N° d'ordre :
N° de série :

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

ECHAHID HAMMA LAKHDAR UNIVERSITY - EL OUED
FACULTY OF EXACT SCIENCES
Computer Science department

End of Study Memory
Presented for the Diploma of

ACADEMIC MASTER

Domain : **Mathematics and Computer Science**
spinneret: **Computer Science**
Speciality : **Artificial Intelligence and Distributed Systems**

Presented by :

- **Hanine Horra**
- **Wiam Ghilani**

Theme

Question Answering Transformer Model for Arabic Language

Supported in: 06-06- 2023 In front of jury:

M.	Sana Sahar Guia	President
N.	Abdel Kamel Ben Berika	Reporter
Dr.	Mohammed Anouar Naoui	Supervisor

Academic Year: 2022/2023

ABSTRACT

Arabic is the 6th most wide-spread natural language in the world with more than 350 million native speakers. The majority of text in Arabic data is unstructured and dispersed across the internet. This text data can yield helpful knowledge if it is properly obtained, aggregated, formatted, and analyzed.

In this work, we extracted and formatted data in Fiqh and Syrah from different resources to create QAFS (Question Answering in Fiqh and Syrah) dataset with 550 questions in the Arabic language. Which is the first of its kind. Later, we trained three BERT models family (BERT, DistilBERT, and ELECTRA) on QAFSv1 using simple transformers library. The evaluation metrics used were (correct, incorrect, and similar). Finally, we obtained relatively good results for BERT & DistilBERT, and ELECTRA.

KEYWORDS: *Transformer, Arabic Language, Question Answering, Dataset.*

الملخص

اللغة العربية هي سادس لغة طبيعية الأكثر انتشارا في العالم مع أكثر من 350 مليون ناطقا بها. غالبية النصوص في البيانات العربية غير منظمة وموزعة عبر الإنترنت. يمكن أن تسفر بيانات النص هذه عن معرفة مفيدة إذا تم الحصول عليها وتجميعها وتنسيقها وتحليلها بشكل صحيح.

في هذا العمل، قمنا باستخراج وتنسيق البيانات في الفقه والسيرة من مصادر مختلفة لإنشاء مجموعة بيانات (الإجابة على الأسئلة في الفقه والسيرة) مع 550 سؤالا باللغة العربية. وهو الأول من نوعه. بعد ذلك ، قمنا بتدريب ثلاثة نماذج من عائلة بيرت (بيرت ، ديستيلبرت ، و إلكترا) على مجموعة البيانات باستخدام مكتبة المحولات البسيطة. مقاييس التقييم المستخدمة كانت(صحيحة وغير صحيحة ومتشابهة) ،حيث كانت النتيجة المتحصلعليها من اجل نماذج بيرت، ديستيلبيرت والكترا جيدة نسبيا.

الكلمات المفتاحية: المحولات ، اللغة العربية ، الإجابة على الأسئلة ، مجموعة البيانات.

DEDICATION

We dedicate this modest work to whom brought us love, support and happiness , to the dearest people to our hearts, To our dear parents, for their; love, sacrifice, patience, moral and material support from our childhood until this day, To all our brothers and sisters , who supported us, encouraged us and who took the way with us along all the courses of our studies, To all our dear nephews, to our husbands, To our dear friends

Hanine & Wiam

KNOWLEDGMENT

“ Thanks to the one God, Light of the heavens and the earth, who help and guide ”

we must first of all thank “ ALLAH ” the almighty, who gave us the power, the will and the patience to develop this work .
Our sincerest thanks to our mentor Mr. Naoui Mohammed Anouar for his availability, his contributions, his valuable guidance and his understanding throughout the development of this memoir.
We would also like to thank the members of the jury very much for accepting, examine and evaluate this modest work .
We also thank all our teachers of Licence and Master in the department of Computing Science.

Hanine & Wiam

Table of Contents

List of Figures

INTRODUCTION

Object presentation

The digital world would have never evolved without Natural Language Processing (NLP) which would have been inefficient without artificial intelligence. NLP enables computers to perform a wide range of natural language-related tasks at all levels, ranging from parsing and part-of-speech (POS) tagging, to machine translation and question answering.

While Arabic language users constitute the fastest growing language group on the web with regard to the number of Internet users the Arabic question-answering systems are gaining great importance due to the increasing amounts of Arabic content on the Internet and the increasing demand for information that regular information retrieval techniques cannot satisfy.

The classic Recurrent Neural Network (RNN) and Lost Short-Term Memory (LSTM) were regarded as the basic architectures for translation. Not even six years back, LSTM-based architectures formed the base of every NLP system.

The transformer architecture was generated to move away from sequential processing. It was originally introduced for machine translation tasks but has spread into many other application areas such as question-answering and sentiment analyses.

In this work, we created a dataset for Arabic question answering extracted from diverse sources in Fiqh and Syrah. This dataset is composed of 550 questions, of which the training file contains 440 dictionaries and the test file contains 110 dictionaries. After that, we trained BERT, DistilBERT, and ELECTRA on the QAFSv1 dataset. For treating a dataset in Arabic question answering using a transformer we followed this work plan.

Work plan

This thesis contains three chapters as following :

- **Chapter One**

  This chapter presents an overview of NLP, components of NLP, and its applications finally the NLP pipeline.

