

Knowledge Processing for Intelligent Systems

Seminar Presentation



Open Domain Question Answering with BERT

Yuqian Lei & Yunlong Wang

Outline



1. Introduction

- Motivation
- Background

2. Related work

3. Model BERT

- Model overview
- Key-points

4. Approach

- Environment
- Experimental setup

5. Results

6. Discussion



Part 1: Introduction

Motivation

What are we interested in?

- Transfer learning from one domain to another domain (QA)
 - Two stages, pretraining and fine-tuning
 - In our work, fine-tuning a pretrained BERT model with GermanQuAD
 - Apply the knowledge obtained by solving a task in a domain to a different but related task or domain
- Basic idea behind the design of BERT

Why transfer learning and BERT?

- Fine-tune the pretrained model to adapt to another domain/task
 - Less resources needed (knowledge reused)
 - Less training time
 - The model would be more versatile since it's not bound to one dataset
- Outstanding performance in modern NLP task

Background

Question answering (QA) system

- Answer questions posed in a natural language in the form of short texts
- Open domain QA system
 - Across different domains
 - No domain-specific knowledge needed to prepare a question
 - e.g. Wikipedia
- Closed domain QA system
 - Restricted to a specific domain
 - Domain-specific knowledge needed
 - e.g. medicine, biology

How does it work in general?

- The model is trained to give the span of an answer with the information provided or retrieved.

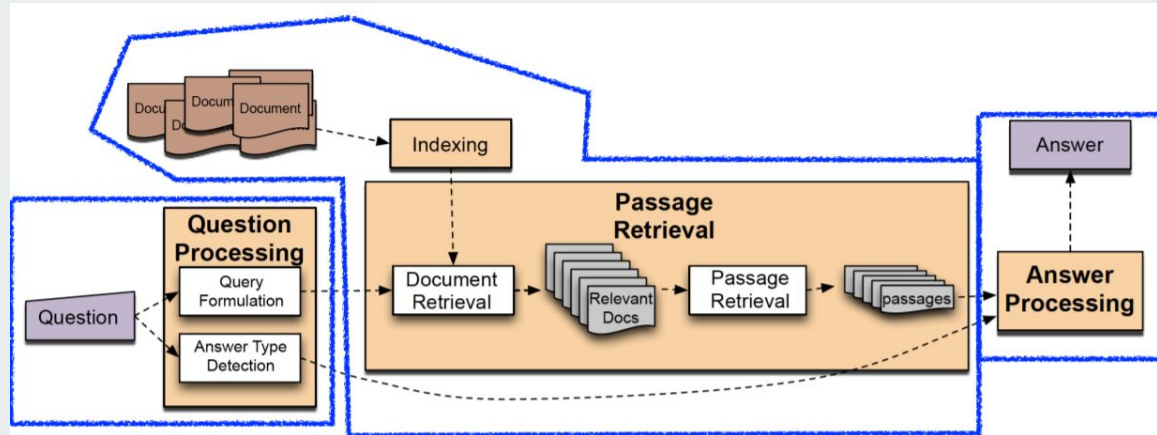


Part 2: Related Work

Early QA system

~~Closed~~ domain only

- A representative QA system
 - Introduced in 1998s
 - 1) Form a query and detect the answer type based on the question posed
 - 2) Formed query + indexed documents -> retrieve relevant documents
 - 3) Detected answer type + retrieved documents -> answer



Recent QA system

Open domain

- DrQA (2017)
 - First representative open-domain QA system
 - 1) Document retrieval using an untrainable algorithm
 - 2) Document reading using a multi-layer recurrent neural network

