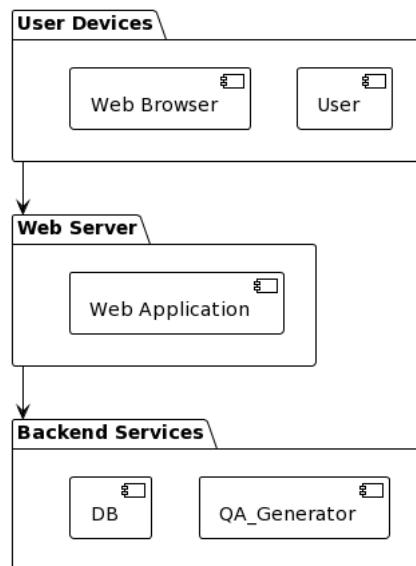


Figure 20: Deployment Diagram

3.3 Interface Demo

3.3.1 Registration Page

The screenshot shows a registration form titled "إنشاء حساب" (Create Account) with the subtitle "للوصول إلى جميع المميزات" (To access all features). The form includes fields for "النسم" (Name), "البريد الإلكتروني" (Email), "كلمة المرور" (Password), and a checkbox for "أوافق على الشروط والأحكام" (I agree to the terms and conditions). A red "إنشاء حساب" (Create Account) button is at the bottom. Below the form, a link "لديك حساب؟ تسجيل الدخول" (Already have an account? Log in) is visible.