- **Chapter Tow**

  This chapter serves as an overview of machine learning and Deep learning and its methods CNN & RNN. Finally an introduction to transformers and the attention mechanism.

- **Chapter Three**

  In this chapter we saw details about QAFSv1, the training steps of BERT, DistilBERT, and ELECTRA. Also discussion of the results.

Finally, we achieve this thesis by a general conclusion where we resume our main contribution and give some orientation for further work.

CHAPTER 1

INTRODUCTION TO NLP

Introduction

Natural Language Processing (NLP) is a subfield of computer science and artificial intelligence that deals with the interaction between computers and human languages. The primary goal of NLP is to enable computers to understand, interpret, natural language.

NLP is aimed to do this job effectively and with accuracy like a human. This chapter presents an overview of NLP, its applications, components of NLP, and finally the NLP pipeline.

1.1 Definition

Natural Language Processing(NLP) as shown in **Figure 1.1** refers to the branch of computer science and more specifically, the branch of artificial intelligence concerned with giving computers the ability to under- stand text and spoken words in much the same way human beings can[13],NLP combines computational linguistics - rule-based modeling of human language.Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, with the speaker or writer's intent and sentiment.

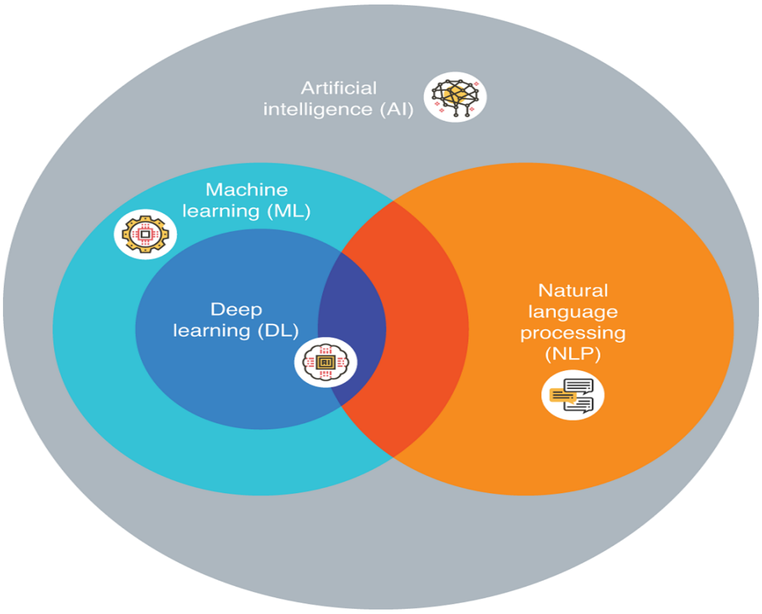

Figure 1.1: Natural Language Processing

1.2 History of NLP

The history of NLP generally starts in the 1950s. In 1950, Alan Turing published an article titled "Machine and Intelligence" which advertised what is now called the Turing test as a sub field of intelligence. Some beneficial and successful Natural language systems were developed in the 1960s were SHRDLU, a natural language system working in restricted "blocks worlds" with restricted vocabularies was written between 1964 to 1966[4].

1.3 Applications

The number of NLP real-world applications is increasing with advancements in the field of machine learning. Here is a list of the most popular NLP applications[13]:

- **Machine Translation:** Machine Translation for information access is probably one of the most common uses of NLP technology, Google Translate alone translates hundreds of billions of words a day between over 100 languages (e.g. Google Translate.

  "I want to talk the dialect of your people. It's no use of talking unless people understand what you say."

Zora Neale Hurston, Moses,Man of the Mountain 1939, p.121

- **Speech Recognition:** The task of recognizing human voice to take actions or to convert into text.
- **Sentiment Analysis:** The task of understanding the emotion in a text piece.
- **Question Answering (QA):** Is a classic application of Natural Language Processing. It is a process that understands user's natural language query and can provide a correct answer and extract it from retrieving relevant documents, data, or knowledge base.

  QA systems are broadly divided into two categories: open-domain (ODQA) system and closed-domain (CDQA) system[17] :

  - Closed-domain: In closed-domain systems, questions belong to a particular domain. They can answer the questions from a single domain only. As an example, a question answering system for the health care domain.
  - Open-domain: In open-domain systems, questions can be from any domain, such as health care, IT, sports, and more. These systems are designed to answer questions from any domain.

  To be able to develop these real-world applications, researchers have to use several NLP methods.

1.4 Steps of natural language processing

There are 5 phases involved in natural language processing:[4]

- **Morphological and Lexical Analysis :** The lexicon of a language is its vocabulary that includes its words and expressions. Morphology depicts analysing, identifying and description of structure of words. Lexical analysis is dividing the whole chunk of text into paragraphs, sentences, and words.

- **Syntactic Analysis :** This involves analysation of the words in a sentence to depict the grammatical structure of the sentence. The words are transformed into a structure that shows how the words are related to each other.

- **Semantic Analysis :** This abstracts the dictionary meaning or the exact meaning from context. The structures which are created by the syntactic analyser are assigned meaning. There is a mapping between the syntactic structures and the objects in the task domain.

