

standford_qa_bert

October 20, 2024

```
[ ]: !pip install transformers tensorflow datasets
```

Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.44.2)

Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-packages (2.17.0)

Collecting datasets

Downloading datasets-3.0.1-py3-none-any.whl.metadata (20 kB)

Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.16.1)

Requirement already satisfied: huggingface-hub<1.0,>=0.23.2 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.24.7)

Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.26.4)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (24.1)

Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.2)

Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.9.11)

Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.32.3)

Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.4.5)

Requirement already satisfied: tokenizers<0.20,>=0.19 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.19.1)

Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.66.5)

Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.4.0)

Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.6.3)

Requirement already satisfied: flatbuffers>=24.3.25 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (24.3.25)

Requirement already satisfied: gast!=0.5.0,!0.5.1,!0.5.2,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.6.0)

Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)

Requirement already satisfied: h5py>=3.10.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.11.0)

Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (18.1.1)

Requirement already satisfied: ml-dtypes<0.5.0,>=0.3.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.4.1)

Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.4.0)

Requirement already satisfied: protobuf!=4.21.0,!4.21.1,!4.21.2,!4.21.3,!4.21.4,!4.21.5,<5.0.0dev,>=3.20.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.20.3)

Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow) (75.1.0)

Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.0)

Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.5.0)

Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (4.12.2)

Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.0)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.64.1)

Requirement already satisfied: tensorboard<2.18,>=2.17 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.17.0)

Requirement already satisfied: keras>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.4.1)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.37.1)

Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (16.1.0)

Collecting dill<0.3.9,>=0.3.0 (from datasets)

 Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)

Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (2.2.2)

Collecting xxhash (from datasets)

 Downloading xxhash-3.5.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)

Collecting multiprocessing (from datasets)

 Downloading multiprocessing-0.70.17-py310-none-any.whl.metadata (7.2 kB)

Requirement already satisfied: fsspec<=2024.6.1,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from fsspec[http]<=2024.6.1,>=2023.1.0->datasets) (2024.6.1)

Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.10.10)

Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->tensorflow)

```

(0.44.0)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (2.4.3)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-
packages (from aiohttp->datasets) (24.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (6.1.0)
Requirement already satisfied: yarl<2.0,>=1.12.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.15.2)
Requirement already satisfied: async-timeout<5.0,>=4.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (4.0.3)
Requirement already satisfied: rich in /usr/local/lib/python3.10/dist-packages
(from keras>=3.2.0->tensorflow) (13.9.2)
Requirement already satisfied: namex in /usr/local/lib/python3.10/dist-packages
(from keras>=3.2.0->tensorflow) (0.0.8)
Requirement already satisfied: optree in /usr/local/lib/python3.10/dist-packages
(from keras>=3.2.0->tensorflow) (0.13.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(2024.8.30)
Requirement already satisfied: markdown>=2.6.8 in
/usr/local/lib/python3.10/dist-packages (from
tensorboard<2.18,>=2.17->tensorflow) (3.7)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
/usr/local/lib/python3.10/dist-packages (from
tensorboard<2.18,>=2.17->tensorflow) (0.7.2)
Requirement already satisfied: werkzeug>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from
tensorboard<2.18,>=2.17->tensorflow) (3.0.4)
INFO: pip is looking at multiple versions of multiprocess to determine which
version is compatible with other requirements. This could take a while.
  Downloading multiprocess-0.70.16-py310-none-any.whl.metadata (7.2 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->datasets) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-
packages (from pandas->datasets) (2024.2)

```

```

Requirement already satisfied: MarkupSafe>=2.1.1 in
/usr/local/lib/python3.10/dist-packages (from
werkzeug>=1.0.1->tensorboard<2.18,>=2.17->tensorflow) (3.0.1)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from
yarl<2.0,>=1.12.0->aiohttp->datasets) (0.2.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.10/dist-packages (from rich->keras>=3.2.0->tensorflow)
(3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.10/dist-packages (from rich->keras>=3.2.0->tensorflow)
(2.18.0)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-
packages (from markdown-it-py>=2.2.0->rich->keras>=3.2.0->tensorflow) (0.1.2)
Downloading datasets-3.0.1-py3-none-any.whl (471 kB)
471.6/471.6 kB
27.5 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
116.3/116.3 kB
11.1 MB/s eta 0:00:00
Downloading multiprocessing-0.70.16-py310-none-any.whl (134 kB)
134.8/134.8 kB
14.2 MB/s eta 0:00:00
Downloading
xxhash-3.5.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
194.1/194.1 kB
18.4 MB/s eta 0:00:00
Installing collected packages: xxhash, dill, multiprocessing, datasets
Successfully installed datasets-3.0.1 dill-0.3.8 multiprocessing-0.70.16
xxhash-3.5.0

```

```

[16]: import tensorflow as tf
from transformers import BertTokenizerFast, TFBertForQuestionAnswering,
      ↪ create_optimizer
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np
from sklearn.metrics import f1_score

```

0.1 Data Collection and Preprocessing:

```

[3]: # 1. Load the JSON data
json_data = pd.read_json('/content/dev-v1.1.json')

# 2. Use json_normalize to flatten the data
df = pd.json_normalize(
    json_data['data'],

```

```

record_path=['paragraphs', 'qas'],
meta=['title', ['paragraphs', 'context']],
errors='ignore'
)
display(df.head())

# count how many answers for each question by counting the number of answers in
↳ the 'answers' column
df['answer_count'] = df['answers'].apply(len)

print(df['answer_count'].value_counts())

```

```

                                answers \
0  [{'answer_start': 177, 'text': 'Denver Broncos...
1  [{'answer_start': 249, 'text': 'Carolina Panth...
2  [{'answer_start': 403, 'text': 'Santa Clara, C...
3  [{'answer_start': 177, 'text': 'Denver Broncos...
4  [{'answer_start': 488, 'text': 'gold'}], {'answ...