Figure 22: Registration Page Demo

3.3.2 Login Page

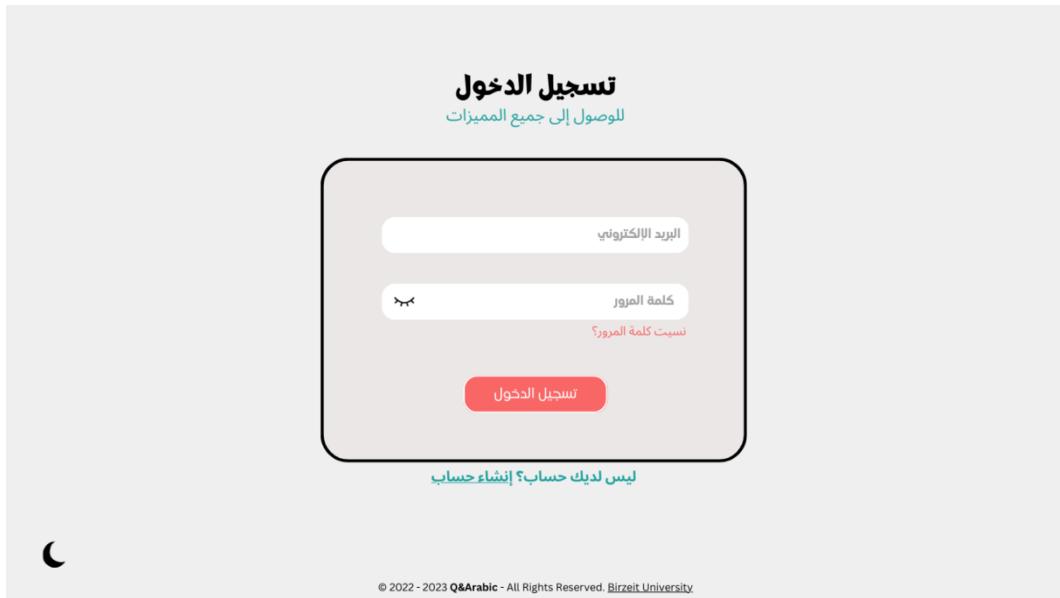


Figure 23: Login Page (Light Mood) Demo

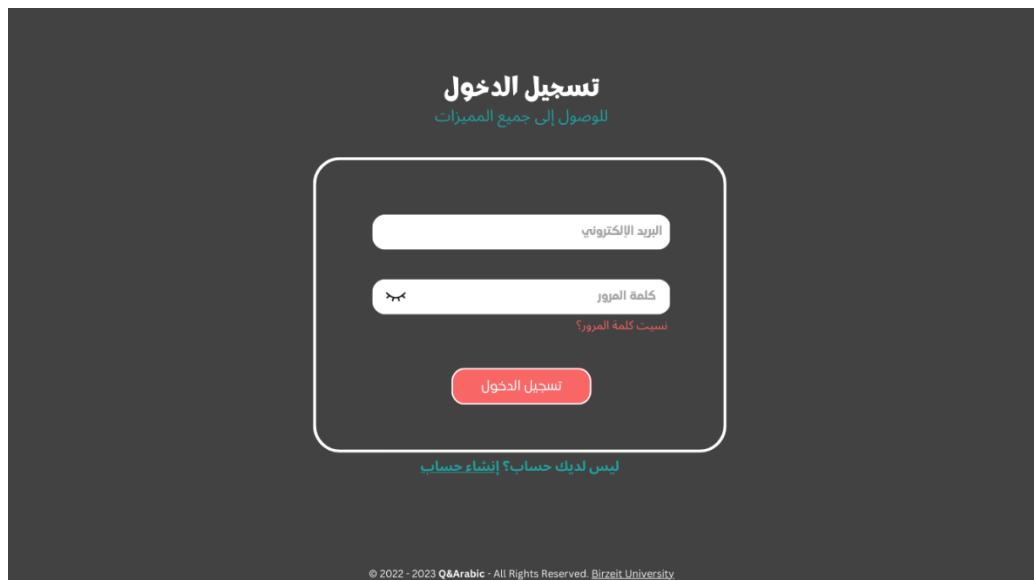


Figure 24: Login Page (Night Mood) Demo

3.3.3 Main Page (Text Input and Q/A generation)

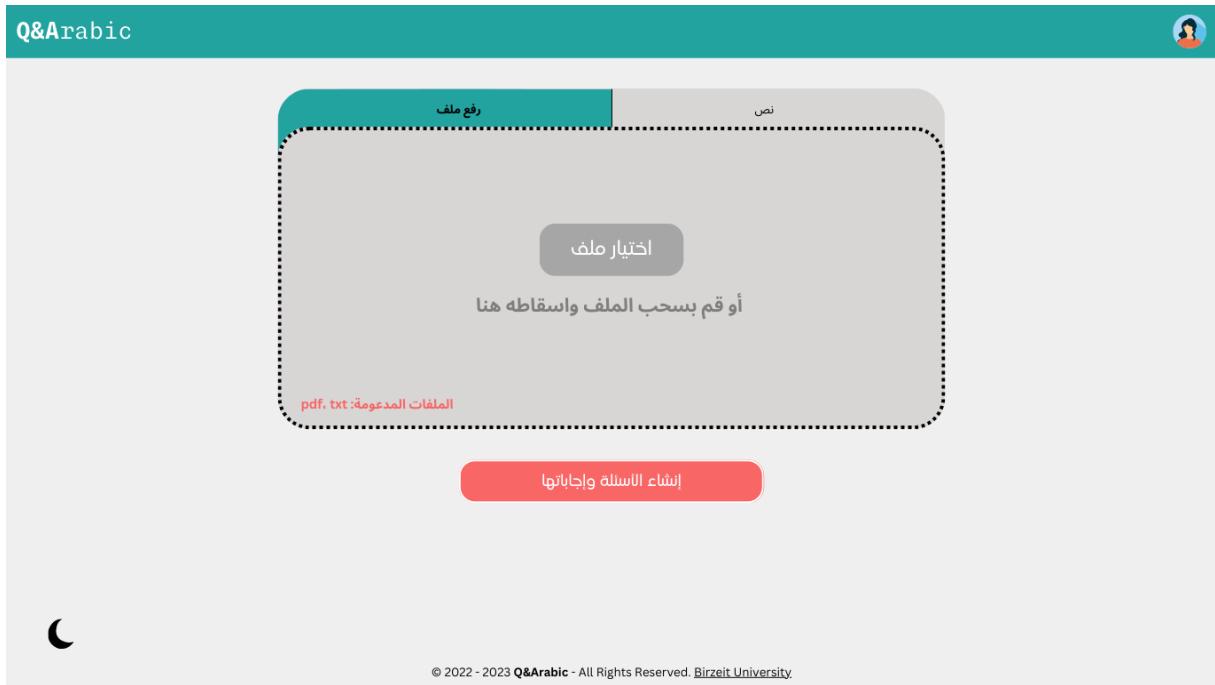


Figure 21: Main Page (File Upload) Demo

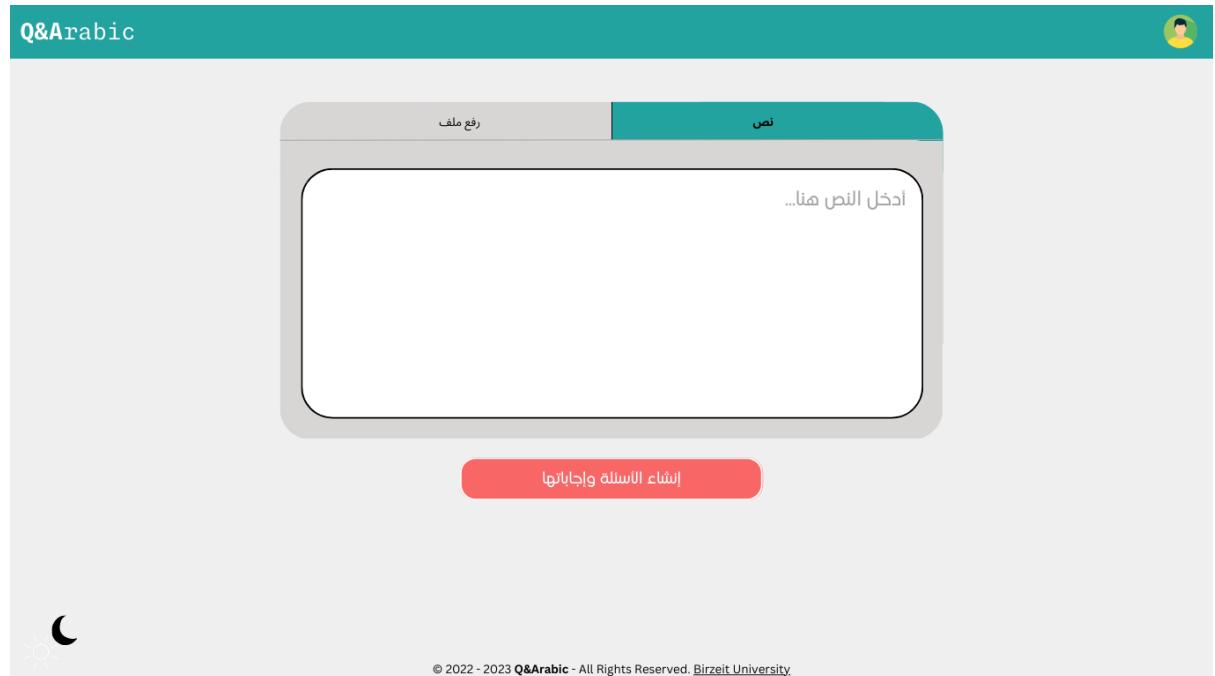


Figure 22: Main Page (Text Input) Demo



Figure 23: Main Page (With Options Pop-up Demo)