- **Discourse Integration :** The meaning of any single sentence depends upon the sentences that precedes it and also invokes the meaning of the sentences that follow it.

- **Pragmatic Analysis :** It means abstracting the purposeful use of the language in situations, importantly those aspects of language which require world knowledge the main focus is what was said is reinterpreted on what it actually means.

1.5 Components of NLP

NLP can be classified into two parts i.e., Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text **Figure 1.2**.Presents the broad classification of NLP, the objective of this section is to discuss the Natural Language Understanding (Linguistic) (NLU) and the Natural Language Generation (NLG)[12].

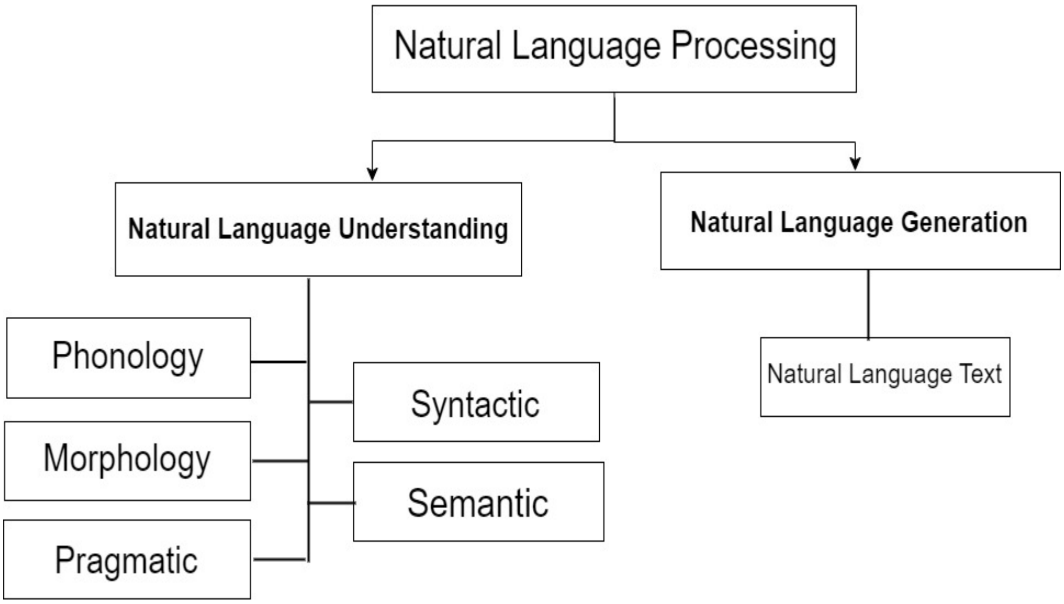

Figure 1.2: Broad classification of NLP

- Natural Language Understanding (NLU) which involves transforming human language into a machine-readable format.
- It helps the machine to understand and analyse human language by extracting the text from large data.
- Natural Language Generation (NLG) acts as a translator that converts the computerized data into natural language representation.
- It mainly involves Text planning, Sentence planning, and Text realization.
- The NLU is harder than NLG.

1.5.1 Natural Language Understanding (NLU)

The most difficult part of NLP is the understanding of Natural Language. The very first step is to convert the natural language into machine understandable language i.e. Binary language. It is the way in which Speech Recognition and Speech to Text systems work. As soon as the data arrives in the text format, the NLU process starts with an objective of taking out the meaning from the text[3].

- Phonology : It is the study of organizing sound systematically.

- Morphology : The study of the formation and internal structure of words.
- Syntactic : The study of the formation and internal structure of sentences.
- Semantic : The study of the meaning of sentences.
- Discourse : It deals with how the immediately preceding sentence can affect the interpretation of the next sentence.
- Pragmatic : It deals with using and understanding sentences in different situations and how the interpretation of the sentence is affected.
- World Knowledge : It includes the general knowledge about the world.

1.5.2 Natural Language Generation (NLG)

The artificial language obtained by the NLU step is then converted into text by the Natural Language Generator (NLG). The task of conversion of this text to audible speech is also done by NLG via text to speech. Various components of NLG are shown in **Figure1.3**[3].

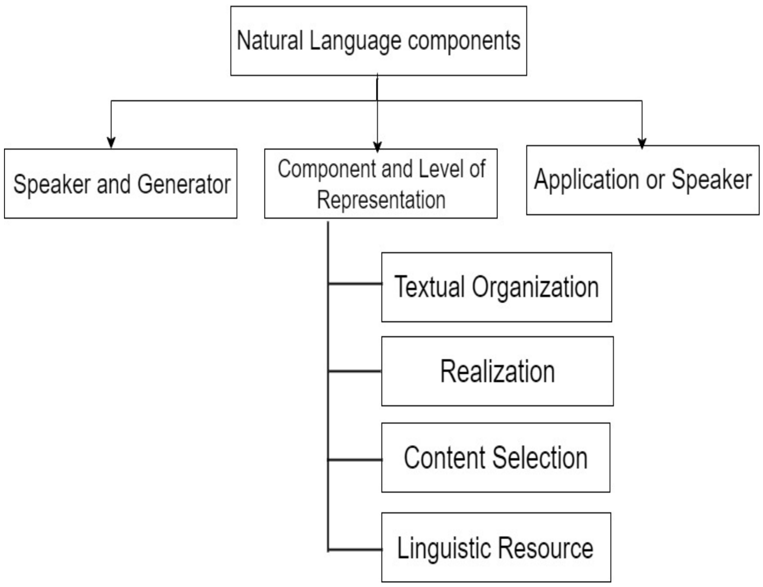

Figure 1.3: Natural Language Generation component

- **Speaker and Generator :** For the generation of text, a speaker and a generator program concentrate the application's objective into smooth phrases relevant to the state.

