

# AI 829 NLP Project: Context-Based Open-Domain Question Answering system

Akshay Mahesh Gudiyawar (MT2020137)

Vikrant Punia (MT2020049)

Shubham Jha (MT2020514)

**Problem Statement:** A model which answers general questions based on a context in text based answering format. We intend to use Wikipedia texts, NCERT textbooks, and Stanford's SQUAD Dataset for training (link given below). The answer to every question would be a segment of text, phrase. For e.g., when trained on context about Amazon Rainforest, the model should be able to answer general questions like - In what country can most of the amazon rainforest be found? Where is the drainage basin of Amazon river? etc.

**Dataset Link:** <https://rajpurkar.github.io/SQuAD-explorer/>

The Stanford Question Answering Dataset (SQuAD) is a dataset for machine comprehension based on Wikipedia. The dataset contains 87k examples for training and 10k for development. Each example is composed of a paragraph extracted from a Wikipedia article and an associated human-generated question. The answer is always a span from this paragraph and a model is given credit if its predicted answer matches it. Two evaluation metrics are used: exact string match (EM) and F1 score, which measures the weighted average of precision and recall at the token level. For the task of evaluating open-domain question answering over Wikipedia, we use the SQuAD development set QA pairs only, and we ask systems to uncover the correct answer spans without having access to the associated paragraphs. That is, a model is required to answer a question given the whole of Wikipedia as a resource and not by just giving the relevant paragraph.

**Solution Approach:** We utilize the Document Reader architecture in DrQA et al. [1] to predict answer text span in the context for the given questions. The approach presented in [1] combines a search component based on bigram hashing and TF-IDF matching with a multi-layer recurrent neural network model trained to detect answers in Wikipedia paragraphs. Using Wikipedia articles as the knowledge source causes the task of question answering (QA) to combine the challenges of both large-scale open-domain QA and of machine comprehension of text. It treats Wikipedia as a collection of articles and does not rely on its internal graph structure. The system DrQA system for MRS (Machine reading at scale) which consists of two components: (1) the Document Retriever module for finding relevant articles and (2) a machine comprehension model, Document Reader, for extracting answers from a single document or a small collection of documents as show in Figure [1].

## Open-domain QA

SQuAD, TREC, WebQuestions, WikiMovies

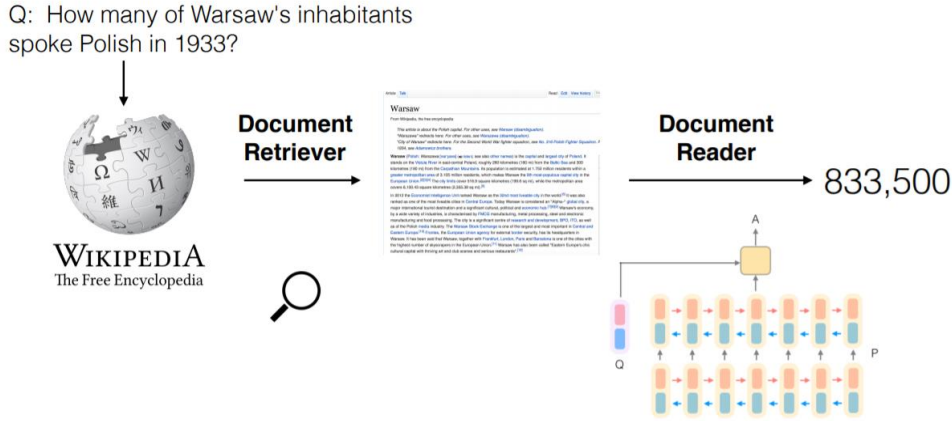


Figure [1] Overview of question answering system DrQA

In Document Retriever, a simple inverted index lookup followed by term vector model scoring is used for the retrieving task. Articles and questions are compared as TF-IDF weighted bag-of-word vectors. Its further improved by taking local word order into account with n-gram features. The Document Retriever is the first part of the full model, by setting it to return 5 Wikipedia articles given any question. Those articles are then processed by Document Reader.

