

Text prediction app documentation - Usage

Ivy Woo

16 February 2021

1 In-/Output

Input:

- Select the number of predictions of the next word desired. (1-5 available, default is 4.)
- Type some text in the input box provided.
- Make a blank space at the end of your text to indicate the completion of input. Without a blank space, the app treats the last word as incomplete and does not take it into account for prediction.

For example, both inputs “This is a gre” and “This is a great” would be read by the app as “This is” only. Inputting “This is a great ” (with a blank space at the end) would guide the app to read all three words completely and make prediction according.

- Upper and lower case characters are not differentiated by the app. For example, “I Have” and “i have” will be treated in the same way by the app.

Output:

- The n predictions of the next word will be displayed, where $n \in \{1, 2, \dots, 5\}$ is chosen by the user.
- Currently only lower case letters will be return, e.g., “i” instead of “I”, “susan” instead of “Susan”.

2 Support

- The following characters are supported:
 - the 26 English alphabets,
 - the punctuation comma (“,”), and
 - the punctuation full stop (“.”) which indicates an end of sentence.

Each punctuation comma or full stop is treated as one “word”. Hence it is possible that one prediction is simply a punctuation.

- The punctuation full stop is not supported when used in abbreviations, such as “Mr.”, “i.e.” and “e.g.”.
- Any input other than those specified above cannot be recognized by the app and a warning message will pop out, although users can continue with the typing.
- The app is intended to make prediction which continue an incomplete sentence. A full stop at the end of the input will confuse the app and a warning message will pop out, although users can continue with the typing.

3 Scope

Text prediction in this flavour is highly dependent on the dataset used for training the model. This app is trained using text chunks from U.S. English blogs, news and twitter posts. As a result, the app will function properly for predictions on daily-used English words/sentence/phrase.

For example, “how are” will return “you” (and also “u” because this is frequently seen in online informal text especially in twitter), “the financial” will return words like “times” and “crisis”, and “from top to” will return “bottom”.

On the other hand, technical terms from a specific discipline can general not be predicted. For instance, typing “Vandermonde” will not result in predictions like “matrix” or “determinant”, as no text from the field of linear algebra is used to train the model. Similarly, do not expect the app to return “field” or “ring” with the input “cyclotomic” because there is no data from the field of number theory.

Further, note that human’s texting habit, including the usage of words, changes with time. The dataset for this app consists of texts from the late 2000s to the early 2010s. Hence, trending words during this period would be included in the vocabulary of the model, but not those after this period. For example, “Barack Obama” is included in the vocabulary but “Donald Trump” is not.