# DistilBERT model download link:

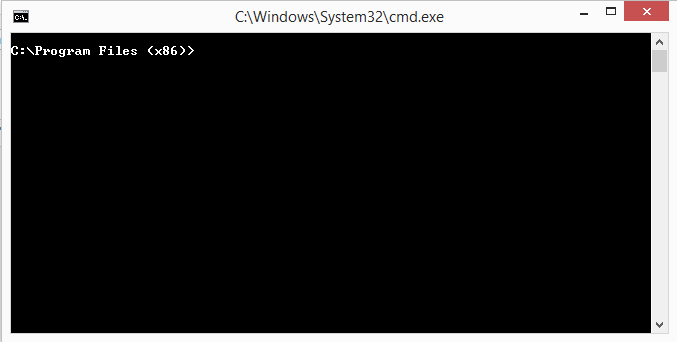
https://drive.google.com/file/d/1QSXinQ1d10L-qQdqjnkmgjNFFIzGgOd9/view?usp=sharing

# cdQA-annotator

In order to facilitate the data annotation, a web-based application, the [cdQA-annotator](https://github.com/cdqa-suite/cdQA-annotator) is used.

In order to use it, you should have your dataset transformed to a JSON file with SQuAD-like format:

from cdqa.utils.converters import df2squad  
json\_data = df2squad(df=df, squad\_version='v1.1', output\_dir='.', filename='dataset-name.json')

Installing the annotator and running it:

Open Command Prompt:

Run the following commands:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*ONE TIME INSTALLATION\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Comment --> # Clone the repo  
git clone <https://github.com/cdqa-suite/cdQA-annotator>

Comment --> # Install dependencies  
cd cdQA-annotator  
npm install

\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\

Note: if npm is not recognised, install nodejs.

Process to be followed:

Go to -> https://nodejs.org/en/download/

Download the applicable version (Windows -> .msi, Linux -> x64)

Install it on the local system (not google colab)

Close the terminal and reopen terminal.

This will also install npm.

now npm should work.

\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*EXECUTE TO USE ANNOTATOR\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Comment --> # Start development server

npm run serve

Open http://localhost:8080/ in your browser

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Upload the json file created using dataframe.

Write a question and mark the answer from the uploaded data.

Once all the paragraphs are done, download the file (say filename “cdqa.json”)

Note: The downloaded file has paragraph, questions and respective answers. It will be in json format.

This downloaded file is further uploaded to colab for training the BERT model.