In Document Reader part, given a question  $q$  consisting of  $i$  tokens  $\{q_1, \dots, q_i\}$  and a document or a small set of documents of  $n$  paragraphs where a single paragraph  $p$  consists of  $m$  tokens  $\{p_1, \dots, p_m\}$ , an RNN model is used which is applied to each paragraph in turn and then finally aggregate the predicted answers. To encode a question, we apply another recurrent neural network on top of the word embeddings of  $q_i$  and combine the resulting hidden units into one single vector. By this we compute  $q = \text{Summation}(b_j \cdot q_j)$  where  $b_j$  encodes the importance of each question word. At the paragraph level, the goal is to predict the span of tokens that is most likely the correct answer. We take the paragraph vectors  $\{p_1, \dots, p_m\}$  and the question vector  $q$  as input, and train two classifiers independently for predicting the two ends of the span. We use the unnormalized exponential and take argmax over all considered paragraph spans for our final prediction.

The two metrics we will be using are exact match and F1. The exact match metric measures the percentage of predictions that match any one of the ground truth answers exactly. The F1 score is a looser metric, which measures the average overlap between the prediction and ground truth answer. The metrics are defined as follows:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Precision} = \frac{TP}{TP + FP} = \frac{\# \text{ of matching tokens}}{\# \text{ of predicted tokens}}$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{\# \text{ of matching tokens}}{\# \text{ of gold tokens}}$$

These are then averaged with the "has answer" prediction where we replace the individual question-answer pair score with a 1 if it correctly predicts that the question is not answerable, and keep the original score otherwise. We use the basic pretrained bert-base-uncased model from huggingface transformer to get the representation for tokens in the (context, question) pairs. Then use a linear classifier to classify answer start/end for tokens in context. The (context, question) pair is pre-processed via huggingface transformer's bert-base-uncased tokenizer to get token\_ids, attention mask, sequence mask (whether is in context or question sequence), then fed into our model. Our model is inherited from huggingface transformer's BertPreTrainedModel class, which is a template class for developers to make new classes based on pre-trained models. Both linear classifier and bert model are fine-tuned.

For hyper-parameters, initial learning rate is 1e-3 and batch size used is 16.

### **Results:**

On validation set, the following result metrics were obtained:

EM = 57.756 and F1 = 67.182 as shown in Figure [2]

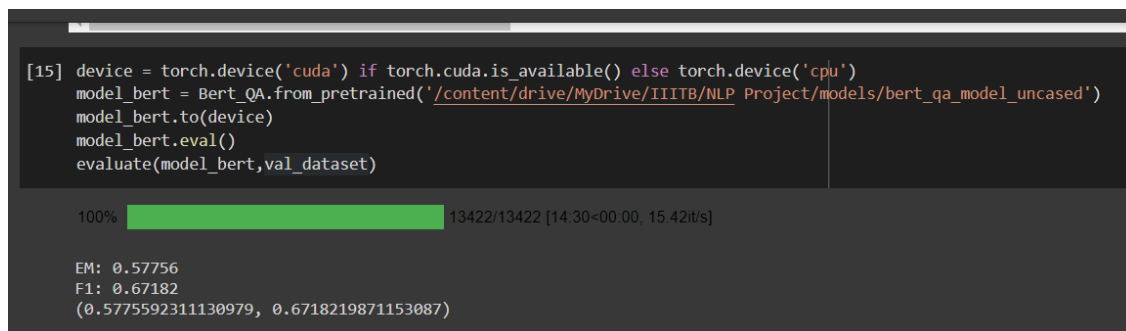


Figure [2] EM and F1 scores on Validation data

The EM scores on has-ans, and no-has groups are as shown in Figure [3]

