IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

Email Notifier using Named Entity Extraction

Dr. B. Muthu Senthil, Logeshwaran M, Lokeshwaran S, Muthtamiz Selvan J
Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chennai, Tamilnadu, India

Abstract: *Email plays an important role in communication due to its flexibility, simplicity, diversified types and low cost of information. Therefore processing a large amount of emails takes a tremendous amount of human power and time. In order to quickly extract the named entities from an email requires a computerized solution. So we had proposed a mechanism to extract named entities from emails. The proposed solution integrates technology like natural language processing and information retrieval. It performs the automatic extraction of named entities from an email, organises the named entities and notifies the user. Named Entity Recognition (NER) is a vital language processing tool for information retrieval from texts like newspapers, blogs and emails. NER performs classification of words and sensing the expression of text from unstructured data. spaCy may be a free, open-source python library for advanced language Processing. It is well known for production use and helps in building the system that performs understanding of text. It helps in information extraction, understanding the systems, and preprocess the text for deep learning..*

Keywords: E-Mail, NLP, Notification, NER, spaCy, iMap, Extraction

I. INTRODUCTION

The goal of the system is to Extract the Named Entities from Business email. These entities are extracted by using spaCy. spaCy is a powerful Stringprocessing library that performs word tokenization and POS tagging. Our system aims to extract the named entities which can be processed further. This system groups the named entities and displays pop-up notification to the user.

Named Entity Recognition (NER) could be a category of knowledge extraction that extracts and classifies the named entities[7] into predefined categories like percentages, monetary values, quantities, time expressions, medical codes, locations, person names, organizations, etc .

Named Entity Recognition is not feasible sometimes, conceptually in implementations. The two problems of Named Entity Recognition are: detection of named entities [1][7] and classification of those named entities and organising the named entities[6]. The first phase is with the process called segmentation problem: named entities that are defined as contiguous spans of tokens, with no nesting, For example "Bank of Australia" is a name, in this name we have a problem regarding the reputation like the substring "Australia" is a reputation. This segmentation problem is also called chunking. Temporal expressions like time of a day, dates and numerical expressions like money value, percentage and other mathematical expressions may be considered as named entities in the Named Entity Recognition process. Sometimes the NER is not strict for some reasons. So it is clear that we cannot achieve or satisfy full accuracy. Instead we can try to increase the efficiency of the system by deploying some algorithms to achieve maximum accurate results.

II. LITERATURE REVIEW

Aaron Li-Feng Han, Xiaodong Zeng, Derek Fai Wong, Lidia Sam Chao et al [1] has proposed a Chinese Named Entity Recognition with Graph-based Semi-supervised Learning Model. Abdallah Z.S., Carman M., Haffari G et al [2] gave the Multi-domain evaluation framework for named entity recognition tools. Einat Minkov, Richard C. Wang, William W. Cohen et al [3] gave procedure for Extracting Personal Names from Email by Applying Named Entity Recognition to Informal Text.

Zhou Guo Dong, and Jian Su [5] proposed a Named entity recognition using an HMM-based chunk tagger. Juan Li , Souvik Sen , Nazia Zaman et al [6] provided procedure for Entity Extraction From Business Emails. Stolfo, Salvatore J, Shlomo Hershkop, Chia-Wei Hu, Wei-Jen Li, Olivier Nimeskern and Ke Wang et al [7] introduced a Behavior-based model and its application to email analysis.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

Saleem Ozair, Latif Seemab et al [8] has proposed Information Extraction from Research Papers by Data Integration and Data Validation from Multiple Header Extraction Sources.

III. EXISTING SYSTEM

In order to quickly extract the named entities from an email requires acomputerized solution. So in the above proposed solutions, it uses a mechanism to extract named entities from emails. The above existing solution uses technology like information retrieval and natural language processing to extract the required entities. It enables the automatic extraction of important entities from an email and makes batch processing to organise and group such email.

3.1 Limitations of Existing System

- This system fails to classify the named entities into predefined categories.
- Existing systems are not dynamic and work only for the use cases theyare designed for.
- This system fails to identify the meaning of words according to thecontext.
- No feedback loop to improve the system's efficiency and accuracy

IV. PROPOSED SYSTEM

Our proposed system uses an "imaplib" library in python, it is going to establish the connection to respective mail servers and it allows us to extract the mail body content. After extracting the mail body contents, these mail contents are imported as text from the ner module. This ner module is going to extract the named entities from mail and it categorises these named entities into groups. Finally, the named entities are displayed as a pop-up notification.

4.1 System Flow

In order to design a service for users, the design of the components has to be done first. Communication between different components belonging to the service also plays a major role in the whole design. Also, the flow of text data from the user to the system and the reverse is also important to be designed before development.