```

```

                                question \
0  Which NFL team represented the AFC at Super Bo...
1  Which NFL team represented the NFC at Super Bo...
2                                Where did Super Bowl 50 take place?
3                                Which NFL team won Super Bowl 50?
4  What color was used to emphasize the 50th anni...

```

```

                                id          title \
0  56be4db0acb8001400a502ec  Super_Bowl_50
1  56be4db0acb8001400a502ed  Super_Bowl_50
2  56be4db0acb8001400a502ee  Super_Bowl_50
3  56be4db0acb8001400a502ef  Super_Bowl_50
4  56be4db0acb8001400a502f0  Super_Bowl_50

```

```

                                paragraphs.context
0  Super Bowl 50 was an American football game to...
1  Super Bowl 50 was an American football game to...
2  Super Bowl 50 was an American football game to...
3  Super Bowl 50 was an American football game to...
4  Super Bowl 50 was an American football game to...

```

```

answer_count
3      8490
5      1147
4       759
2       136
6        35
1         3
Name: count, dtype: int64

```

```
[4]: # 3. Expand the dataframe to handle multiple answers
df['answer_count'] = df['answers'].apply(len)
df_expanded = df.explode('answers').reset_index(drop=True)

# Extract the start position of the answer
df_expanded['answer_start'] = df_expanded['answers'].apply(lambda ans: ans[
    ↪ans['answer_start']])

# Extract the answer text
df_expanded['answer'] = df_expanded['answers'].apply(lambda ans: ans['text'])

# Drop unnecessary columns
df_expanded = df_expanded.drop(columns=['answers', 'answer_count'])

# Drop duplicates and missing values
df_expanded = df_expanded.dropna(subset=['answer', 'question', 'answer_start'])

display(df_expanded.head())
```

	question \
0	Which NFL team represented the AFC at Super Bo...
1	Which NFL team represented the AFC at Super Bo...
2	Which NFL team represented the AFC at Super Bo...
3	Which NFL team represented the NFC at Super Bo...
4	Which NFL team represented the NFC at Super Bo...

	id	title \
0	56be4db0acb8001400a502ec	Super_Bowl_50
1	56be4db0acb8001400a502ec	Super_Bowl_50
2	56be4db0acb8001400a502ec	Super_Bowl_50
3	56be4db0acb8001400a502ed	Super_Bowl_50
4	56be4db0acb8001400a502ed	Super_Bowl_50

	paragraphs.context	answer_start \
0	Super Bowl 50 was an American football game to...	177
1	Super Bowl 50 was an American football game to...	177
2	Super Bowl 50 was an American football game to...	177
3	Super Bowl 50 was an American football game to...	249
4	Super Bowl 50 was an American football game to...	249

	answer
0	Denver Broncos
1	Denver Broncos
2	Denver Broncos
3	Carolina Panthers
4	Carolina Panthers

```
[5]: print(df_expanded.shape)
```

(34726, 6)

```
[6]: # 4. Prepare training and validation datasets

# take 2000 rows from df_expanded
df_expanded = df_expanded.head(2000)

# do validation and train split
train_df, val_df = train_test_split(df_expanded, test_size=0.2, random_state=42)

# save train_texts, train_contexts, train_starts, train_answers
train_texts = train_df['question'].tolist()
train_contexts = train_df['paragraphs.context'].tolist()
train_starts = train_df['answer_start'].tolist()
train_answers = train_df['answer'].tolist()

# save val_texts, val_contexts, val_starts, val_answers
val_texts = val_df['question'].tolist()
val_contexts = val_df['paragraphs.context'].tolist()
val_starts = val_df['answer_start'].tolist()
val_answers = val_df['answer'].tolist()

display(train_texts[:5])
display(train_contexts[:5])
display(train_starts[:5])
display(train_answers[:5])
```

```
['What was the final score of the game between the Broncos and Steelers?',
 'Which network broadcasted Super Bowl 50 in the U.S.?',
 'Who had the most rushing yards on the Broncos?',
 'How many touchdowns did Jonathan Stewart have in 13 games?',
 'When were the finalists announced?']
```

```
["The Broncos defeated the Pittsburgh Steelers in the divisional round, 23-16,
↳by scoring 11 points in the final three minutes of the game. They then beat
↳the defending Super Bowl XLIX champion New England Patriots in the AFC
↳Championship Game, 20-18, by intercepting a pass on New England's 2-point
↳conversion attempt with 17 seconds left on the clock. Despite Manning's
↳problems with interceptions during the season, he didn't throw any in their
↳two playoff games.",
 'CBS broadcast Super Bowl 50 in the U.S., and charged an average of $5 million
↳for a 30-second commercial during the game. The Super Bowl 50 halftime show
↳was headlined by the British rock group Coldplay with special guest performers
↳Beyoncé and Bruno Mars, who headlined the Super Bowl XLVII and Super Bowl
↳XLVIII halftime shows, respectively. It was the third-most watched U.S.
↳broadcast ever.',
```

"Manning finished the year with a career-low 67.9 passer rating, throwing for 2,249 yards and nine touchdowns, with 17 interceptions. In contrast, Osweiler threw for 1,967 yards, 10 touchdowns and six interceptions for a rating of 86.4. Veteran receiver Demaryius Thomas led the team with 105 receptions for 1,304 yards and six touchdowns, while Emmanuel Sanders caught 76 passes for 1,135 yards and six scores, while adding another 106 yards returning punts. Tight end Owen Daniels was also a big element of the passing game with 46 receptions for 517 yards. Running back C. J. Anderson was the team's leading rusher 863 yards and seven touchdowns, while also catching 25 passes for 183 yards. Running back Ronnie Hillman also made a big impact with 720 yards, five touchdowns, 24 receptions, and a 4.7 yards per carry average. Overall, the offense ranked 19th in scoring with 355 points and did not have any Pro Bowl selections.",