Has ans EM = 51.83, Has ans F1 = 65.53

no-ans EM = 70.88, no-ans F1 = 70.88

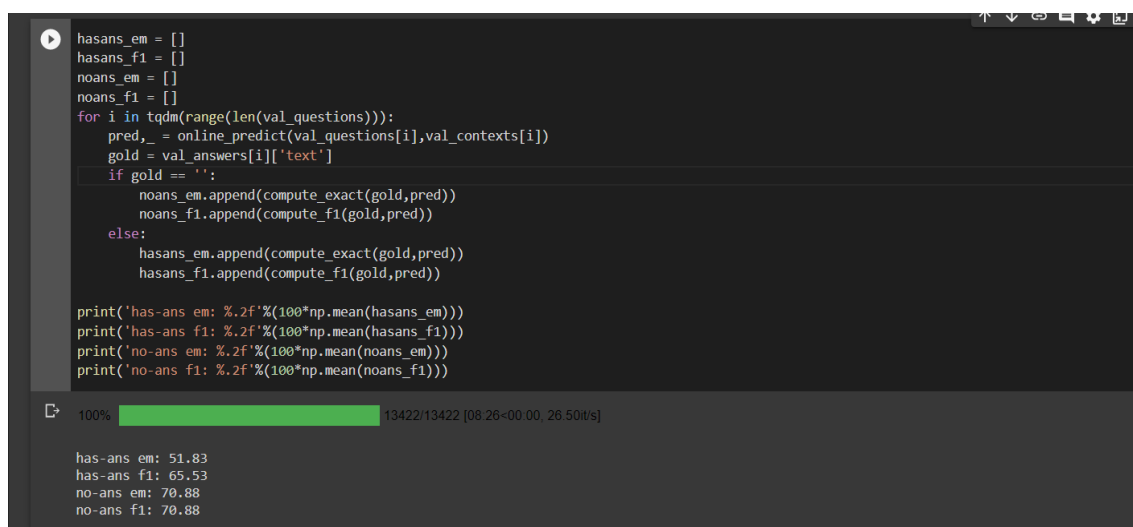


Figure [3] EM and F1 scores on has-answer and no-answer groups

Here are some example cases where answers were predicted as phrases to questions asked on context. We can compare the bert predicted answer with ground truth answer (named as gold answer).

```
context:
Southern Italy was also part of the Norman kingdom but great mosaics did not survive in this area except the fine mosaic pavement of the Otranto Cathedral from 1166, with mosaics tied into a tree of life.
question:
The mosaics in churches of Ravello prove what?
gold answer: mosaic art was widespread
bert answer: mosaic art was widespread in southern italy

context:
The Qing showed that the Manchus valued military skills in propaganda targeted towards the Ming military to get them to defect to the Qing, since the Ming civilian political system discriminated against them.
question:
What was the three Han Bannermen position called?
gold answer: viceroys
bert answer: liaodong

context:
In Kazakhstan Russian is not a state language, but according to article 7 of the Constitution of Kazakhstan its usage enjoys equal status to that of the Kazakh language in state and local administration.
question:
What was the population of Kazakhstan in 2004?
gold answer:
bert answer: 4, 200, 000

context:
Natives of North America began practicing farming approximately 4,000 years ago, late in the Archaic period of North American cultures. Technology had advanced to the point that pottery was becoming widespread.
question:
How did Archaic Indians use fire?
gold answer: a controlled manner
bert answer: in a controlled manner

context:
As conceptualized by the Polizeiwissenschaft, according to Foucault the police had an administrative, economic and social duty ("procuring abundance"). It was in charge of demographic concerns and neighborhood control.
question:
Which economic theory did Foucault say supported the police?
gold answer: mercantilist theory
bert answer:

context:
Each camp had its own religious personnel, standard bearers, priestly officers and their assistants, including a haruspex, and housekeepers of shrines and images. A senior magistrate-commander (sacred king) presided over the ceremonies.
question:
Even in other places, of what did the legions keep observance?
gold answer: Rome's official festivals
bert answer: rome's official festivals

context:
Compared to neighbouring countries, Namibia has a large degree of media freedom. Over the past years, the country usually ranked in the upper quarter of the Press Freedom Index of Reporters without Borders.
question:
Compared to neighbouring countries, what freedom does Namibia have?
gold answer: media
bert answer: media freedom

context:
```

```
context:
Sociologist Rodney Stark states that Jehovah's Witness leaders are "not always very democratic" and that members "are expected to conform to rather strict standards," but adds that "enforcement tends to be lenient."
question:
What profession did Rodney Stark grow up to become?
gold answer: Sociologist
bert answer:

context:
As the Industrial Revolution spread across Europe, chemists and manufacturers sought new red dyes that could be used for large-scale manufacture of textiles. One popular color imported into Europe from France in the 18th century was cochineal.
question:
What popular color was imported from France in the 18th century?
gold answer:
bert answer:

context:
Eisenhower retired to the place where he and Mamie had spent much of their post-war time, a working farm adjacent to the battlefield at Gettysburg, Pennsylvania, only 70 miles from his ancestral home.
question:
What did Barry Goldwater call Eisenhower?
gold answer: dime-store New Dealer
bert answer: a dime - store new dealer "

context:
Namibia follows a largely independent foreign policy, with persisting affiliations with states that aided the independence struggle, including Cuba. With a small army and a fragile economy, the Namibian government has been able to maintain its independence.
question:
When did Namibia join the UN?