- **Components and Levels of Representation :** The language generation process includes the following interconnected tasks:
  - Content Selection: The gathered information is compiled into a set. It is then analyzed into representational units for the removal of some part of this unit and the addition of default parts.
  - Textual organization: The data is then arranged textually according to the grammar.
  - Linguistic Resources: Linguistic resources are chosen to support the information's realization. These resources come down to specific words, idioms etc. at the end.
  - Realization: The chosen and codified resources are realized into actual text.
- **Application or Speaker :** It is for supporting the model of the condition. This speaker only starts the process and does not take any initiation in the language generation.

1.6 Steps of NLP pipeline

There should be some way to teach a machine the basic concepts so that they can read, process, and understand text. The following are some of the steps of an NLP pipeline shown in **Figure1.4** [18].

Figure 1.4: NLP pipeline

1. **Sentence Segmentation:** The first step in the pipeline is to segment the text snippet into individual sentences.

The implementation for sentence segmentation was very easy, you just need to segment the text based on punctuation or 'full point'. Sometimes it fails, when documents or a piece of text was not formatted correctly or was grammatically incorrect.

There are some advanced NLP methods such as sequential learning which cuts off a part of the text even when there is no full point. Extract phrases by dividing the text using semantic comprehension along with grammatical comprehension.

2. **Tokenization:** The next task in the NLP pipeline is tokenization. In this task, we break each of the sentences into multiple tokens. A token can be a character, a word, or a phrase. The basic methodology used in tokenization is to split a sentence into separate words whenever there is a space between them.

   There are some advanced tokenization methods such as Markov chain models that can extract phrases out of a sentence.

3. **Parts of Speech Tagging:** POS tagging shown in **Figure 1.5** is the next step to determine parts of speech for each of the tokens or words extracted from the tokenization step. This helps us to identify the use of each word and its significance in a sentence. It also introduces first steps toward the actual understanding of the meaning of a sentence. Imparting a POS tag can increase the dimension of the word, to give better detail of the meaning the given word is trying to impart. as shown in **Figure 1.5**.

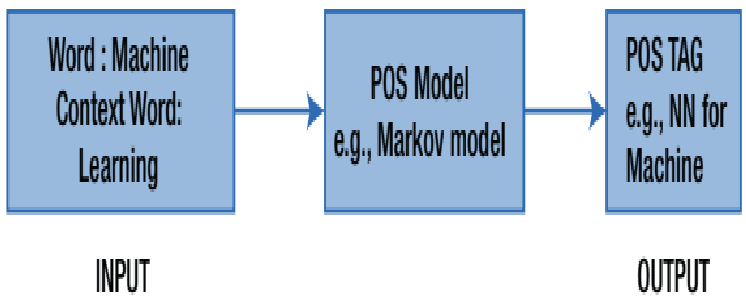

Figure 1.5: POS Tagging

These models are trained on a huge corpus of (millions or billions) sentences of literature in the target language where each word along with its POS tag is used as training data for the POS classifier.

4. **Stemming and Lemmatization :** Sometimes the same word occurs in multiple sentences in different forms. Stemming can be defined as the process of reducing words to their root or base form by removing suffixes. Here, the reduced words can be dictionary words or non dictionary words.

5. **Identification of Stop Words :** Text snippets contain important as well as filler words. For example, these are the filler words. [**"be", "use", "in", "a", "such", "a", "and"**] These words introduce noise into your text and it is important to manage them, as they appear very frequently in the text and will have a much higher frequency and less importance than other words.

    - Flag words as stop words on the basis of their frequency of occurrence.
    - Flag words as stop words if they are quite common across all documents.

1.7 Conclusion

This chapter dealt with Natural Language Processing which is an area of computer science and artificial intelligence that focuses on the interaction between computers and humans. NLP has already proven to be useful in multiple areas. Its various applications can help companies save time and improve efficiency .

CHAPTER 2

MACHINE LEARNING FOR NLP

Introduction

Artificial Intelligence (AI) and Machine learning (ML) represent an important evolution in computer science and data processing systems [20] whose methods such as CNN & RNN have been increasingly popular in NLP for a long time.

But in terms of performance for the tasks of comprehending and creating natural languages, the Transformer model outperformed competing neural models such as convolutional and recurrent neural networks[21].

This chapter serves as an overview of machine learning and Deep learning and its methods CNN & RNN. Finally an introduction to transformers and the attention mechanism.

2.1 Machine Learning

Machine learning (ML) enables inferring models or relationships by learning from data. ML is finding major applications in process modeling, computer vision, speech recognition, and language understanding[7].