"The Panthers offense, which led the NFL in scoring (500 points), was loaded with talent, boasting six Pro Bowl selections. Pro Bowl quarterback Cam Newton had one of his best seasons, throwing for 3,837 yards and rushing for 636, while recording a career-high and league-leading 45 total touchdowns (35 passing, 10 rushing), a career-low 10 interceptions, and a career-best quarterback rating of 99.4. Newton's leading receivers were tight end Greg Olsen, who caught a career-high 77 passes for 1,104 yards and seven touchdowns, and wide receiver Ted Ginn, Jr., who caught 44 passes for 739 yards and 10 touchdowns; Ginn also rushed for 60 yards and returned 27 punts for 277 yards. Other key receivers included veteran Jerricho Cotchery (39 receptions for 485 yards), rookie Devin Funchess (31 receptions for 473 yards and five touchdowns), and second-year receiver Corey Brown (31 receptions for 447 yards). The Panthers backfield featured Pro Bowl running back Jonathan Stewart, who led the team with 989 rushing yards and six touchdowns in 13 games, along with Pro Bowl fullback Mike Tolbert, who rushed for 256 yards and caught 18 passes for another 154 yards. Carolina's offensive line also featured two Pro Bowl selections: center Ryan Kalil and guard Trai Turner.",

"The league announced on October 16, 2012, that the two finalists were Sun Life Stadium and Levi's Stadium. The South Florida/Miami area has previously hosted the event 10 times (tied for most with New Orleans), with the most recent one being Super Bowl XLIV in 2010. The San Francisco Bay Area last hosted in 1985 (Super Bowl XIX), held at Stanford Stadium in Stanford, California, won by the home team 49ers. The Miami bid depended on whether the stadium underwent renovations. However, on May 3, 2013, the Florida legislature refused to approve the funding plan to pay for the renovations, dealing a significant blow to Miami's chances."

[70, 0, 578, 1026, 24]

['23-16', 'CBS', 'Anderson', 'six', 'October 16, 2012']

```
[7]: tokenizer = BertTokenizerFast.from_pretrained('bert-base-uncased') #  
      ↪ Initialize tokenizer  
      model = TFBertForQuestionAnswering.from_pretrained('bert-base-uncased')
```



```

/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:89:
UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab
(https://huggingface.co/settings/tokens), set it as secret in your Google Colab
and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access
public models or datasets.

```

```

warnings.warn(

tokenizer_config.json: 0%|          | 0.00/48.0 [00:00<?, ?B/s]
vocab.txt: 0%|          | 0.00/232k [00:00<?, ?B/s]
tokenizer.json: 0%|          | 0.00/466k [00:00<?, ?B/s]
config.json: 0%|          | 0.00/570 [00:00<?, ?B/s]

```

```

/usr/local/lib/python3.10/dist-
packages/transformers/tokenization_utils_base.py:1601: FutureWarning:
`clean_up_tokenization_spaces` was not set. It will be set to `True` by default.
This behavior will be depracted in transformers v4.45, and will be then set to
`False` by default. For more details check this issue:
https://github.com/huggingface/transformers/issues/31884

```

```

warnings.warn(

model.safetensors: 0%|          | 0.00/440M [00:00<?, ?B/s]

```

All PyTorch model weights were used when initializing
TFBertForQuestionAnswering.

Some weights or buffers of the TF 2.0 model TFBertForQuestionAnswering were not
initialized from the PyTorch model and are newly initialized:
['qa_outputs.weight', 'qa_outputs.bias']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.

```

[8]: # 5. Tokenize the inputs (questions and contexts)
def encode_data(questions, contexts, answers, answer_starts):
    encodings = tokenizer(questions, contexts, truncation=False, padding=True,
    ↪return_tensors='tf', return_offsets_mapping=True,)
    start_positions = []
    end_positions = []

    for i, answer in enumerate(answers):
        answer_start = answer_starts[i]
        answer_end = answer_start + len(answer)

        # get offset_mapping
        offset_mapping = encodings.offset_mapping[i]

```

```

start_position = None
end_position = None

for j, (start, end) in enumerate(offset_mapping):
    if start <= answer_start and end >= answer_start:
        start_position = j
    if start <= answer_end and end >= answer_end:
        end_position = j
    # break if both start and end positions are found
    if start_position is not None and end_position is not None:
        break

# Debug output for first 5 answers
if i < 5:
    # Convert token IDs back to the original tokens for better
    ↪ visualization
    tokens = tokenizer.convert_ids_to_tokens(encodings['input_ids'][i])
    print(f"Answer: {answer}")
    print(f"Answer start: {answer_start}")
    print(f"Answer end: {answer_end}")
    print("Tokens:", tokens)
    print("Answer tokens:", tokens[start_position:end_position + 1])
    print()

    if start_position is None or end_position is None:
        start_positions.append(0)
        end_positions.append(0)
        continue

    start_positions.append(start_position)
    end_positions.append(end_position)

    encodings.update({'start_positions': tf.convert_to_tensor(start_positions),
    ↪ 'end_positions': tf.convert_to_tensor(end_positions)})
    return encodings

# Encode training and validation datasets
train_encodings = encode_data(train_texts, train_contexts, train_answers,
    ↪ train_starts)
val_encodings = encode_data(val_texts, val_contexts, val_answers, val_starts)

```

Answer: 23-16

Answer start: 70

Answer end: 75

Tokens: ['[CLS]', 'what', 'was', 'the', 'final', 'score', 'of', 'the', 'game',

[illegible]

Answer tokens: ['23', '-', '16']

Answer: CBS

Answer start: 0

Answer end: 3

[illegible]


```

' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ',
' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ',
' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ',
' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ',
' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ',
' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ', ' [PAD] ',
' [PAD] ', ' [PAD] ', ' [PAD] '
tf.Tensor(27, shape=(), dtype=int32)
tf.Tensor(29, shape=(), dtype=int32)