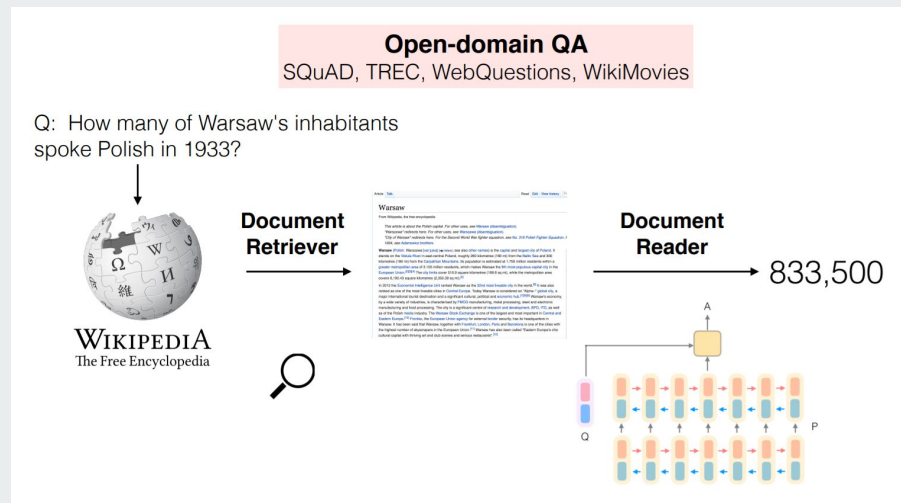


figure taken from [here](#)

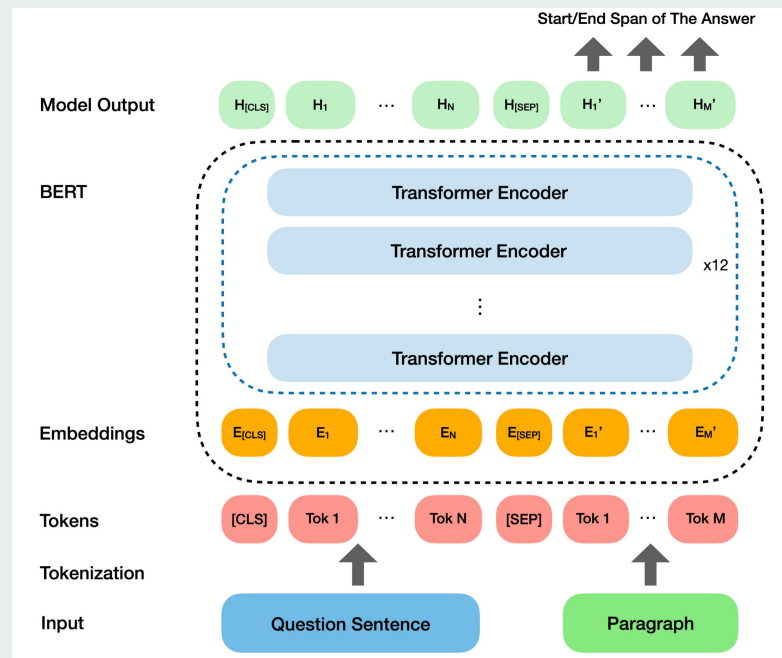


Part 3: BERT

General structure for QA using BERT

German BERT (same structure)

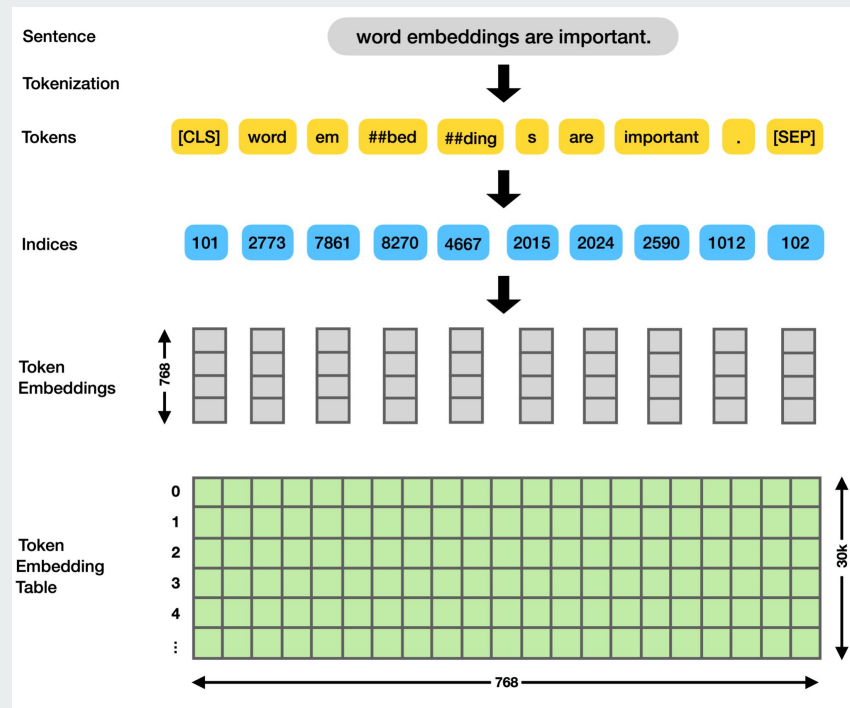
- Pretrained using a German corpus instead of an English corpus
- Three main components
 - Tokenization
 - Embeddings
 - Transformer Encoders
 - 12 in base BERT
 - Self-attention layer
 - How a token is related to all other tokens and itself
 - Feed forward network