Machine learning, a subcategory of AI while AI refers to the simulation of human intelligence processes by a computer, machine learning refers to computational models that effectively perform a specific task without using explicit instructions but learning from data to make predictions or decisions[7]. ML is composed of three areas: supervised learning, unsupervised learning, and reinforcement learning.

2.1.1 ML Models

A machine learning model is a program that can find patterns or make decisions from a previously unseen dataset. This part introduces you to the most prominent learning models. The primary learning models that will be considered are shown in **Figure 2.1**[16]:

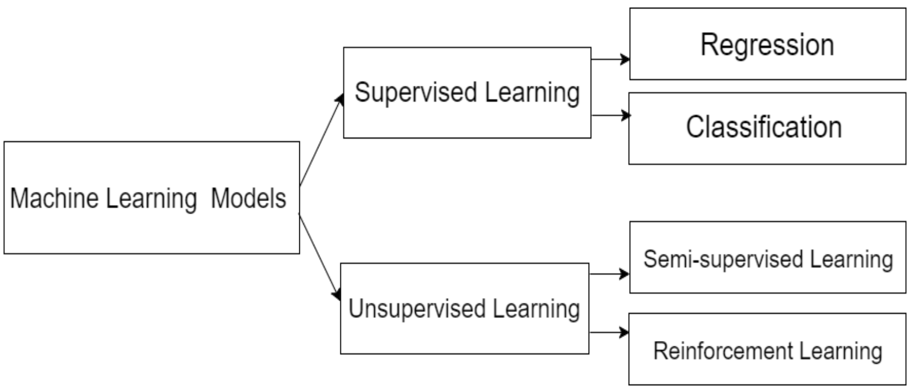

Figure 2.1: Machine Learning Model

- **Supervised Learning:**

  In supervised learning (SL), the machine is given a dataset, along with the right answers to a question corresponding to the data points. The learning algorithm is provided with a huge set of data points with answers . The algorithm has to learn the key characteristics within each data point in the dataset to determine the answer. So, next time a new data point is provided to the algorithm, the algorithm should be able to predict the right answer. Supervised learning falls mostly in two categories they are [16]:

  - Classification Problem: Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.
    The task of the classification algorithm is to find the mapping function to map the input(x) to the discrete output(y).
  - Regression Problem: Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables.
    The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).

- **Unsupervised Learning:** In unsupervised learning, the machine is provided with a set of data and is not provided with any right answer. Given the huge amount of data, the machine may identify trends of similarity. The algorithm will identify groups of similar items or similarity of new items with existing groups, etc.
  Methods of unsupervised learning.

  - Semi-supervised Learning: Semi-supervised learning falls somewhere between supervised learning and unsupervised learning. Here, the device is given a large dataset, in which only a few data points are labeled. The algorithm will use clustering techniques to identify clusters within the given dataset and use the few labeled data points within Each group to provide labels for other data points in the same group.
  - Reinforcement Learning: In reinforcement learning, the algorithm is made to train itself using many trial and error experiments. Reinforcement learning happens when the algorithm interacts continually with the environment, rather than relying on training data. One of the most popular examples of reinforce-

ment learning is autonomous driving.

2.2 Deep Learning

Deep learning is a subset of machine learning, which is artificial neural networks (ANN) with several hidden layers (three or more layers) between the input and output layers known as deep neural networks (DNN).
Deep learning neural networks attempt to mimic the human brain through a combination of data inputs, weights, and bias. These elements work together to accurately recognize, classify, and describe objects within the data.
DL models are categorized based on their neural network topologies, such as recurrent neural networks (RNN) and convolutional neural networks (CNN). RNN detects patterns over time, while CNN can identify patterns over space[1].

Recurrent neural network (RNN) is one of the most widely used algorithms in deep learning. more details in next section.

2.2.1 Recurrent Neural Network

Recurrent neural networks have been an important focus of research and development during the 1990s. A recurrent net is a neural network with feedback (closed loop) connections. Its techniques have been applied to a wide variety of problems[15].

Text is viewed as a series of words by RNN models designed to capture word relationships and sentence patterns for text analytics.

Recurrent neural networks utilize training data to learn. They are distinguished by their "memory" as they take information from prior inputs to influence the current input and output[14]. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (NLP).

The following points briefly discuss simple RNN, LSTM, and GRU models.

- Simple RNN network: can be viewed as a single layer recurrent neural network where the activation is delayed and fed back synchronously with the external input[19].

Simple RNN Architecture showing in **Figure 2.2**[5] .

- LSTM network: is a modified version of recurrent neural networks, which makes it easier to remember past data in memory[5].

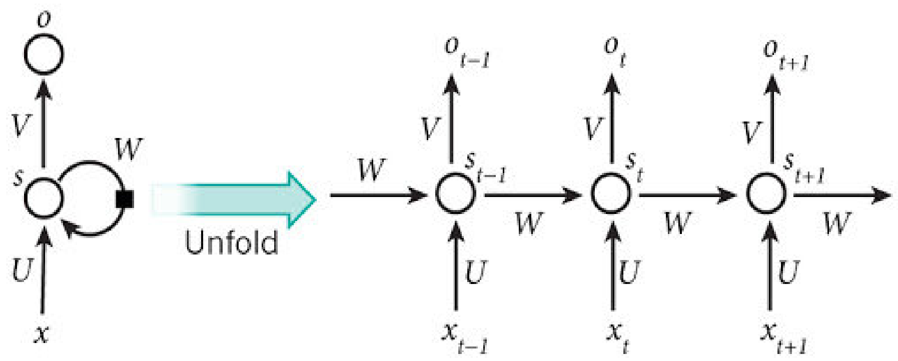

Figure 2.2: Simple RNN Architecture

2.2.2 Convolution Neural Network

CNN is a neural network with many successes and inventions in image processing and computer vision.

