**Task Description**

The user will first type in or paste training text with some words prefixed with \* and !. The model will train on this text. Then, the user will enter test text without the prefixes. The program will tag same or related words and output a color coded version of the evaluation text with accuracy metrics.

UI is unimportant, the mockup below is for demonstration purposes of this document only. In the example below, the user is giving examples of cities and countries with ! prefix and pharmaceutical drugs with \* prefix. Other “taggable” examples could be animals, automobile brands, species of dinosaurs or winter sports.

**Mockup:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User input:**   |  |  | | --- | --- | | **Enter Training Text. Use \* and ! prefix for special words.** | *I lived in \*Munich last summer. \*Germany has a relaxing, slow summer lifestyle. One night, I got food poisoning and couldn't find !Tylenol to make the pain go away, they insisted I take !aspirin instead.* |   *>> model trained. Self evaluation:*   |  | | --- | | *I lived in Munich last summer. Germany has a relaxing, slow summer lifestyle. One night, I got food poisoning and couldn't find Tylenol to make the pain go away, they insisted I take aspirin instead.* |  |  |  | | --- | --- | | **Enter evaluation text.** | *When I lived in Paris last year, France was experiencing a recession. The night life was too fun, I developed an addiction to Adderall and Ritalin.* |   *>> text evaluated:*   |  | | --- | | *When I lived in Paris last year, France was experiencing a recession. The night life was too fun, I developed an addiction to Adderall and Ritalin.* | |

Robustness and Evaluation

With just one example sentence, the system has learned Cities and Pharmaceutical drugs. Think about how to evaluate the limitations of this model and improve robustness. Try to come up with some evaluation framework to help users explore the limitations of the model.

How well will your model handle longer, more difficult texts, with more diverse topics? Does it crash if the users enter some nonsensical evaluation input?