Tokenization & token embeddings

Tokenizer

- Sentence/paragraph to tokens
 - A word can be broken down into several tokens
 - e.g. 'embeddings' -> 'em', '##bed', '##ding', 's'
 - Help the model understand a new word by breaking it down into the words it has already learned
- Two special tokens
 - [CLS] - classification task
 - [SEP] - separate sentences



Embedding layers

Three embeddings for each token

- Token embedding
- Segment embedding
 - Identify which sentence a token belongs to
- Position embedding
 - Locate a token in a sentence

Sentences	CLS	I	buy	a	hat	[SEP]	it	looks	very	nice	[SEP]
Token Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}
Input Embeddings	$E_{[CLS]}$	E_1	E_2	E_3	E_4	$E_{[SEP]}$	E'_1	E'_2	E'_3	E'_4	$E'_{[SEP]}$



Part 4: Approach

Environments

Hardware:

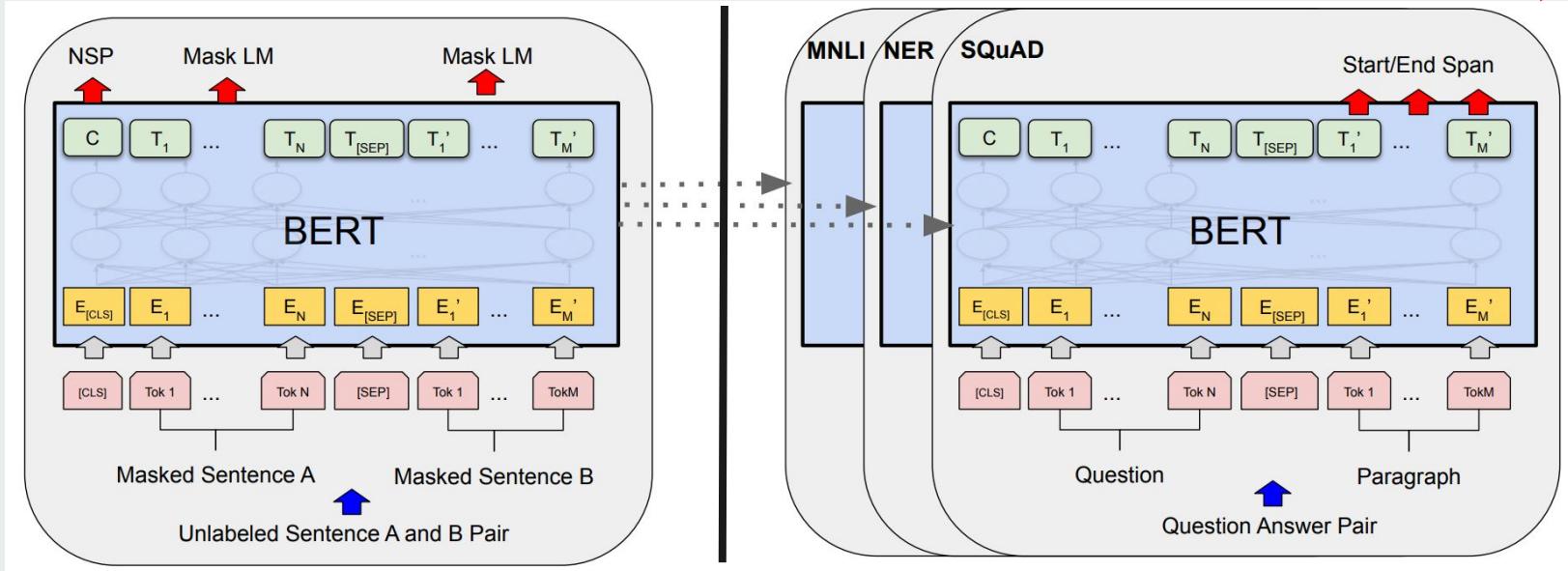
- CPU: Intel(R) Core(TM) i7-4930K CPU @ 3.40GHz
- GPU: NVIDIA GeForce GTX 1080

Software:

- Ubuntu 18.04 x86_64
- Pycharm IDE
- Anaconda virtual env
 - Python 3.9.7
 - **Pytorch**
 - **Transformers**
 - Numpy
 - ...