This screenshot shows the Q&Arabic main page with generated Q/A pairs. The top navigation and sidebar are identical to Figure 23. The main content area displays the following Q/A pairs:

- 1. عن مَاذا تعلن جامعة بيرزيت؟**
تعلن جامعة بيرزيت عن البدء باستقبال طلبات الالتحاق ببرنامج الدكتوراه في علم الحاسوب للفصل الدراسي الأول 2023-2024. ابتداء من يوم الثلاثاء الموافق 7/2/2023 لحين استيفاء العدد المطلوب (5 طلاب) أو لموعد أقصاه يوم الخميس الموافق 15/6/2023. علماً بأن رسوم الساعة المعتمدة لهذا البرنامج هو 250 دينار أردني.
يمكن لحاملي درجة الماجستير في علم الحاسوب أو الحقوق المعرفية المتصلة بها التقدم بطلب التحاق للبرنامج، على أن تكون درجة الماجستير قد تم الحصول عليها على أساس الدوام المنتظم من جامعة معترف بها بتقدير "جيد جداً" (أو ما يعاده...).
- 2. متى ستبدأ الجامعة في استقبال الطلبات وإلى متى؟**
سوف تبدأ الجامعة باستقبال طلبات الالتحاق ببرنامج الدكتوراه في علم الحاسوب ابتداء من يوم الثلاثاء الموافق 7/2/2023 لحين استيفاء العدد المطلوب (5 طلاب) أو لموعد أقصاه يوم الخميس الموافق 15/6/2023.
- 3. كم تبلغ رسوم الساعة المعتمدة لبرنامج الدكتوراه في علم الحاسوب؟**
تبلغ رسوم الساعة المعتمدة لبرنامج الدكتوراه في علم الحاسوب 250 دينار أردني.
- 4. ما هي شروط التقدم بطلب الالتحاق ببرنامج الدكتوراه في علم الحاسوب؟**
يشترط أن يكون المتقدم من حملة الماجستير في علم الحاسوب أو الحقوق المعرفية المتصلة بها، على...

In the bottom right corner of the main content area, there is a small crescent moon icon.

Figure 24: Main Page (With Q/A Generated) Demo

3.3.4 History Page

The screenshot shows a mobile application interface for 'Q&Arabic'. At the top, there is a teal header bar with the 'Q&Arabic' logo on the left and a user profile icon on the right. Below the header, the word 'السجل' (The Record) is displayed in large, bold black font. Underneath, there are three horizontal cards, each representing a historical entry:

- Entry 1:** 'تعلن جامعة بيرزيت عن البدء باستقبال طلبات الالتحاق ببرنامج الدكتوراة في' (Birzeit University announces the start of applications for the PhD program in ...). It includes icons for a document and a bookmark, and the text 'عدد الأسئلة: 4' (Number of questions: 4).
- Entry 2:** 'ساهمت الجامعة في تهيئة الطلبة ليكونوا مواطنين فاعلين في مجتمعهم، ولعبت' (The university contributed to preparing students to be active citizens in their community, and played a role in ...). It includes icons for a document and a bookmark, and the text 'عدد الأسئلة: 8' (Number of questions: 8).
- Entry 3:** A download link labeled 'birzeit-uni.pdf' with a document icon, indicating 22 pages. It includes icons for a document and a bookmark, and the text 'عدد الأسئلة: 22' (Number of questions: 22).

At the bottom of the screen, there is a small crescent moon icon on the left and a copyright notice on the right: '© 2022 - 2023 Q&Arabic - All Rights Reserved. Birzeit University'.

Figure 25: History Page Demo

Chapter 4: Conclusion and Future works

4.1 Review of the project

In this work, we presented a cutting-edge solution for Question Generation and Answering in the Arabic language. By leveraging the latest NLP technology, our proposed system aims to create a powerful and efficient platform that can generate Arabic questions and provide accurate answers. The system will make use of existing Arabic T5 models for question generation and Arabic BERT models for question answering, and will be implemented as a web-based application using Python for NLP tasks, Python Django for the backend, ReactJS for the frontend, and a MySQL database for data storage.

Furthermore, we conducted a comprehensive review of the latest language transformer models, including GPT and its Arabic counterparts, as well as the advancements in T5 and BERT models for the Arabic language. We also reviewed relevant research papers on the topics of question generation, question answering, and their combination to gain a better understanding of the current state of the art in NLP-based question generation and answering systems for the Arabic language. Also, a similar related application introduced.

Our report also includes a detailed System Analysis of the proposed Question Generation and Answering system, outlining its design and functionality. This analysis helped us gain a deeper understanding of the system's capabilities and limitations, as well as the steps required for its successful implementation. A sample mockup of the web interface was also introduced, showcasing the user-friendly experience that the system will offer.

In conclusion, this report presents a comprehensive proposal for developing a Question Generation and Answering system for the Arabic language, combining the latest advancements in NLP technology with a user-friendly interface. Our solution aims to facilitate better communication and knowledge transfer in the Arabic language, filling a gap in the current state of the art where relevant past research has only focused on either question generation or question answering.

4.2 Future Works

- Mobile app version: we are considering developing a mobile app version of the system in the future to get more people to use the system.
- Feedback system: a feedback system that allows users to provide feedback on the accuracy and relevance of the generated questions and answers. This feedback can be used to improve the machine learning models and accuracy of the questions and answers being generated later.