```

```

[11]: # 6. Convert to TensorFlow Dataset
train_dataset = tf.data.Dataset.from_tensor_slices((
    {
        'input_ids': train_encodings['input_ids'],
        'attention_mask': train_encodings['attention_mask']
    },
    {
        'start_positions': train_encodings['start_positions'],
        'end_positions': train_encodings['end_positions']
    }
))

val_dataset = tf.data.Dataset.from_tensor_slices((
    {
        'input_ids': val_encodings['input_ids'],
        'attention_mask': val_encodings['attention_mask']
    },
    {
        'start_positions': val_encodings['start_positions'],
        'end_positions': val_encodings['end_positions']
    }
))

# 7. Batch the datasets
train_dataset = train_dataset.batch(8)
val_dataset = val_dataset.batch(8)

```

1 Model Selection and Training

```

[12]: # 8. Define optimizer and learning rate schedule
num_train_steps = len(train_dataset) * 5
optimizer, schedule = create_optimizer(
    init_lr=2e-5,
    num_warmup_steps=0,
    num_train_steps=num_train_steps
)

```

```

# 9. Compile the model with the optimizer and loss function
loss = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)
model.compile(optimizer=optimizer, loss=loss, metrics=['accuracy'])

# 10. Train the model
model.fit(
    train_dataset,
    validation_data=val_dataset,
    epochs=5
)

# 11. Save the model and tokenizer
model.save_pretrained('./saved_qa_model')
tokenizer.save_pretrained('./saved_qa_model')

```

Epoch 1/5

200/200 [=====] - 72s 166ms/step - loss: 7.1991 - end_logits_loss: 3.5773 - start_logits_loss: 3.6218 - end_logits_accuracy: 0.1838 - start_logits_accuracy: 0.1419 - val_loss: 5.1766 - val_end_logits_loss: 2.5950 - val_start_logits_loss: 2.5815 - val_end_logits_accuracy: 0.3375 - val_start_logits_accuracy: 0.3100

Epoch 2/5

200/200 [=====] - 25s 123ms/step - loss: 3.8111 - end_logits_loss: 1.9047 - start_logits_loss: 1.9063 - end_logits_accuracy: 0.5200 - start_logits_accuracy: 0.4825 - val_loss: 3.0900 - val_end_logits_loss: 1.5162 - val_start_logits_loss: 1.5738 - val_end_logits_accuracy: 0.6350 - val_start_logits_accuracy: 0.5825

Epoch 3/5

200/200 [=====] - 25s 123ms/step - loss: 2.2717 - end_logits_loss: 1.1293 - start_logits_loss: 1.1425 - end_logits_accuracy: 0.7013 - start_logits_accuracy: 0.6619 - val_loss: 2.6655 - val_end_logits_loss: 1.3141 - val_start_logits_loss: 1.3514 - val_end_logits_accuracy: 0.7050 - val_start_logits_accuracy: 0.6650

Epoch 4/5

200/200 [=====] - 25s 123ms/step - loss: 1.6970 - end_logits_loss: 0.8346 - start_logits_loss: 0.8624 - end_logits_accuracy: 0.7700 - start_logits_accuracy: 0.7281 - val_loss: 2.5023 - val_end_logits_loss: 1.2554 - val_start_logits_loss: 1.2468 - val_end_logits_accuracy: 0.7375 - val_start_logits_accuracy: 0.6750

Epoch 5/5

200/200 [=====] - 25s 123ms/step - loss: 1.4248 - end_logits_loss: 0.6987 - start_logits_loss: 0.7261 - end_logits_accuracy: 0.8025 - start_logits_accuracy: 0.7494 - val_loss: 2.4196 - val_end_logits_loss: 1.2001 - val_start_logits_loss: 1.2195 - val_end_logits_accuracy: 0.7350 - val_start_logits_accuracy: 0.6775


```
[12]: ( './saved_qa_model/tokenizer_config.json',
        './saved_qa_model/special_tokens_map.json',
        './saved_qa_model/vocab.txt',
        './saved_qa_model/added_tokens.json',
        './saved_qa_model/tokenizer.json')
```

1.1 Model Evaluation

```
[14]: # 12. test the model
model = TFBertForQuestionAnswering.from_pretrained('./saved_qa_model')
tokenizer = BertTokenizerFast.from_pretrained('./saved_qa_model')

example_context = "Architecturally, the school has a Catholic character. Atop
↳the Main Building's gold dome is a golden statue of the Virgin Mary.
↳Immediately in front of the Main Building and facing it, is a copper statue
↳of Christ with arms upraised with the legend \"Venite Ad Me Omnes\". Next to
↳the Main Building is the Basilica of the Sacred Heart. Immediately behind
↳the basilica is the Grotto, a Marian place of prayer and reflection. It is a
↳replica of the grotto at Lourdes, France where the Virgin Mary reputedly
↳appeared to Saint Bernadette Soubirous in 1858. At the end of the main drive
↳(and in a direct line that connects through 3 statues and the Gold Dome), is
↳a simple, modern stone statue of Mary."
question = "What is atop the Main Building's gold dome?"

inputs = tokenizer(question, example_context, return_tensors='tf')
outputs = model(inputs)

answer_start = tf.argmax(outputs.start_logits, axis=1).numpy()[0]
answer_end = tf.argmax(outputs.end_logits, axis=1).numpy()[0]

answer = tokenizer.convert_tokens_to_string(tokenizer.
↳convert_ids_to_tokens(inputs['input_ids'][0][answer_start:answer_end+1]))

print(f"Question: {question}")
print(f"Answer: {answer}")
```

All model checkpoint layers were used when initializing TFBertForQuestionAnswering.

All the layers of TFBertForQuestionAnswering were initialized from the model checkpoint at ./saved_qa_model.

If your task is similar to the task the model of the checkpoint was trained on, you can already use TFBertForQuestionAnswering for predictions without further training.