The main architecture of CNN is depicted in **Figure 2.3** [**1**]. A CNN consists of several layers: an input layer, a convolutional layer, a pooling layer, a fully connected layer. The input layer receives the image pixel value as input and passes it to the convolutional layer[1].

The convolution layer computes output using kernel or filter values, subsequently transferred to the pooling layer. The pooling layer shrinks the representation size and speeds up computation.

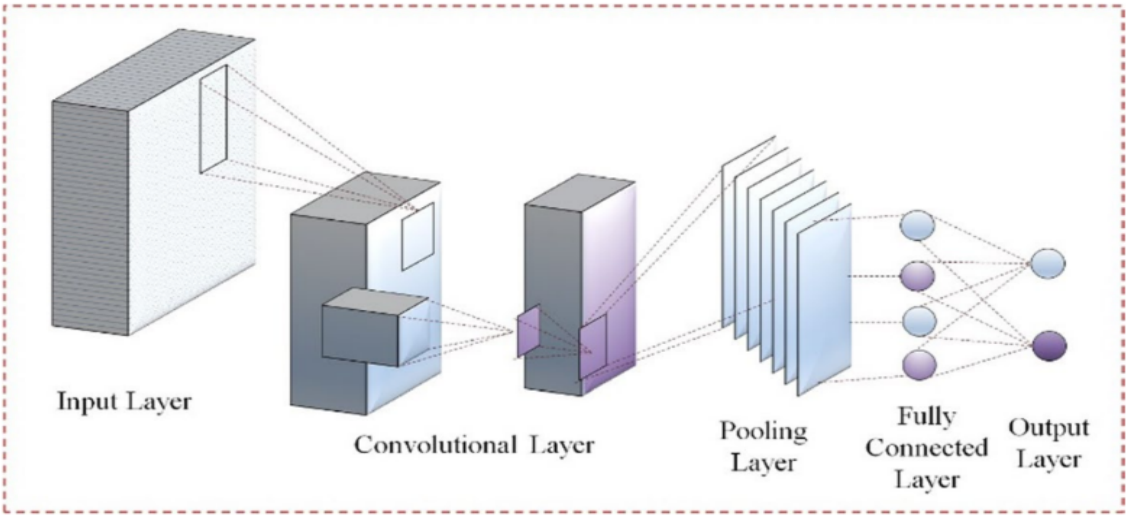

Figure 2.3: Architecture of CNN.

2.2.3 Word Embeddings

A word embedding is a fixed-length vector of real values representing the word meaning in such a way that it can be distinguishable from all the other words.

The word embedding technique represented by deep learning has received much attention. It has been of benefit to a number of NLP tasks such as Named Entity Recognition (NER)[11] and Sentiment Analysis (SA)[10].

Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have the drawback of processing text sequentially, making them unsuitable for parallel processing with contemporary hardware such as graphics processing units (GPUs). In terms of performance for the tasks of comprehending and creating natural languages, the Transformer model outperformed competing neural models such as convolutional and recurrent neural networks.

2.3 Overview of Transformer:

On 12 June 2017 paper called “Attention Is All You Need” in which a new neural architecture called the transformer was proposed. The main highlight of this work was a mechanism called self-attention[21].

Transformers intended to change sequential processing by providing the whole sequence as input at one shot to the network and allowing the network to learn one whole sentence at once. The transformer architecture's main point was to only use self-attention for capturing the dependencies between the words in a sequence and not depend on any of the RNN- or LSTM-based approaches[9].

An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key[21].

2.4 Transformers Model Architecture

In this section, we will briefly introduce the mechanisms of transformers. A high-level architecture of the transformer is shown in the following **Figure 2.4**[**21**].