References

1. Liddy, Elizabeth D. "Natural language processing." (2001).
2. Winston, Patrick Henry. Artificial intelligence. Adison-Wesley Longman Publishing Co., Inc., 1984.
3. Jusoh, Shaidah. "A STUDY ON NLP APPLICATIONS AND AMBIGUITY PROBLEMS." *Journal of Theoretical & Applied Information Technology* 96, no. 6 (2018).
4. Mitchell, Tom M. Machine learning. Vol. 1, no. 9. New York: McGraw-hill, 1997.
5. Davis, Randall, and Jonathan J. King. "The origin of rule-based systems in AI." Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project (1984).
6. Goertzel, Ben. Artificial general intelligence. Edited by Cassio Pennachin. Vol. 2. New York: Springer, 2007.
7. Gill, Karamjit S. "Artificial super intelligence: beyond rhetoric." *Ai & Society* 31, no. 2 (2016): 137-143.
8. Ayodele, Taiwo Oladipupo. "Types of machine learning algorithms." *New advances in machine learning* 3 (2010): 19-48.
9. Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." *Emerging artificial intelligence applications in computer engineering* 160, no. 1 (2007): 3-24.
10. Gentleman, Robert, and Vincent J. Carey. "Unsupervised machine learning." In *Bioconductor case studies*, pp. 137-157. Springer, New York, NY, 2008.
11. Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement learning: A survey." *Journal of artificial intelligence research* 4 (1996): 237-285.
12. Shinde, Pramila P., and Seema Shah. "A review of machine learning and deep learning applications." In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, pp. 1-6. IEEE, 2018.
13. Hirschberg, Julia, and Christopher D. Manning. "Advances in natural language processing." *Science* 349, no. 6245 (2015): 261-266.
14. Munot, Nikita, and Sharvari S. Govilkar. "Comparative study of text summarization methods." *International Journal of Computer Applications* 102, no. 12 (2014).
15. Brants, Thorsten. "Natural Language Processing in Information Retrieval." *CLIN* 111 (2003).
16. Feldman, Ronen. "Techniques and applications for sentiment analysis." *Communications of the ACM* 56, no. 4 (2013): 82-89.
17. Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal* 5, no. 4 (2014): 1093-1113.
18. Pundge, Ajitkumar M., S. A. Khillare, and C. Namrata Mahender. "Question answering system, approaches and techniques: a review." *International Journal of Computer Applications* 141, no. 3 (2016): 0975-8887.
19. Du, Xinya, Junru Shao, and Claire Cardie. "Learning to ask: Neural question generation for reading comprehension." *arXiv preprint arXiv:1705.00106* (2017).
20. Kalady, Saidalavi, Ajeesh Elikkottil, and Rajarshi Das. "Natural language question generation using syntax and keywords." In *Proceedings of QG2010: The Third Workshop on Question Generation*, vol. 2, pp. 5-14. questiongeneration. org, 2010.
21. Fabbri, Alexander R., Patrick Ng, Zhiguo Wang, Ramesh Nallapati, and Bing Xiang. "Template-based question generation from retrieved sentences for improved unsupervised question answering." *arXiv preprint arXiv:2004.11892* (2020).
22. Tang, Duyu, Nan Duan, Tao Qin, Zhao Yan, and Ming Zhou. "Question answering and question generation as dual tasks." *arXiv preprint arXiv:1706.02027* (2017).
23. Bebis, George, and Michael Georgopoulos. "Feed-forward neural networks." *Ieee Potentials* 13, no. 4 (1994): 27-31.

24. Medsker, Larry R., and L. C. Jain. "Recurrent neural networks." *Design and Applications* 5 (2001): 64-67.
25. Wen, Tsung-Hsien, and Steve Young. "Recurrent neural network language generation for spoken dialogue systems." *Computer Speech & Language* 63 (2020): 101017.
26. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
27. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
28. Antoun, Wissam, Fady Baly, and Hazem Hajj. "Arabert: Transformer-based model for arabic language understanding." *arXiv preprint arXiv:2003.00104* (2020).
29. Safaya, Ali, Moutasem Abdullatif, and Deniz Yuret. "Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media." In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pp. 2054-2059. 2020.
30. Lan, Wuwei, Yang Chen, Wei Xu, and Alan Ritter. "An empirical study of pre-trained transformers for Arabic information extraction." *arXiv preprint arXiv:2004.14519* (2020).
31. Abdul-Mageed, Muhammad, Chiyu Zhang, AbdelRahim Elmadany, and Lyle Ungar. "Toward micro-dialect identification in diaglossic and code-switched environments." *arXiv preprint arXiv:2010.04900* (2020).
32. Chowdhury, Shammur Absar, Ahmed Abdelali, Kareem Darwish, Jung Soon-Gyo, Joni Salminen, and Bernard J. Jansen. "Improving Arabic text categorization using transformer training diversification." In *Proceedings of the fifth arabic natural language processing workshop*, pp. 226-236. 2020.
33. Inoue, Go, Bashar Alhafni, Nurpeiiis Baimukan, Houda Bouamor, and Nizar Habash. "The interplay of variant, size, and task type in Arabic pre-trained language models." *arXiv preprint arXiv:2103.06678* (2021).
34. Ghaddar, Abbas, Yimeng Wu, Ahmad Rashid, Khalil Bibi, Mehdi Rezagholizadeh, Chao Xing, Yasheng Wang et al. "Jaber and saber: Junior and senior arabic bert." *arXiv e-prints* (2021): arXiv-2112.
35. Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. "Improving language understanding by generative pre-training." (2018).
36. Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language models are unsupervised multitask learners." *OpenAI blog* 1, no. 8 (2019): 9.
37. Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.
38. Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang et al. "Training language models to follow instructions with human feedback." *arXiv preprint arXiv:2203.02155* (2022).
39. Thompson, A. D. (2022). GPT-3.5 + ChatGPT: An illustrated overview: An Independent Report by Dr Alan D. Thompson <https://LifeArchitect.ai/chatgpt>
40. Antoun, Wissam, Fady Baly, and Hazem Hajj. "Aragpt2: Pre-trained transformer for arabic language generation." *arXiv preprint arXiv:2012.15520* (2020).
41. Nagoudi, El Moatez Billah, Muhammad Abdul-Mageed, AbdelRahim Elmadany, Alcides Alcoba Inciarte, and Md Tawkat Islam Khondaker. "JASMINE: Arabic GPT Models for Few-Shot Learning." *arXiv preprint arXiv:2212.10755* (2022).
42. Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. "Exploring the limits of transfer learning with a unified text-to-text transformer." *The Journal of Machine Learning Research* 21, no. 1 (2020): 5485-5551.

