CIS468 Project Report - Auto Tagger Todo-List DB

Zekai Lin
Summer 2025
Aug.8 2025

1 Introduction

As the advance of the NLP, during the learning of the course, NLP may deeply change the way of how people interact with computers. Specifically speaking, as the nature of NLP and related workflow, instead of filling in tables, AI may be able to understand the requirement of the user, so that user can achieve their goal, by just talking.The context of this project is to try to design a potentially more convenient way of human computer interface. This project called `Auto Tagger Todo-List`, it approaches to build up an workflow of:

Accepting user's input with a paragraph description of a thing they would like to do.

Named Entity Recognition (NER) Database Builder
Identify names, places, dates, and organizations in texts or other documents, and transform the information into a database as an Todo List backend

Auto labeling in given tags (classifier)
add labels to corresponding db item while saving to the db with ranges of words(sports, cook, video game, and etc),

Save the feedback from the user and managed to self-implemented

This project should potentially act as a front end to easily fit popular scheduling apps' interface to improve the efficiency of the human computer interface.

2 Workflow Design and C

The initial tagger and extractor service is provided by the trained model.

Initial Input:

[complete] The user will give an initial input of the a thing want to remember
[complete] The initial input with the current time will be sent to the NER extractor

NER Recognition:

[ing…] for analysis of the actual time and the special Named identities.
[todo] letting user to check and provided answer, if model made mistakes

Tag Allocation:

[todo] A list of tags should be saved in the Database and managed by user, easily
[complete] Allocate the tag for given input (text, and list of several tags)
[complete] Allow user to fix the tagging if there are errors

Log as Data Set for Model Further Training

[complete] The result of each request will be saved for further implement for NLP models
[complete] Periodically train the NER Recognition & Tag Allocation [todo]

3 Sample Results

Generally Speaking, all of the codes functioning correctly, there is no fatal error occurs.

``` shell
Mounted at /content/drive/
Loaded model: en_core_web_lg

=== Auto-Tagging Todo System ===
1. Add new todo item
2. View history
3. Exit system
Please select an option: 1

=== 2025-08-08 23:06:48 EDT-0400 ===


--- Step 1: Enter Todo Description ---
Please describe what you plan to do: I plan to go play with my younger brother Bob

=== 2025-08-08 23:07:30 EDT-0400 ===


--- Step 2: Confirm Description ---
Your input: I plan to go play with my younger brother Bob
Is this description correct? (y/n): y

=== 2025-08-08 23:07:34 EDT-0400 ===


--- Step 3: Auto Analysis ---
Device set to use cpu

--- Named Entity Recognition Results ---
Event Time: Not detected
Persons: Bob

--- Tag Classification Results ---
Supported tags: sports, cooking, gaming, work, study, shopping
Predicted tags: sports, gaming

--- Step 4: Confirm Results ---
Are the results correct? (y-correct/n-revise/c-cancel): y
```

```
=== 2025-08-08 23:08:07 EDT-0400 ===


--- Processing Complete ---
Todo item saved successfully! Record ID: 1
Content: I plan to go play with my younger brother Bob
Time: 2025-08-08 23:08:07 EDT-0400
Tags: sports, gaming

=== 2025-08-08 23:08:07 EDT-0400 ===


=== Auto-Tagging Todo System ===
1. Add new todo item
2. View history
3. Exit system
Please select an option: 2

=== 2025-08-08 23:08:10 EDT-0400 ===


Found 1 historical records:

ID: 1
Creation Time: 2025-08-08 23:08:07 EDT-0400
Content: I plan to go play with my younger brother Bob
Predicted Tags: sports, gaming
Persons: Bob

=== 2025-08-08 23:08:10 EDT-0400 ===


=== Auto-Tagging Todo System ===
1. Add new todo item
2. View history
3. Exit system
Please select an option: 3

=== 2025-08-08 23:09:03 EDT-0400 ===

Thank you for using our system. Goodbye!
```

4 Review & Further Improve

Generally speaking I design a lot of modules and features, closely corresponding to each other. I keep coming up with great ideas, while designing and completing the program. For the sake of achieving reliable experiences, there are several further implementation is available:

General

On the one hand, To further improve the correctness of the module, an auto trainer function is required, for both the NERs and Tagger. On the other hand, a correctness recorder for the program is required.

NERs

Besides, there is a weakness in Time Extracting Logic. The NERs is currently extracting only two types of identities: People and Time. The time extractor can't extract all the time terms, process them wisely and give a reasonable speculation of the actual time of the event. So the further specialized logic for time identity collecting and process is required, especially when multiple time is given. In addition, the data structure of the return should also be changed, to allow saving both the start time and duration for example. I comment out recognition code for the other NERs identities,(Location, Organization, etc.), for the similar reason.

DataBase

Users should be able to do the Add, delete, modify and check to tags list, as if the tags were also saved in db instead of hard coding to the program as demo.

UI

A command line UI is just for the demo. An interactive UI will further improve the accessibilities. For example, enabling users to modify the chosen tags by just clicking the tag element in a webpage.

Furthermore, an Extra interface is required to let users manage their log and tags.

5 Conclusion

The result of the project shows the great potential of the NLP may influence the way humans interact with computers. As known, conversation based apps and platforms have achieved great technical advances and financial success. Besides, It was a great project that enabled me to have a general view.Researching and learning the AI related concepts, and completion of programming tasks, brings me a great chance to have a knowledge of fundamental techniques supporting the flourishing of the NLP industry.

References