gold answer: 23 April 1990
bert answer: 23 april 1990

context:
Portugal spearheaded European exploration of the world and the Age of Discovery. Prince Henry the Navigator, son of King João I, became the main sponsor and patron of this endeavour. During this period, Portugal discovered the Atlantic coast of Africa, the sea route to India, and Brazil.
question:
Who was the Father of Prince Henry the Navigator?
gold answer: King João I
bert answer: king joao i

context:
Current Governor of the Reserve Bank of India Raghuram Rajan had predicted the crisis in 2005 when he became chief economist at the International Monetary Fund. In 2005, at a celebration honouring A. B. Vaidya, he delivered a paper titled "Has Financial Development Made the World Riskier?"
question:
What was the name of Raghuram Rajan's controversial paper delivered in 2005?
gold answer: "Has Financial Development Made the World Riskier?"
bert answer: " has financial development made the world riskier? "

context:
Although the initial focus of the humanist scholars in the university was the discovery, exposition and insertion of ancient texts and languages into the university, and the ideas of those texts into the curriculum, the study of the self resulted in what field of study?
question:
A concentration on the study of the self resulted in what field of study?
gold answer: the humanities
bert answer: humanities

context:
```

```

context:
Current Governor of the Reserve Bank of India Raghuram Rajan had predicted the crisis in 2005 when he became chief economist at the International Monetary Fund.In 2005, at a celebration honouring A
question:
What was the name of Raghuram Rajan's controversial paper delivered in 2005?
gold answer: "Has Financial Development Made the World Riskier?"
bert answer: " has financial development made the world riskier? "

context:
Although the initial focus of the humanist scholars in the university was the discovery, exposition and insertion of ancient texts and languages into the university, and the ideas of those texts in
question:
A concentration on the study of the self resulted in what field of study?
gold answer: the humanities
bert answer: humanities

context:
Egypt was producing 691,000 bbl/d of oil and 2,141.05 Tcf of natural gas (in 2013), which makes Egypt as the largest oil producer not member of the Organization of the Petroleum Exporting Countries
question:
Where is Egypt planning to build is first nuclear power plant?
gold answer: El Dabaa city
bert answer: el dabaa city, northern egypt

context:
The Communist Party of China, founded in 1920, adopted the red flag and hammer and sickle emblem of the Soviet Union, which became the national symbols when the Party took power in China in 1949. U
question:
What has eighty members?
gold answer:
bert answer:

context:
The Third Great Awakening refers to a hypothetical historical period that was marked by religious activism in American history and spans the late 1850s to the early 20th century. It affected pietis
question:
What new groups were formed as a result of the Third Great Awakening?
gold answer: the Holiness, Nazarene, and Christian Science movements
bert answer: holiness, nazarene, and christian science movements

context:
The cartridge media of the console is officially referred to as Game Pak in most Western regions, and as Cassette (カセット, Kasetto?) in Japan and parts of Latin America. While the SNES can address
question:
How much data space can an SNES cartridge use?
gold answer: 117.75 Mbit
bert answer: 128 mbit

context:
Eisenhower did provide France with bombers and non-combat personnel. After a few months with no success by the French, he added other aircraft to drop napalm for clearing purposes. Further requests
question:
Along with the Vice President and MGS head, who urged Eisenhower to intervene after the fall of Dien Bien Phu?
gold answer: Chairman of the Joint Chiefs
bert answer: chairman of the joint chiefs

```

## References:

[1] <https://arxiv.org/abs/1704.00051> (Reading Wikipedia to Answer Open-Domain Questions, by: Danqi Chen, Adam Fisch, Jason Weston, Antoine Bordes)