43. Nagoudi, El Moatez Billah, AbdelRahim Elmadany, and Muhammad Abdul-Mageed. "Arat5: Text-to-text transformers for arabic language generation." arXiv preprint arXiv:2109.12068 (2021).
44. Klein, Tassilo, and Moin Nabi. "Learning to answer by learning to ask: Getting the best of gpt-2 and bert worlds." arXiv preprint arXiv:1911.02365 (2019).
45. Lukovnikov, Denis, Asja Fischer, and Jens Lehmann. "Pretrained transformers for simple question answering over knowledge graphs." In *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I* 18, pp. 470-486. Springer International Publishing, 2019.
46. Ngai, Hillary, Yoona Park, John Chen, and Mahboobeh Parsapoor. "Transformer-based models for question answering on COVID19." arXiv preprint arXiv:2101.11432 (2021).
47. Srivastava, Pragya, Tanuja Ganu, and Saikat Guha. "Towards Zero-Shot and Few-Shot Table Question Answering using GPT-3." arXiv preprint arXiv:2210.17284 (2022).
48. Rajpurkar, Pranav, Robin Jia, and Percy Liang. "Know what you don't know: Unanswerable questions for SQuAD." arXiv preprint arXiv:1806.03822 (2018).
49. Mozannar, Hussein, Karl El Hajal, Elie Maamary, and Hazem Hajj. "Neural Arabic question answering." arXiv preprint arXiv:1906.05394 (2019).
50. Clark, Jonathan H., Eunsol Choi, Michael Collins, Dan Garrette, Tom Kwiatkowski, Vitaly Nikolaev, and Jennimaria Palomaki. "TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages." *Transactions of the Association for Computational Linguistics* 8 (2020): 454-470.
51. Lewis, Patrick, Barlas Oğuz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. "MLQA: Evaluating cross-lingual extractive question answering." arXiv preprint arXiv:1910.07475 (2019).
52. Atef, Adel, Bassam Mattar, Sandra Sherif, Eman Elrefai, and Marwan Torki. "AQAD: 17,000+ arabic questions for machine comprehension of text." In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1-6. IEEE, 2020.
53. Nguyen, Tri, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. "MS MARCO: A human generated machine reading comprehension dataset." choice 2640 (2016): 660.
54. Joshi, Mandar, Eunsol Choi, Daniel S. Weld, and Luke Zettlemoyer. "Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension." *arXiv preprint arXiv:1705.03551* (2017).
55. Reddy, Siva, Danqi Chen, and Christopher D. Manning. "Coqa: A conversational question answering challenge." *Transactions of the Association for Computational Linguistics* 7 (2019): 249-266.
56. Bordes, Antoine, Nicolas Usunier, Sumit Chopra, and Jason Weston. "Large-scale simple question answering with memory networks." *arXiv preprint arXiv:1506.02075* (2015).
57. Webster, Jonathan J., and Chunyu Kit. "Tokenization as the initial phase in NLP." In *COLING 1992 volume 4: The 14th international conference on computational linguistics*. 1992.
58. Jivani, Anjali Ganesh. "A comparative study of stemming algorithms." *Int. J. Comp. Tech. Appl.* 2, no. 6 (2011): 1930-1938.
59. Voutilainen, Atro. "Part-of-speech tagging." *The Oxford handbook of computational linguistics* (2003): 219-232.
60. Sidorov, Grigori, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. "Syntactic n-grams as machine learning features for natural language processing." *Expert Systems with Applications* 41, no. 3 (2014): 853-860.
61. Goddard, Cliff, and Andrea C. Schalley. "Semantic Analysis." (2010): 93-120.
62. Das, Rubel, Antariksha Ray, Souvik Mondal, and Dipankar Das. "A rule based question generation framework to deal with simple and complex sentences." In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 542-548. IEEE, 2016.
63. Alhashedi, Saleh, Norhaida Mohd Suaib, and Aryati Bakri. Arabic Automatic Question Generation Using Transformer Model. No. 8588. EasyChair, 2022.