Question: What is atop the Main Building's gold dome?
 Answer: golden statue of the virgin mary

```
[17]: # 12. Evaluate the model
def evaluate_model(model, tokenizer, val_texts, val_contexts, val_answers, val_starts):
    total = len(val_texts)
    exact_match = 0
    f1 = 0.0

    for i in range(total):
        question = val_texts[i]
        context = val_contexts[i]
        true_answer = val_answers[i]

        # Tokenize question and context
        inputs = tokenizer(question, context, return_tensors='tf')
        outputs = model(inputs)

        # Get predicted answer start and end positions
        answer_start = tf.argmax(outputs.start_logits, axis=1).numpy()[0]
        answer_end = tf.argmax(outputs.end_logits, axis=1).numpy()[0]

        # Extract the predicted answer
        predicted_answer = tokenizer.convert_tokens_to_string(
            tokenizer.convert_ids_to_tokens(inputs['input_ids'][0][answer_start:
            answer_end+1]))

        # Remove extra spaces in both true and predicted answers
        true_answer = true_answer.strip()
        predicted_answer = predicted_answer.strip()

        # Calculate exact match
        if true_answer == predicted_answer:
            exact_match += 1

        # Calculate F1 score
        true_tokens = set(true_answer.split())
        predicted_tokens = set(predicted_answer.split())

        common_tokens = true_tokens & predicted_tokens
        if len(common_tokens) == 0:
            f1 += 0
        else:
            precision = len(common_tokens) / len(predicted_tokens)
            recall = len(common_tokens) / len(true_tokens)
            f1 += 2 * (precision * recall) / (precision + recall)

    exact_match_score = exact_match / total * 100
```

```

f1_score_value = f1 / total * 100

print(f"Exact Match: {exact_match_score:.2f}%")
print(f"F1 Score: {f1_score_value:.2f}%")
return exact_match_score, f1_score_value

# Call the evaluate function
exact_match_score, f1_score_value = evaluate_model(model, tokenizer, val_texts,
↪val_contexts, val_answers, val_starts)

```

Exact Match: 19.00%

F1 Score: 25.22%

1.2 Summary

- We trained a BERT-based model for question answering using a subset of 2,000 data points from the SQuAD dataset.
- The data was preprocessed to extract relevant question-answer pairs:
 - We tokenized both the questions and contexts using the BertTokenizerFast
 - For training, we used TensorFlow’s TFBertForQuestionAnswering model and compiled it with an Adam optimizer and a sparse categorical cross-entropy loss function.
- After training for 5 epochs, we evaluated the model on the validation set using Exact Match (EM) and F1 score metrics.

The results on the validation set were as follows:

- **Exact Match (EM):** 19.00%
- **F1 Score:** 25.22%

1.2.1 Analysis of Results:

The Exact Match (EM) score of 19.00% and F1 score of 25.22% are relatively low, which suggests that the model struggles to correctly identify the precise span of the answers.

There could be several potential reasons for this:

- **Data Size:** We trained the model on a small subset of 2,000 data points. BERT models typically require larger datasets to capture the nuances of language and provide accurate predictions.
- **Tokenization Misalignment:** The tokenizer might have difficulties precisely mapping token offsets, especially for longer or more complex answers. Small misalignments in token boundaries can significantly affect the EM score.
- **Model Hyperparameters:** The learning rate and the number of training epochs could have been suboptimal. Additionally, using more advanced learning rate schedules or warmup steps might improve performance.
- **Truncation Issues:** We set truncation to False during tokenization, which could lead to the context being too long for the model’s input size. This could result in the model missing relevant parts of the context, leading to lower accuracy.

1.2.2 Future Improvements:

To improve the performance of the model, we could consider the following steps:

- **Increase the Dataset Size:** Training on a larger dataset would help the model generalize better and capture more varied patterns in the data.
- **Fine-tuning Hyperparameters:** We can further experiment with different learning rates, batch sizes, and epochs. Additionally, introducing learning rate warmup and weight decay could improve optimization.
- **Tokenization Refinement:** Adjusting the tokenization process, including the use of truncation and padding, might help the model handle long contexts more effectively. We could also explore different strategies for handling token offset mapping more accurately.
- **Data Augmentation:** Using data augmentation techniques to increase the diversity and size of the training dataset could help improve the model's robustness.

With these refinements, we expect the model's ability to predict accurate answer spans and improve its overall performance on the task.

1.3 Google Collab Utils

```
[18]: import shutil

# Create a zip file of the saved model directory
shutil.make_archive('saved_qa_model', 'zip', './saved_qa_model')
```

```
[18]: '/content/saved_qa_model.zip'
```

```
[19]: from google.colab import files

# Download the zip file
files.download('saved_qa_model.zip')
```

<IPython.core.display.Javascript object>

<IPython.core.display.Javascript object>

```
[20]: !apt-get install -y texlive-xetex texlive-fonts-recommended_
      ↪texlive-plain-generic
      !pip install nbconvert
```

Reading package lists... Done

Building dependency tree... Done

Reading state information... Done

The following additional packages will be installed:

dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre

fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-java

```

libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53
libtexluajit2 libwoff1
libzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre
texlive-base texlive-binaries texlive-latex-base texlive-latex-extra texlive-
latex-recommended
texlive-pictures tipa xfonts-encodings xfonts-utils
Suggested packages:
fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java poppler-
utils ghostscript
fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic | fonts-
ipafont-gothic
fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper
gv
| postscript-viewer perl-tk xpdf | pdf-viewer xzdec texlive-fonts-recommended-
doc
texlive-latex-base-doc python3-pygments icc-profiles libfile-which-perl
libspreadsheet-parseexcel-perl texlive-latex-extra-doc texlive-latex-
recommended-doc
texlive-luatex texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
default-jre-headless
tipa-doc
The following NEW packages will be installed:
dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-
texgyre
fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-
java
libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53
libtexluajit2 libwoff1
libzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre
texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra
texlive-latex-recommended texlive-pictures texlive-plain-generic texlive-xetex
tipa
xfonts-encodings xfonts-utils
0 upgraded, 54 newly installed, 0 to remove and 49 not upgraded.
Need to get 182 MB of archives.
After this operation, 571 MB of additional disk space will be used.

