

## Assignment 3: Question Answering via Reading Comprehension

Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable. It is a well-known NLP competition, and you can check it out via <https://rajpurkar.github.io/SQuAD-explorer/>. The top ranked solutions are also available here. For a quick review, you can read [this introduction](#).

In this assignment, you will have a chance to design your own question-answering model by using NLP + Deep Learning approach. Before you start, please check this video tutorial very carefully: [Applying BERT to Question Answering \(SQuAD v1.1\)](#) and the related document + code can be found at <https://colab.research.google.com/drive/1uSIWtJdZmLrI3FCNIIUHFxwAJiSu2JO->

You can download the dataset [here](#) or [here](#). Using a small subset is acceptable.

Download a copy of the dataset (distributed under the CC BY-SA 4.0 license):

Training Set v2.0 (40 MB)

Dev Set v2.0 (4 MB)

### Task 1: Implement a question-answering model

1. Please find the official evaluation script [here](#). You only need to report a final F1 score of the model.
2. Optional. Follow the [tutorial](#) and submit your model and results to the official SQuAD leaderboard. (This is good for your resume.)

### Task 2: Unanswerable question detection

Some questions can be unanswerable. For instance, the reference text is focusing on Massachusetts demographics information, but the question is “What’s the population of Indiana?” Then, this question can be answerable. Please read this paper for [SQUADRUN](#) dataset.

In this task, please propose a model to detect those unanswerable questions, e.g., binary classification. Note that you don’t need to implement your model. However, please give me your hypothesis, model design, and key parameter settings (if there is any). You can review related works, but I would like to read your own design.

Please provide **detail** of your proposed model, an overall design figure is recommended.

**Task 3 [optional: 15 extra credit]:** Implement your design from Task 2, and give me algorithm evaluation scores.