64. Rothe, Sascha, Shashi Narayan, and Aliaksei Severyn. "Leveraging pre-trained checkpoints for sequence generation tasks." *Transactions of the Association for Computational Linguistics* 8 (2020): 264-280.
65. Mihalcea, Rada, and Paul Tarau. "Textrank: Bringing order into text." In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 404-411. 2004.
66. Bousmaha, Kheira Zineb, Nour H. Chergui, Mahfoud Sid Ali Mbarek, and Lamia Hadrich Belguith. "AQG: Arabic Question Generator." *Rev. d'Intelligence Artif.* 34, no. 6 (2020): 721-729.
67. Paul, Kingsbury, and Palmer Martha. "From treebank to propbank." In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*. 2002.
68. Pasha, Arfath, Mohamed Al-Badrashiny, Mona Diab, Ahmed El Kholy, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow, and Ryan Roth. "Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic." In *Proceedings of the ninth international conference on language resources and evaluation (LREC'14)*, pp. 1094-1101. 2014.
69. Belguith, Lamia Hadrich, Leila Baccour, and Mourad Ghassan. "Segmentation de textes arabes basée sur l'analyse contextuelle des signes de ponctuations et de certaines particules." In *Actes de la 12ème conférence sur le Traitement Automatique des Langues Naturelles. Articles courts*, pp. 451-456. 2005.
70. Vargas-Vera, Maria, and Miltiadis D. Lytras. "Aqua: A closed-domain question answering system." *Information systems management* 27, no. 3 (2010): 217-225.
71. Alsubhi, Kholoud, Amani Jamal, and Areej Alhothali. "Deep learning-based approach for Arabic open domain question answering." *PeerJ Computer Science* 8 (2022): e952.
72. Benajiba, Yassine, and Paolo Rosso. "Arabic question answering." Diploma of advanced studies. Technical University of Valencia, Spain (2007).
73. Hammo, Bassam, Hani Abu-Salem, Steven L. Lytinen, and Martha Evens. "QARAB: A Question answering system to support the Arabic language." In *Proceedings of the ACL-02 workshop on Computational approaches to semitic languages*. 2002.
74. Mohammed, F. A., Khaled Nasser, and Hany M. Harb. "A knowledge based Arabic question answering system (AQAS)." *ACM SIGART Bulletin* 4, no. 4 (1993): 21-30.
75. Abouenour, Lahsen, Karim Bouzoubaa, and Paolo Rosso. "Improving Q/A using Arabic wordnet." In *Proc. The 2008 International Arab Conference on Information Technology (ACIT'2008)*, Tunisia, December. 2008.
76. Bhogal, Jagdev, Andrew MacFarlane, and Peter Smith. "A review of ontology based query expansion." *Information processing & management* 43, no. 4 (2007): 866-886.
77. Elkateb, Sabri, William J. Black, Piek Vossen, David Farwell, Horacio Rodríguez, Adam Pease, Musa Alkhalfa, and Christiane Fellbaum. "Arabic WordNet and the challenges of Arabic." In *Proceedings of the International Conference on the Challenge of Arabic for NLP/MT*. 2006.

INTRO GRADUATION REPORT

ORIGINALITY REPORT



PRIMARY SOURCES

| | | |
|---|--|------------|
| 1 | medium.com Internet Source | 1 % |
| 2 | easychair.org Internet Source | 1 % |
| 3 | www.researchgate.net Internet Source | 1 % |
| 4 | paperswithcode.com Internet Source | 1 % |
| 5 | aclanthology.org Internet Source | 1 % |
| 6 | Kheira Z. Bousmaha, Nour H. Chergui, Mahfoud Sid Ali Mbarek, Lamia Belguith Hadrich. "AQG: Arabic Question Generator", <i>Revue d'Intelligence Artificielle</i> , 2020 Publication | <1 % |
| 7 | www.mlq.ai Internet Source | <1 % |
| 8 | gist.github.com Internet Source | <1 % |

- 9 www.getz.pro [Internet Source](#) <1 %
- 10 nursingsolved.com [Internet Source](#) <1 %
- 11 www.arxiv-vanity.com [Internet Source](#) <1 %
- 12 F. A. Mohammed, Khaled Nasser, H. M. Harb.
"A knowledge based Arabic question
answering system (AQAS)", ACM SIGART
Bulletin, 1993 [Publication](#) <1 %
- 13 Xinyi Wang, Zhenye Gong, Guoxin Wang,
Jingdan Jia, Ying Xu, Jialu Zhao, Qingye Fan,
Shaun Wu, Weiguo Hu, Xiaoyang Li. "ChatGPT
Performs on the Chinese National Medical
Licensing Examination", Research Square
Platform LLC, 2023 [Publication](#) <1 %
- 14 www.datasetlist.com [Internet Source](#) <1 %
- 15 etd.uwc.ac.za [Internet Source](#) <1 %
- 16 ro.ecu.edu.au [Internet Source](#) <1 %
- 17 "ECAI 2020", IOS Press, 2020 [Publication](#) <1 %

| | | |
|-------------|---|------|
| 18 | "XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019", Springer Science and Business Media LLC, 2020 | <1 % |
| Publication | | |
| 19 | mytrickynotes.wordpress.com | <1 % |
| | Internet Source | |
| 20 | www.theses.fr | <1 % |
| | Internet Source | |
| 21 | dad.uni-bielefeld.de | <1 % |
| | Internet Source | |
| 22 | link.springer.com | <1 % |
| | Internet Source | |
| 23 | doaj.org | <1 % |
| | Internet Source | |
| 24 | repository.usahidsolo.ac.id | <1 % |
| | Internet Source | |
| 25 | assets.researchsquare.com | <1 % |
| | Internet Source | |
| 26 | github.com | <1 % |
| | Internet Source | |
| 27 | www.aclweb.org | <1 % |
| | Internet Source | |
| 28 | www.blogs.zappyhire.com | <1 % |
| | Internet Source | |

| | | |
|----|---|------|
| 29 | jurnal.unidha.ac.id Internet Source | <1 % |
| 30 | ethesisarchive.library.tu.ac.th Internet Source | <1 % |
| 31 | openresearch-repository.anu.edu.au Internet Source | <1 % |
| 32 | thesai.org Internet Source | <1 % |
| 33 | finance.dailyherald.com Internet Source | <1 % |
| 34 | personales.upv.es Internet Source | <1 % |
| 35 | nexttech.edu.au Internet Source | <1 % |
| 36 | Nanda Diaz Arizona, Yulia Yulia, Oktavianus Jaka. "Penerapan Metode ASD (Adaptive Software Development) Dalam Pembuatan Sistem Informasi Akuntansi Kas Masuk Dan Kas Keluar Pada CV.Bela Usaha", Jurnal Sistem Informasi Akuntansi, 2022 Publication | <1 % |
| 37 | www.blopig.com Internet Source | <1 % |
| 38 | "Intelligent Systems and Applications", Springer Science and Business Media LLC, | <1 % |