```

Get:1 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-droid-fallback all 1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-lato all 2.0-2.1 [2,696 kB]
Get:3 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 poppler-data all 0.4.11-1 [2,171 kB]
Get:4 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 tex-common all 6.17 [33.7 kB]
Get:5 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-urw-base35 all 20200910-1 [6,367 kB]
Get:6 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libgs9-common all 9.55.0~dfsg1-0ubuntu5.9 [752 kB]
Get:7 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libidn12 amd64 1.38-4ubuntu1 [60.0 kB]
Get:8 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libijs-0.35 amd64 0.35-15build2 [16.5 kB]
Get:9 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libjbig2dec0 amd64 0.19-3build2 [64.7 kB]
Get:10 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libgs9 amd64 9.55.0~dfsg1-0ubuntu5.9 [5,033 kB]
Get:11 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libkpathsea6 amd64 2021.20210626.59705-1ubuntu0.2 [60.4 kB]
Get:12 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libwoff1 amd64 1.0.2-1build4 [45.2 kB]
Get:13 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 dvisvgm amd64 2.13.1-1 [1,221 kB]
Get:14 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-lmodern all 2.004.5-6.1 [4,532 kB]
Get:15 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-noto-mono all 20201225-1build1 [397 kB]
Get:16 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-texgyre all 20180621-3.1 [10.2 MB]
Get:17 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libapache-pom-java all 18-1 [4,720 B]
Get:18 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-parent-java all 43-1 [10.8 kB]
Get:19 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-logging-java all 1.2-2 [60.3 kB]
Get:20 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libfontenc1 amd64 1:1.1.4-1build3 [14.7 kB]
Get:21 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libptexenc1 amd64 2021.20210626.59705-1ubuntu0.2 [39.1 kB]
Get:22 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rubygems-integration all 1.18 [5,336 B]
Get:23 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby3.0 amd64 3.0.2-7ubuntu2.7 [50.1 kB]
Get:24 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-rubygems all 3.3.5-2 [228 kB]

Get:25 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby amd64 1:3.0~exp1
[5,100 B]
Get:26 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rake all 13.0.6-2 [61.7
kB]
Get:27 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:28 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-webrick
all 1.7.0-3ubuntu0.1 [52.1 kB]
Get:29 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-xmldrpc all
0.3.2-1ubuntu0.1 [24.9 kB]
Get:30 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libruby3.0
amd64 3.0.2-7ubuntu2.7 [5,113 kB]
Get:31 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libsynchronet2
amd64 2021.20210626.59705-1ubuntu0.2 [55.6 kB]
Get:32 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libteckit0 amd64
2.5.11+ds1-1 [421 kB]
Get:33 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexlua53
amd64 2021.20210626.59705-1ubuntu0.2 [120 kB]
Get:34 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexluajit2
amd64 2021.20210626.59705-1ubuntu0.2 [267 kB]
Get:35 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libzip-0-13 amd64
0.13.72+dfsg.1-1.1 [27.0 kB]
Get:36 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-encodings all
1:1.0.5-0ubuntu2 [578 kB]
Get:37 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-utils amd64
1:7.7+6build2 [94.6 kB]
Get:38 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 lmodern all
2.004.5-6.1 [9,471 kB]
Get:39 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:40 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:41 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:42 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:43 <http://archive.ubuntu.com/ubuntu> jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.2 [9,860 kB]
Get:44 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:45 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:46 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:47 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:48 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]

```

Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 182 MB in 16s (11.1 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages ...
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 123629 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Selecting previously unselected package fonts-lato.
Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
Unpacking fonts-lato (2.0-2.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
Unpacking poppler-data (0.4.11-1) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.17_all.deb ...
Unpacking tex-common (6.17) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
Unpacking fonts-urw-base35 (20200910-1) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.9_all.deb ...
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.9) ...
Selecting previously unselected package libidn12:amd64.
Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15build2) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.9_amd64.deb ...
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.9) ...
Selecting previously unselected package libkpathsea6:amd64.

```



```

Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build4) ...
Selecting previously unselected package dvisvgm.
Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
Unpacking dvisvgm (2.13.1-1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
Unpacking fonts-lmodern (2.004.5-6.1) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
Unpacking fonts-noto-mono (20201225-1build1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...
Unpacking fonts-texgyre (20180621-3.1) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../17-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../18-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../19-libfontenc1_1%3a1.1.4-1build3_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../20-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../21-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...
Selecting previously unselected package ruby3.0.
Preparing to unpack .../22-ruby3.0_3.0.2-7ubuntu2.7_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.7) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../23-ruby-rubygems_3.3.5-2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../24-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../25-rake_13.0.6-2_all.deb ...

```

```

Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../26-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../27-ruby-webrick_1.7.0-3ubuntu0.1_all.deb ...
Unpacking ruby-webrick (1.7.0-3ubuntu0.1) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../28-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../29-libruby3.0_3.0.2-7ubuntu2.7_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.7) ...
Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../30-libsyntax2_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libsyntax2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../31-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../32-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
.../33-libtexluajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../34-libzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../35-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../36-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../37-lmodern_2.004.5-6.1_all.deb ...
Unpacking lmodern (2.004.5-6.1) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../38-preview-latex-style_12.2-1ubuntu1_all.deb ...
Unpacking preview-latex-style (12.2-1ubuntu1) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../39-t1utils_1.41-4build2_amd64.deb ...
Unpacking t1utils (1.41-4build2) ...
Selecting previously unselected package teckit.
Preparing to unpack .../40-teckit_2.5.11+ds1-1_amd64.deb ...