Git Link (on mafiasi): https://git.mafiasi.de/21wang/BERT_on_GermanQuAD.git

Experimental Workflow



Unsupervised Pretraining

Task: NSP and Masked LM

Dataset: Wiki, OpenLegalData, News in germany

Supervised Fine-tuning

Task: extractive QA with Information retrieved

Dataset: [GermanQuAD](#)

figure taken from [here](#)

Fine-tuning

1. Data preprocessing stage

- a. Splitting into train-set, valivation-set, test-set
- b. Tokenization
- c. Building question-answer pairs for further inputs

2. Training stage

Loop each epoch:

for batch in dataset:

output = model(bacth) # front-forward

loss = cross-entropy(output, ground-truth)

loss.backward() #calculate gradient

optimizer.step() # update the parameters of model

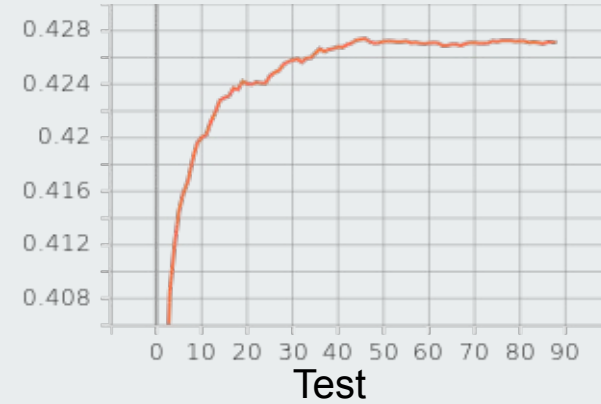
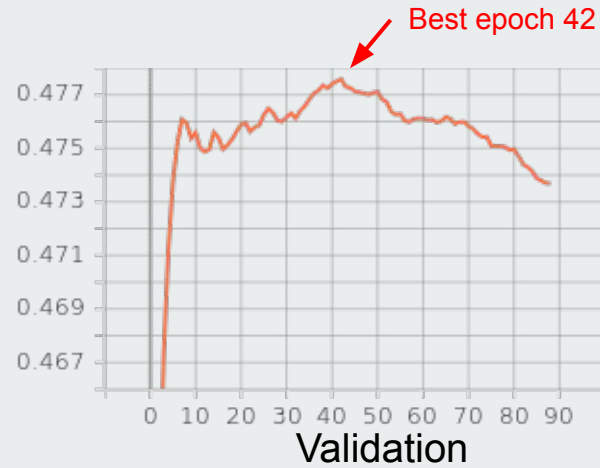
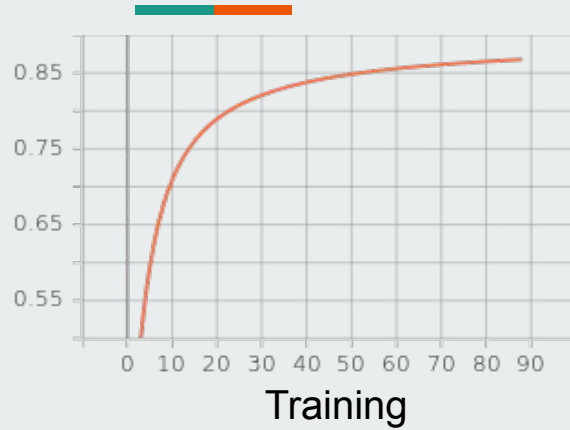
Notice:

- All parameters are opened for fine-tuning
- Use small learning rate at the beginning to avoid too much knowledge-forgetting by fine tuning stage



Part 5: Results

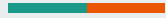
F1 Score



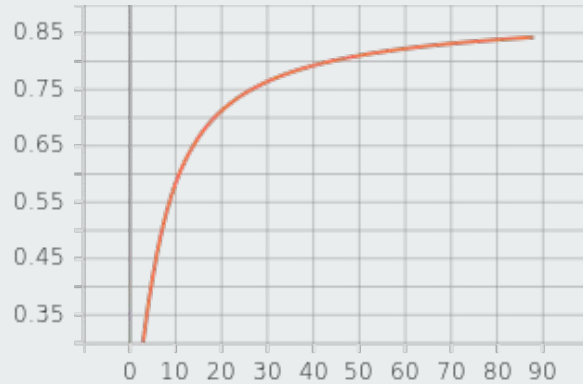
Metrics in epoch 42:

	F1 Score	Exact Match
Training	0.7968	0.7968
Validation	0.4776	0.2747
Test	0.4270	0.2243

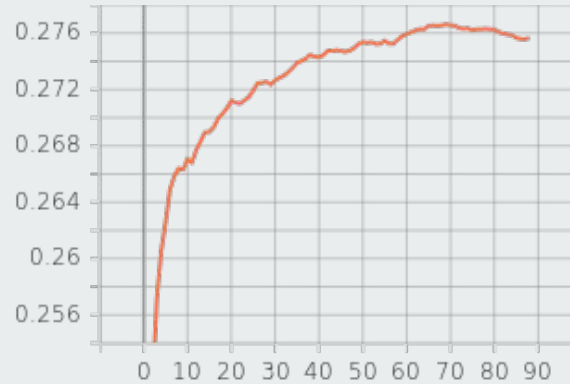
Exact Match (EM)



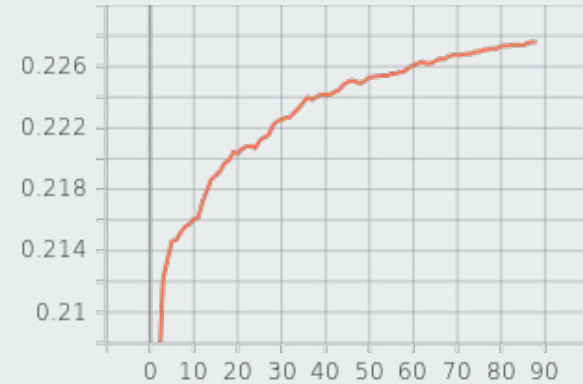
Training



Validation



Test



Prediction Top-1

```
Run: main x
ssh://1wang@wtmstuc2.informatik.uni-hamburg.de:22/data/1wang/anaconda3/envs/trf/bin/python3.9 -u /export/home/1wang/projects/qa/main.py
Context: Pizza ist ein Essen mit einer hohen religiösen Bedeutung, vergleichbar mit Manna im Christentum. Oftmals werden sie in dunklen Räumen unter Einwirkung von LED-Licht konsumiert, ein religiöses Ritual, bei dem man sich vor das LED-Licht beugt und Gebete spricht. Die Pizza genießt unter diesen Wesen ein hohes Maß an Verehrung, viele Lichtsymbole haben die gleiche Form wie eine Pizza. Pizza ist rund, weil die sie konsumierenden Wesen fast nie etwas mit runden Sachen (bspw. sich hinter Stoff verbergende Haut) zu tun haben, die Rundheit ist also gewissermaßen ein Ausgleich. Pizza ist heiß wie alle anderen religiösen Lebensmittel (z.B. Kaffee). Der Pizza Bestellprozess ist eine höchst komplexe Angelegenheit, da hier sehr viel soziale Interaktion gefordert wird und Ausdrucksstärke, wenn man antwortet, für welches Reliquium man sich entschieden hat. Jemand, der die Pizza-Bestellungen entgegennimmt, ist gewissermaßen ein Pfarrer, da er*sie sich um seine Schäfchen kümmern und sie trösten muss, wenn die betreffende Person seit 12h mit niemandem geredet oder anderweitig sozial interagiert hat (von der Teilnahme an einem kollektiven Massenmord aka Counterstrike mal abgesehen). Die Pizza führt zur Auferstehung, sie ist der Wein des Lebens.
please input your question: Was ist Pizza?
Was ist Pizza?
Answer Top-1: [der Wein des Lebens]
please input your question: Was ist Bestellprozess der Pizza?
Was ist Bestellprozess der Pizza?
Answer Top-1: [Der Pizza Bestellprozess ist eine höchst komplexe Angelegenheit, da hier sehr viel soziale Interaktion gefordert wird und Ausdrucksstärke, wenn man antwortet, für welches Reliquium man sich entschieden hat.]
please input your question: Wie kann man Pizza essen?
Wie kann man Pizza essen?
Answer Top-1: [Oftmals werden sie in dunklen Räumen unter Einwirkung von LED-Licht konsumiert, ein religiöses Ritual, bei dem man sich vor das LED-Licht beugt und Gebete spricht.]
please input your question: Warum Pizza ist rund?
Warum Pizza ist rund?
Answer Top-1: [weil die sie konsumierenden Wesen fast nie etwas mit runden Sachen (bspw. sich hinter Stoff verbergende Haut) zu tun haben]
please input your question: exit
exit

Process finished with exit code 0
```

Prediction Top-1 (Text Version)

Missed!