2022

Publication

- 39 R. Hariharan, M. Dhilsath Fathima, A. Kameshwaran, A. Bersika M. C. Sweety et al. "Chapter 23 Abstractive Summarization is Improved by Learning Via Semantic Similarity", Springer Science and Business Media LLC, 2023 <1 %
Publication
- 40 Sachin Patil. "My learning assistant", interactions, 1/1/2007 <1 %
Publication
- 41 anjana669.wordpress.com <1 %
Internet Source
- 42 www.microsoft.com <1 %
Internet Source
- 43 Navin Sabharwal, Amit Agrawal. "Hands-on Question Answering Systems with BERT", Springer Science and Business Media LLC, 2021 <1 %
Publication
- 44 Samer Albahra, Tom Gorbett, Scott Robertson, Giana D'Aleo et al. "Artificial Intelligence and Machine Learning Overview in Pathology & Laboratory Medicine: A General Review of Data Preprocessing and <1 %

Basic Supervised Concepts", Seminars in Diagnostic Pathology, 2023

Publication

-
- 45 asmen.icopy.site <1 %
Internet Source
-
- 46 "Advanced Computing", Springer Science and Business Media LLC, 2021 <1 %
Publication
-
- 47 "Formalising Natural Languages: Applications to Natural Language Processing and Digital Humanities", Springer Science and Business Media LLC, 2021 <1 %
Publication
-
- 48 "The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)", Springer Science and Business Media LLC, 2020 <1 %
Publication
-
- 49 Yongqi Li, Wenjie Li, Liqiang Nie. "Dynamic Graph Reasoning for Conversational Open-Domain Question Answering", ACM Transactions on Information Systems, 2022 <1 %
Publication
-
- 50 dev.knmi.nl <1 %
Internet Source
-
- 51 downloads.hindawi.com <1 %
Internet Source
-

- 52 hdl.handle.net <1 %
Internet Source
-
- 53 solveforce.com <1 %
Internet Source
-
- 54 sportold.ubbcluj.ro <1 %
Internet Source
-
- 55 Sarah Saad Alanazi, Nazar Elfadil, Mutsam Jarajreh, Saad Algarni. "Question Answering Systems: A Systematic Literature Review", International Journal of Advanced Computer Science and Applications, 2021 <1 %
Publication
-
- 56 bura.brunel.ac.uk <1 %
Internet Source
-
- 57 etda.libraries.psu.edu <1 %
Internet Source
-
- 58 jpt.spe.org <1 %
Internet Source
-
- 59 kth.diva-portal.org <1 %
Internet Source
-
- 60 www.rahular.com <1 %
Internet Source
-
- 61 Lei Sun, Xiwei Feng, Pengcheng Hua, Chi Zhao, Chaoqi Wang, Wei Hou. "Research on Intelligent Question Answering algorithm" <1 %

based on machine learning", Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering, 2021

Publication

- 62 Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, Jianfeng Gao. "Deep Learning--based Text Classification", ACM Computing Surveys, 2021
Publication <1 %
- 63 escholarship.org <1 %
Internet Source
- 64 library.oapen.org <1 %
Internet Source
- 65 nlp.fi.muni.cz <1 %
Internet Source
- 66 quintagroup.com <1 %
Internet Source
- 67 repository.nwu.ac.za <1 %
Internet Source
- 68 scholar.archive.org <1 %
Internet Source
- 69 science.vu.nl <1 %
Internet Source
- 70 "Text, Speech, and Dialogue", Springer Science and Business Media LLC, 2020 <1 %

Publication

- 71 Advances in Intelligent Systems and Computing, 2016. <1 %
Publication
- 72 Mark Last, Guy Danon. "Automatic question generation", WIREs Data Mining and Knowledge Discovery, 2020 <1 %
Publication
- 73 Nutthanit Wiwatbuttsiri, Atiwong Suchato, Proadpran Punyabukkana, Nuengwong Tuaycharoen. "Question Generation in the Thai Language Using MT5", 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2022 <1 %
Publication
- 74 Sane Mo Yagi, Youssef Mansour, Firuz Kamalov, Ashraf Elnagar. "Evaluation of Arabic-Based Contextualized Word Embedding Models", 2021 International Conference on Asian Language Processing (IALP), 2021 <1 %
Publication
- 75 Wissal Brini. "An Arabic question-answering system for factoid questions", 2009 International Conference on Natural Language Processing and Knowledge Engineering, 09/2009 <1 %