```

```

Unpacking teckit (2.5.11+ds1-1) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../41-tex-gyre_20180621-3.1_all.deb ...
Unpacking tex-gyre (20180621-3.1) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../42-texlive-
binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../43-texlive-base_2021.20220204-1_all.deb ...
Unpacking texlive-base (2021.20220204-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../44-texlive-fonts-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-fonts-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../45-texlive-latex-base_2021.20220204-1_all.deb ...
Unpacking texlive-latex-base (2021.20220204-1) ...
Selecting previously unselected package libfontbox-java.
Preparing to unpack .../46-libfontbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libfontbox-java (1:1.8.16-2) ...
Selecting previously unselected package libpdfbox-java.
Preparing to unpack .../47-libpdfbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libpdfbox-java (1:1.8.16-2) ...
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack .../48-texlive-latex-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-latex-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../49-texlive-pictures_2021.20220204-1_all.deb ...
Unpacking texlive-pictures (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../50-texlive-latex-extra_2021.20220204-1_all.deb ...
Unpacking texlive-latex-extra (2021.20220204-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../51-texlive-plain-generic_2021.20220204-1_all.deb ...
Unpacking texlive-plain-generic (2021.20220204-1) ...
Selecting previously unselected package tipa.
Preparing to unpack .../52-tipa_2%3a1.3-21_all.deb ...
Unpacking tipa (2:1.3-21) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../53-texlive-xetex_2021.20220204-1_all.deb ...
Unpacking texlive-xetex (2021.20220204-1) ...
Setting up fonts-lato (2.0-2.1) ...
Setting up fonts-noto-mono (20201225-1build1) ...
Setting up libwoff1:amd64 (1.0.2-1build4) ...
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libijs-0.35:amd64 (0.35-15build2) ...
Setting up libtexluaajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libfontbox-java (1:1.8.16-2) ...

```

```

Setting up rubygems-integration (1.18) ...
Setting up libzzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Setting up fonts-urw-base35 (20200910-1) ...
Setting up poppler-data (0.4.11-1) ...
Setting up tex-common (6.17) ...
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-1build3) ...
Setting up libjbig2dec0:amd64 (0.19-3build2) ...
Setting up libteckit0:amd64 (2.5.11+ds1-1) ...
Setting up libapache-pom-java (18-1) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up t1utils (1.41-4build2) ...
Setting up libidn12:amd64 (1.38-4ubuntu1) ...
Setting up fonts-texgyre (20180621-3.1) ...
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up ruby-webrick (1.7.0-3ubuntu0.1) ...
Setting up fonts-lmodern (2.004.5-6.1) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Setting up libsynchronet2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.9) ...
Setting up teckit (2.5.11+ds1-1) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.9) ...
Setting up preview-latex-style (12.2-1ubuntu1) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.13.1-1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) ...
Setting up texlive-base (2021.20220204-1) ...
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps

```

```

tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
Setting up tex-gyre (20180621-3.1) ...
Setting up texlive-plain-generic (2021.20220204-1) ...
Setting up texlive-latex-base (2021.20220204-1) ...
Setting up texlive-latex-recommended (2021.20220204-1) ...
Setting up texlive-pictures (2021.20220204-1) ...
Setting up texlive-fonts-recommended (2021.20220204-1) ...
Setting up tipa (2:1.3-21) ...
Setting up texlive-latex-extra (2021.20220204-1) ...
Setting up texlive-xetex (2021.20220204-1) ...
Setting up rake (13.0.6-2) ...
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.7) ...
Setting up ruby3.0 (3.0.2-7ubuntu2.7) ...
Setting up ruby (1:3.0~exp1) ...
Setting up ruby-rubygems (3.3.5-2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.4) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

Processing triggers for tex-common (6.17) ...
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.
    This may take some time... done.

```

Requirement already satisfied: nbconvert in /usr/local/lib/python3.10/dist-packages (6.5.4)

Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert) (4.9.4)

Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (4.12.3)

Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert) (6.1.0)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.7.1)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.4)

Requirement already satisfied: jinja2>=3.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (3.1.4)

Requirement already satisfied: jupyter-core>=4.7 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (5.7.2)

Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.3.0)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (3.0.1)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.8.4)

Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.10.0)

Requirement already satisfied: nbformat>=5.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (5.10.4)

Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from nbconvert) (24.1)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (1.5.1)

Requirement already satisfied: pygments>=2.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (2.18.0)

Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (1.3.0)

Requirement already satisfied: traitlets>=5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert) (5.7.1)

Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.7->nbconvert) (4.3.6)

Requirement already satisfied: jupyter-client>=6.1.12 in /usr/local/lib/python3.10/dist-packages (from nbclient>=0.5.0->nbconvert) (6.1.12)

Requirement already satisfied: fastjsonschema>=2.15 in /usr/local/lib/python3.10/dist-packages (from nbformat>=5.1->nbconvert) (2.20.0)

Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.10/dist-packages (from nbformat>=5.1->nbconvert) (4.23.0)

Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert) (2.6)

Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert) (1.16.0)

Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert) (0.5.1)

Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=5.1->nbconvert) (24.2.0)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=5.1->nbconvert) (2024.10.1)

Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=5.1->nbconvert) (0.35.1)

Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=5.1->nbconvert) (0.20.0)

Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.10/dist-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (24.0.1)

Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.10/dist-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (2.8.2)

Requirement already satisfied: tornado>=4.1 in /usr/local/lib/python3.10/dist-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (6.3.3)

```
[21]: !jupyter nbconvert --to pdf standford_qa_bert.ipynb
```

```
[NbConvertApp] WARNING | pattern 'standford_qa_bert.ipynb' matched no files
This application is used to convert notebook files (*.ipynb)
to various other formats.
```

```
WARNING: THE COMMANDLINE INTERFACE MAY CHANGE IN FUTURE RELEASES.
```

Options

```
=====
```

The options below are convenience aliases to configurable class-options, as listed in the "Equivalent to" description-line of the aliases.