Pizza ist ein Essen mit einer hohen religiösen Bedeutung, vergleichbar mit Manna im Christentum. Oftmals werden sie in dunklen Räumen unter Einwirkung von LED-Licht konsumiert, ein religiöses Ritual, bei dem man sich vor das LED-Licht beugt und Gebete spricht. Die Pizza genießt unter diesen Wesen ein hohes Maß an Verehrung, viele Lichtsymbole haben die gleiche Form wie eine Pizza. Pizza ist rund, weil die sie konsumierenden Wesen fast nie etwas mit runden Sachen (bspw. sich hinter Stoff verbergende Haut) zu tun haben, die Rundheit ist also gewissermaßen ein Ausgleich. Pizza ist heiß wie alle anderen religiösen Lebensmittel (z.B. Kaffee). Der Pizza Bestellprozess ist eine höchst komplexe Angelegenheit, da hier sehr viel soziale Interaktion gefordert wird und Ausdrucksstärke, wenn man antwortet, für welches Reliquium man sich entschieden hat. Jemand, der die Pizza-Bestellungen entgegennimmt, ist gewissermaßen ein Pfarrer, da er*sie sich um seine Schäfchen kümmern und sie trösten muss, wenn die betreffende Person seit 12h mit niemandem geredet oder anderweitig sozial interagiert hat (von der Teilnahme an einem kollektiven Massenmord aka Counterstrike mal abgesehen). Die Pizza führt zur Auferstehung sie ist der Wein des Lebens.

– From Wiki of UHH Informatiku

Q: Was ist Pizza?

A: der Wein des Lebens

Q: Was ist Bestellprozess der Pizza?

A: Der Pizza Bestellprozess ist eine höchst komplexe Angelegenheit, da hier sehr viel soziale Interaktion gefordert wird und Ausdrucksstärke, wenn man antwortet, für welches Reliquium man sich entschieden hat.

Q: Wie kann man Pizza essen?

A: Oftmals werden sie in dunklen Räumen unter Einwirkung von LED-Licht konsumiert, ein religiöses Ritual, bei dem man sich vor das LED-Licht beugt und Gebete spricht.

Q: Warum Pizza ist rund?

A: weil die sie konsumierenden Wesen fast nie etwas mit runden Sachen (bspw. sich hinter Stoff verbergende Haut) zu tun haben

Prediction Top-3 (Text Version)



please input your question: Was ist Pizza?

Answer Top-1: [der Wein des Lebens]

Answer Top-2: [ein Essen mit einer hohen religiösen Bedeutung, vergleichbar mit Manna im Christentum. Oftmals werden sie in dunklen Räumen unter Einwirkung von LED-Licht konsumiert, ein religiöses Ritual, bei dem man sich vor das LED-Licht beugt und Gebete spricht.]

Answer Top-3: [Oftmals werden sie in dunklen Räumen unter Einwirkung von LED-Licht konsumiert, ein religiöses Ritual, bei dem man sich vor das LED-Licht beugt und Gebete spricht. Die Pizza genießt unter diesen Wesen ein hohes Maß an Verehrung, viele Lichtsymbole haben die gleiche Form wie eine Pizza. Pizza ist rund, weil die sie konsumierenden Wesen fast nie etwas mit runden Sachen (bspw. sich hinter Stoff verbergende Haut) zu tun haben, die Rundheit ist also gewissermaßen ein Ausgleich.]

Both hit !



Part 6: Discussion

Discussion



Basically, the model we fine-tuned is usable.

And it still can be optimized by:

- **Dataset:** More preprocessing, e.g. denoise, augmentation.
- **Model:** Use advantage model structure, e.g. [RoBERTa](#), [ELECTRA](#).
- **Training:** Further Hyperparameters tuning and some other strategy, e.g. Warmup.

Summary



- Introduce some concepts of the QA system
- Dive into the details of the SOTA model BERT
- Show the workflow of pretraining and fine-tuning by NLP task
- Train a model that has usable performance
- Further strategies for improvement



Thank you for your attention !



Questions ?