Publication

| | | |
|----|--|------|
| 76 | bonndoc.ulb.uni-bonn.de | <1 % |
| 77 | ccrpc.org | <1 % |
| 78 | dspace.fsktm.um.edu.my | <1 % |
| 79 | library.iugaza.edu.ps | <1 % |
| 80 | mafiadoc.com | <1 % |
| 81 | repository.tudelft.nl | <1 % |
| 82 | res.mdpi.com | <1 % |
| 83 | s3-eu-west-1.amazonaws.com | <1 % |
| 84 | scholarspace.manoa.hawaii.edu | <1 % |
| 85 | shodhgangotri.inflibnet.ac.in | <1 % |
| 86 | technodocbox.com | <1 % |
| 87 | www.jatit.org | |

| | | |
|----|---|------|
| | Internet Source | <1 % |
| 88 | www.jjcit.org Internet Source | <1 % |
| 89 | www.lrec-conf.org Internet Source | <1 % |
| 90 | www.ru.nl Internet Source | <1 % |
| 91 | www.theseus.fi Internet Source | <1 % |
| 92 | "Machine Learning and Data Science", Wiley, 2022 Publication | <1 % |
| 93 | "Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2018 Publication | <1 % |
| 94 | "The Semantic Web – ISWC 2019", Springer Science and Business Media LLC, 2019 Publication | <1 % |
| 95 | Adel Atef, Bassam Mattar, Sandra Sherif, Eman Elrefai, Marwan Torki. "AQAD: 17,000+ Arabic Questions for Machine Comprehension of Text", 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), 2020 | <1 % |

Publication

- 96 Adnan Masood, Adnan Hashmi. "Cognitive Computing Recipes", Springer Science and Business Media LLC, 2019 <1 %
Publication
- 97 Boudabous, Mohamed Mahdi, Lamia Hadrich Belguith, and Fatiha Sadat. "Exploiting the Arabic Wikipedia for semi-automatic construction of a lexical ontology", International Journal of Metadata Semantics and Ontologies, 2013. <1 %
Publication
- 98 Joseph Worsham, Jugal Kalita. "Multi-task learning for natural language processing in the 2020s: where are we going?", Pattern Recognition Letters, 2020 <1 %
Publication
- 99 Mariam M. Biltawi, Sara Tedmori, Arafat Awajan. "ARABIC QUESTION ANSWERING SYSTEMS: GAP ANALYSIS", IEEE Access, 2021 <1 %
Publication
- 100 Rasha M. Albalawi, Amani T. Jamal, Alaa O. Khadidos, Areej M. Alhothali. "Multimodal Arabic Rumors Detection", IEEE Access, 2023 <1 %
Publication
- 101 aclanthology.lst.uni-saarland.de <1 %
Internet Source

| | | |
|-----|--|------|
| 102 | chauff.github.io Internet Source | <1 % |
| 103 | core.ac.uk Internet Source | <1 % |
| 104 | direct.mit.edu Internet Source | <1 % |
| 105 | docplayer.net Internet Source | <1 % |
| 106 | dokumen.pub Internet Source | <1 % |
| 107 | doras.dcu.ie Internet Source | <1 % |
| 108 | educationdocbox.com Internet Source | <1 % |
| 109 | eprints.utar.edu.my Internet Source | <1 % |
| 110 | helda.helsinki.fi Internet Source | <1 % |
| 111 | ijirt.org Internet Source | <1 % |
| 112 | ikee.lib.auth.gr Internet Source | <1 % |
| 113 | instrepo-prod.cc.uregina.ca Internet Source | <1 % |

| | | |
|-----|--|------|
| 114 | koreascience.kr Internet Source | <1 % |
| 115 | louis.uah.edu Internet Source | <1 % |
| 116 | static.frontiersin.org Internet Source | <1 % |
| 117 | tel.archives-ouvertes.fr Internet Source | <1 % |
| 118 | uwspace.uwaterloo.ca Internet Source | <1 % |
| 119 | web.archive.org Internet Source | <1 % |
| 120 | www.bibsonomy.org Internet Source | <1 % |
| 121 | www.iosrjournals.org Internet Source | <1 % |
| 122 | www.mdpi.com Internet Source | <1 % |
| 123 | www.torontomachinelearning.com Internet Source | <1 % |
| 124 | "Text, Speech and Dialogue", Springer Science and Business Media LLC, 2014 Publication | <1 % |

- 125 "The Semantic Web – ISWC 2020", Springer Science and Business Media LLC, 2020 **<1 %**
Publication
-
- 126 Ali Bou Nassif, Ashraf Elnagar, Omar Elgendi, Yaman Afadar. "Arabic fake news detection based on deep contextualized embedding models", Neural Computing and Applications, 2022 **<1 %**
Publication
-
- 127 Andrew D. White, Glen M. Hocky, Heta A. Gandhi, Mehrad Ansari et al. "Assessment of chemistry knowledge in large language models that generate code", Digital Discovery, 2023 **<1 %**
Publication
-
- 128 arxiv.org **<1 %**
Internet Source
-
- 129 "Data Science", Springer Science and Business Media LLC, 2019 **<1 %**
Publication
-
- 130 Jie Dong, Dong-Sheng Cao, Hong-Yu Miao, Shao Liu et al. "ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation", Journal of Cheminformatics, 2015 **<1 %**
Publication
-
- 131 Lecture Notes in Computer Science, 2015.

| | | |
|-----|--|------|
| | Publication | <1 % |
| 132 | Mohammad Bani-Almarjeh, Mohamad-Bassam Kurdy. "Arabic abstractive text summarization using RNN-based and transformer-based architectures", Information Processing & Management, 2023 Publication | <1 % |
| 133 | microsoft.github.io Internet Source | <1 % |
| 134 | Mihai Alexandru Niculescu, Stefan Ruseti, Mihai Dascalu. "RoGPT2: Romanian GPT2 for Text Generation", 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), 2021 Publication | <1 % |

Exclude quotes

On

Exclude matches

Off

Exclude bibliography

On