To see all configurable class-options for some <cmd>, use:

```
<cmd> --help-all
```

--debug

```
set log level to logging.DEBUG (maximize logging output)
```

```
Equivalent to: [--Application.log_level=10]
```

--show-config

```
Show the application's configuration (human-readable format)
```

```
Equivalent to: [--Application.show_config=True]
```

--show-config-json

```
Show the application's configuration (json format)
```

```
Equivalent to: [--Application.show_config_json=True]
```

--generate-config

```
generate default config file
```

Equivalent to: [--JupyterApp.generate_config=True]

-y
Answer yes to any questions instead of prompting.
Equivalent to: [--JupyterApp.answer_yes=True]

--execute
Execute the notebook prior to export.
Equivalent to: [--ExecutePreprocessor.enabled=True]

--allow-errors
Continue notebook execution even if one of the cells throws an error and include the error message in the cell output (the default behaviour is to abort conversion). This flag is only relevant if '--execute' was specified, too.
Equivalent to: [--ExecutePreprocessor.allow_errors=True]

--stdin
read a single notebook file from stdin. Write the resulting notebook with default basename 'notebook.*'
Equivalent to: [--NbConvertApp.from_stdin=True]

--stdout
Write notebook output to stdout instead of files.
Equivalent to: [--NbConvertApp.writer_class=StdoutWriter]

--inplace
Run nbconvert in place, overwriting the existing notebook (only relevant when converting to notebook format)
Equivalent to: [--NbConvertApp.use_output_suffix=False
--NbConvertApp.export_format=notebook --FilesWriter.build_directory=]

--clear-output
Clear output of current file and save in place, overwriting the existing notebook.
Equivalent to: [--NbConvertApp.use_output_suffix=False
--NbConvertApp.export_format=notebook --FilesWriter.build_directory=
--ClearOutputPreprocessor.enabled=True]

--no-prompt
Exclude input and output prompts from converted document.
Equivalent to: [--TemplateExporter.exclude_input_prompt=True
--TemplateExporter.exclude_output_prompt=True]

--no-input
Exclude input cells and output prompts from converted document.
This mode is ideal for generating code-free reports.
Equivalent to: [--TemplateExporter.exclude_output_prompt=True
--TemplateExporter.exclude_input=True
--TemplateExporter.exclude_input_prompt=True]

--allow-chromium-download
Whether to allow downloading chromium if no suitable version is found on the system.
Equivalent to: [--WebPDFExporter.allow_chromium_download=True]

--disable-chromium-sandbox
Disable chromium security sandbox when converting to PDF..
Equivalent to: [--WebPDFExporter.disable_sandbox=True]

--show-input

Shows code input. This flag is only useful for dejavu users.
 Equivalent to: [--TemplateExporter.exclude_input=False]

--embed-images
 Embed the images as base64 dataurls in the output. This flag is only useful for the HTML/WebPDF/Slides exports.
 Equivalent to: [--HTMLExporter.embed_images=True]

--sanitize-html
 Whether the HTML in Markdown cells and cell outputs should be sanitized..
 Equivalent to: [--HTMLExporter.sanitize_html=True]

--log-level=<Enum>
 Set the log level by value or name.
 Choices: any of [0, 10, 20, 30, 40, 50, 'DEBUG', 'INFO', 'WARN', 'ERROR', 'CRITICAL']
 Default: 30
 Equivalent to: [--Application.log_level]

--config=<Unicode>
 Full path of a config file.
 Default: ''
 Equivalent to: [--JupyterApp.config_file]

--to=<Unicode>
 The export format to be used, either one of the built-in formats
 ['asciidoc', 'custom', 'html', 'latex', 'markdown', 'notebook', 'pdf', 'python', 'rst', 'script', 'slides', 'webpdf']
 or a dotted object name that represents the import path for an
 ``Exporter`` class
 Default: ''
 Equivalent to: [--NbConvertApp.export_format]

--template=<Unicode>
 Name of the template to use
 Default: ''
 Equivalent to: [--TemplateExporter.template_name]

--template-file=<Unicode>
 Name of the template file to use
 Default: None
 Equivalent to: [--TemplateExporter.template_file]

--theme=<Unicode>
 Template specific theme(e.g. the name of a JupyterLab CSS theme distributed as prebuilt extension for the lab template)
 Default: 'light'
 Equivalent to: [--HTMLExporter.theme]

--sanitize_html=<Bool>
 Whether the HTML in Markdown cells and cell outputs should be sanitized.This should be set to True by nbviewer or similar tools.
 Default: False
 Equivalent to: [--HTMLExporter.sanitize_html]

--writer=<DottedObjectName>
 Writer class used to write the
 results of the conversion


```
> jupyter nbconvert mynotebook.ipynb --to html
```

Options include ['asciidoc', 'custom', 'html', 'latex', 'markdown', 'notebook', 'pdf', 'python', 'rst', 'script', 'slides', 'webpdf'].

```
> jupyter nbconvert --to latex mynotebook.ipynb
```

Both HTML and LaTeX support multiple output templates. LaTeX includes

'base', 'article' and 'report'. HTML includes 'basic', 'lab' and 'classic'. You can specify the flavor of the format used.

```
> jupyter nbconvert --to html --template lab mynotebook.ipynb
```

You can also pipe the output to stdout, rather than a file

```
> jupyter nbconvert mynotebook.ipynb --stdout
```

PDF is generated via latex

```
> jupyter nbconvert mynotebook.ipynb --to pdf
```

You can get (and serve) a Reveal.js-powered slideshow

```
> jupyter nbconvert myslides.ipynb --to slides --post serve
```

Multiple notebooks can be given at the command line in a couple of different ways:

```
> jupyter nbconvert notebook*.ipynb
```

```
> jupyter nbconvert notebook1.ipynb notebook2.ipynb
```

or you can specify the notebooks list in a config file, containing::

```
c.NbConvertApp.notebooks = ["my_notebook.ipynb"]
```

```
> jupyter nbconvert --config mycfg.py
```

To see all available configurables, use `--help-all`.

```
[ ]: from google.colab import files
files.download('standford_qa_bert.pdf')
```