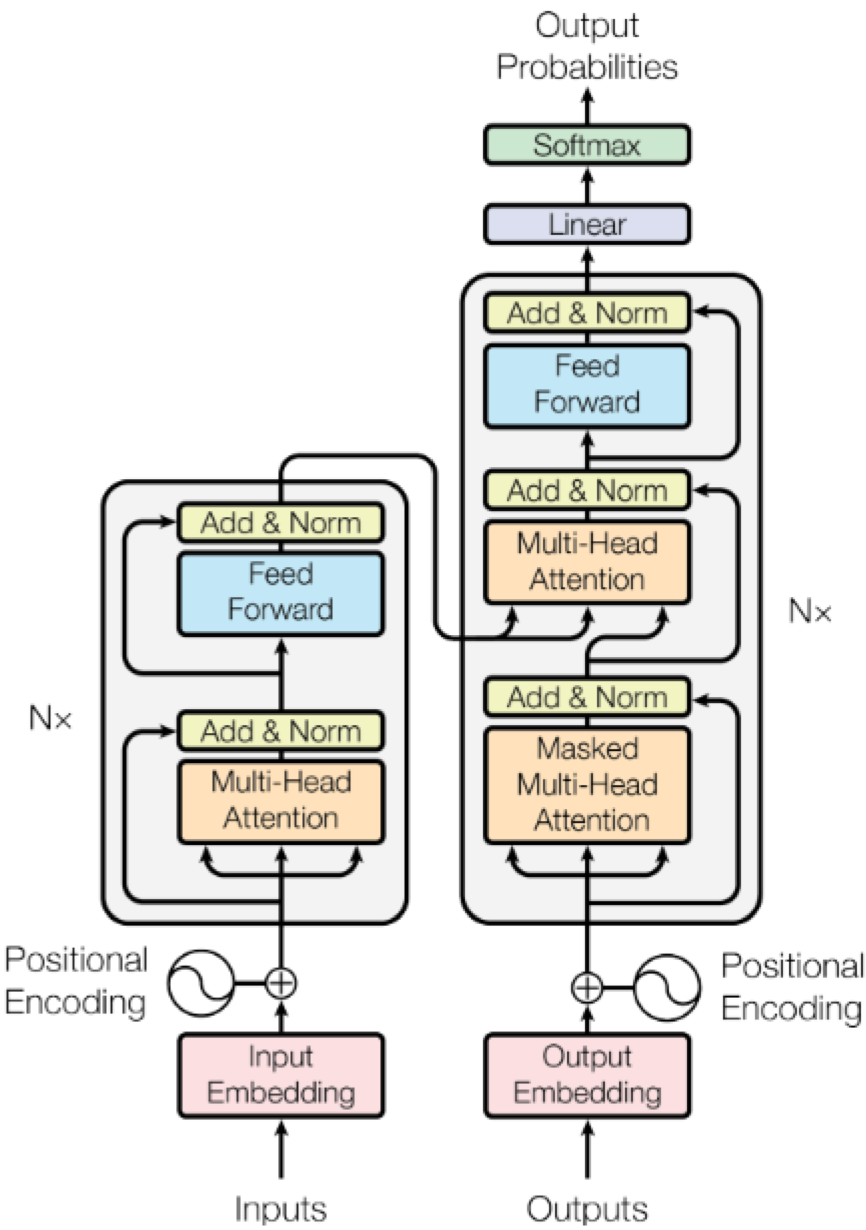

Figure 2.4: Transformer Global Architecture.

2.4.1 Multi-Head Attention and Self-Attention

Multi-head attention is an implementation of parallel computing for scaled dot-product attention (SDP)[6]. The mathematical expression of SDP can be written as[21]:

$$\mathrm{SDP}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{n}}\right)\mathbf{V} \tag{1}$$

Where (K, V) is the key-value pair, Q is the query, and n is the sequence length.Based on Equation (1), multi-head attention computes the SDP multiple times in parallel and subsequently concatenates the independent outputs[6]. In particular, a linear transformation is usually implied to change the results into expected dimensions.
The mathematical expression of multi-headAtt[21]:

$$\text{multi-headAtt}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{Concat}\left[\mathrm{head}_1; \ldots; \mathrm{head}_h\right]\mathbf{W}^O \tag{2}$$

Where:

$$\mathrm{head}_{\mathrm{i}} = SDP\left(\mathbf{QW}_i{}^Q, \mathbf{KW}_i{}^K, \mathbf{VW}_i{}^V\right) \tag{3}$$

and

$$\mathbf{QW}_i{}^Q, \mathbf{KW}_i{}^K, \mathbf{VW}_i{}^V, \mathbf{W}^O$$

are all network parameters to be learned. On the basis of multi-headAtt, self-attention is a technique that simply utilizes the input X to generate Q, K, V in Eq. (2). This mechanism allows the inputs to interact with each other (self) and help the model pin out where they should pay more attention to[6].as shown in **Figure 2.5[21]**

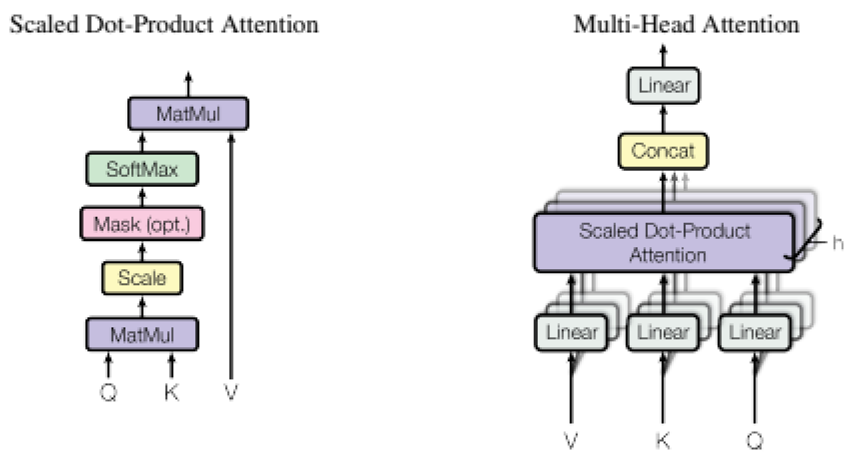

Figure 2.5: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention.

2.4.2 Encoder and Decoder

The encoder is composed of a stack of N = 6 identical layers. Each layer has two sub-layers that have a residual connection[8] around them, followed by layer normalization[2]. All sub-layers in the model, as well as the embedding layers, produce outputs of dimension d-model = 512.
The first sub-layers is a multi-head self-attention mechanism, and the second is a simple, position- wise fully connected feed-forward network.

In addition to all sublayers in the encoder, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack[21]. The self-attention sub-layer in the decoder stack prevents positions from attending to subsequent positions.

2.5 Conclusion

In this chapter, we saw an overview of machine learning and Deep learning and its methods CNN & RNN. Finally an introduction to transformers and the attention mechanism.

CONCLUSION

Natural language processing employs computational techniques for the purpose of learning, understanding, and producing human language content. The aim of NLP tasks is not only to understand single words individually but to be able to understand the context of those words. The common NLP tasks represent sentiment analysis, question answering, and translation..etc.

Question Answering models can retrieve the answer to a question from a given text, which is useful for searching for an answer in a context or document.

The Transformers is a neural network with zero recurrences and just uses the attention mechanism. It was originally introduced for machine translation tasks, but it has spread into many other application areas such as question answering. The transformers outperformed RNN and LSTM.

In this work, we presented the first Arabic dataset in Fiqh and Syrah. QAFSv1 contains 550 questions where 440 for the training file and 110 for the test file. We trained 3 different transformer models on dataset (BERT, DistlBERT, and ELECTRA) using the simple transformer library. After that we evaluate the models using evaluation metrics (correct, incorrect, and similar). Finally, we obtained relatively good results for BERT & DistilBERT, and ELECTRA.

For future work, we have plans to continue working on the project and complete the dataset in the near future, Also to experiment more with the training and Hyper-parameter optimization of BERT, DistilBERT, and ELECTRA transformer models, and even the possibility of fine-tuning or pretraining a transformer model from scratch on exclusively QAFS data.

REFERENCES

[1] Asudani, D. S., Nagwani, N. K., and Singh, P. Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review* (2023), 1–81.

[2] Ba, J. L., Kiros, J. R., and Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).

[3] Bhardwaj, A., Khanna, P., Kumar, S., et al. Generative model for nlp applications based on component extraction. *Procedia Computer Science 167* (2020), 918–931.

[4] Chopra, A., Prashar, A., and Sain, C. Natural language processing. *International journal of technology enhancements and emerging engineering research 1*, 4 (2013), 131–134.

[5] Dey, P., Hossain, E., Hossain, M., Chowdhury, M., Alam, M., Hossain, M., and Andersson, K. Comparative analysis of recurrent neural networks in stock price prediction for different frequency domains. *Algorithms 14* (08 2021), 251.

[6] Fan, R. Transformer-based deep learning method for the prediction of ventilator pressure. In *2022 IEEE 2nd International Conference on Information Communication and Software Engineering (ICICSE)* (2022), pp. 25–28.

[7] Ghosh, S., and Dasgupta, R. *Machine Learning in Biological Sciences: Updates and Future Prospects*. Springer Nature, 2022.

[8] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[9] Jain, S. M. *Introduction to Transformers*. Apress, Berkeley, CA, 2022, pp. 19–36.

[10] Kaibi, I., Nfaoui, E. H., and Satori, H. Sentiment analysis approach based on combination of word embedding techniques. In *Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco* (2020), Springer, pp. 805–813.

[11] Khalifa, M., and Shaalan, K. Character convolutions for arabic named entity recognition with long short-term memory networks. *Computer Speech & Language 58* (2019), 335–346.

[12] Khurana, D., Koli, A., Khatter, K., and Singh, S. Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications 82*, 3 (2023), 3713–3744.

[13] Liddy, E. D. Natural language processing.

[14] Ma, Y., Yu, J., Ji, B., Chen, J., Zhao, S., and Chen, J. Three-way decisions based rnn models for sentiment classification. In *Rough Sets: International Joint Conference, IJCRS 2021, Bratislava, Slovakia, September 19–24, 2021, Proceedings* (2021), Springer, pp. 247–258.

[15] Medsker, L. R., and Jain, L. Recurrent neural networks. *Design and Applications 5* (2001), 64–67.

[16] Rebala, G., Ravi, A., Churiwala, S., Rebala, G., Ravi, A., and Churiwala, S. Machine learning definition and basics. *An introduction to machine learning* (2019), 1–17.

[17] Sabharwal, N., and Agrawal, A. Hands-on question answering systems with bert: applications in neural networks and natural language processing. *(No Title)* (2021).

[18] Sabharwal, N., Agrawal, A., Sabharwal, N., and Agrawal, A. Neural networks for natural language processing. *Hands-on Question Answering Systems with BERT: Applications in Neural Networks and Natural Language Processing* (2021), 15–39.

[19] Salem, F. M. *Network Architectures.* Springer International Publishing, Cham, 2022, pp. 3–19.

[20] Soori, M., Arezoo, B., and Dastres, R. Machine learning and artificial intelligence in cnc machine tools, a review. *Sustainable Manufacturing and Service Economics* (2023), 100009.

[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, ., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems 30* (2017).