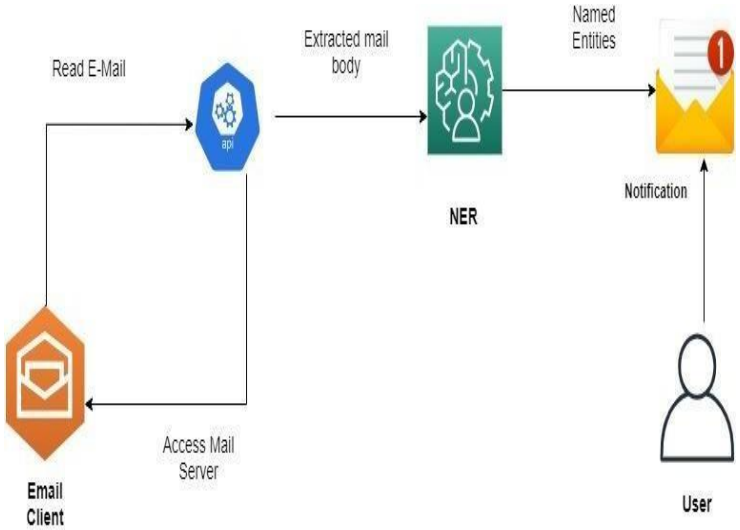

Figure 1

MODULES

The proposed solution can be divided into three main modules:

- Mail Access Protocol
- Named Entity Recognition
- Desktop Notifier

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

MAIL ACCESS PROTOCOL

The Mail Access Protocol consists of imaplib which may be a module in python. Internet Message Access Protocol provides the client program to access and read content of the mail. The email is stored and maintained by the remote server.It provides some features such as organise, delete and manipulate remote message folders from an email which is called as mailboxes. Here the users can access and read the emails without the need of downloading it. It can be accessed to multiple mailboxes on multiple mail servers. After establishing a connection to the mail server, the inbox of a mail will be accessed. Particular organisation mails were selected and also the body content of those emails are extracted. From these body contents named entities are extracted and organised.

NAMED ENTITY RECOGNITION

In NER where the named entities are extracted from the email body. The text is cleaned from unnecessary information that is present in the content example like HTML tags and the named entities are segmented into sentences. Sequences of tokens can be converted by tokenizer. After tokenization, all the words that present in the sentence are converted into lowercase letters to easily extract the named entities.To improve efficiency and effectiveness of study the words are removed from the text. Reducing a word to its root or simpler form 34 performed within the text is called Stemming. NLTK is employed during this system to tag the parts of speech of every word. After parts of speech tagging is finished, spaCy library is employed to spot the phrase and categorise the phrase into their respective category.

DESKTOP NOTIFIER

This Desktop Notifier is the module where the named entities are received, organise these named entities and make crop up notification to the user. By employing a package available in python named win10toast, we will create desktop notifications. It's simple thanks to get notified when an incident occurs. So when the mail is received from any respective organisations, using the NER module named entities are extracted and it gets notified as a desktop notification to the user. The notifications are shown for the time that we specify. When the required time is exceeded, then the notification will disappear from the notification area. The contents that displayed within the notification area are maid id as notification title and named entities as notification body.

V. RESULT AND CONCLUSION

The Named Entity Recognition using email was completed. They were integrated and the content extraction from email and notifying the user was completed and tested with several test cases.

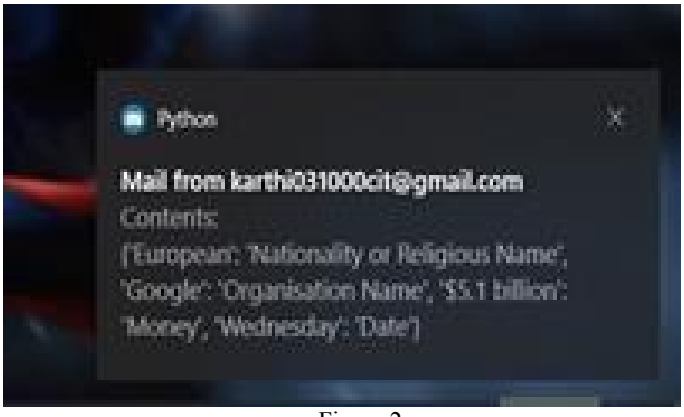

Figure 2

In our lifestyle, Email plays an important role in the industry world. Thereare plenty of business organisations in which they use email as a communication standard to communicate with their clients. So that we had proposed a mechanism to extract named entities from emails. Our system uses efficient algorithms to extract the named entities from the user and classifying the namedentities. Our solution is the efficient, effective and reliable one. Future works include training the system to acknowledge the content more accurately and efficiently. These named entities are further processed to hold out more different operations.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

REFERENCES

[1]. Han, Li-Feng Aaron, Wong, Zeng, Xiaodong, Derek Fai, Chao, Lidia Sam. (2015). Chinese Named Entity Recognition with Graph-based Semi-supervised Learning Model. In Proceedings of SIGHAN workshop in ACL-IJCNLP. 2015.

[2]. Abdallah, Z.S., Carman, M., Haffari, G.: Multi-domain evaluation framework for named entity recognition tools. Comput. Speech Lang. 43, 34–55 (2017) .

[3]. Einat Minkov, Richard C. Wang, William W. Cohen " Extracting Personal Names from Email: Applying Named Entity Recognition to Informal Text "Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing October 2005

[4]. Swapnil, V., Jayshree, A.: Natural language processing preprocessing techniques. Int. J. Comput. Eng. Appl. XI(Special Issue) (2017). http://www.ijcea.com/. ISSN 2321-3469 .

[5]. Zhou, GuoDong, and Jian Su. "Named entity recognition using an HMM-based chunk tagger." proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics,2002.

[6]. Juan Li , Souvik Sen , Nazia Zaman . "Entity Extraction From Business Emails" August 2015 International Journal of Information Technology and Computer Science 7(9):15-22

[7]. Stolfo, Salvatore J., Shlomo Hershkop, Chia-Wei Hu,Wei-Jen Li, Olivier Nimeskern, and Ke Wang. "Behavior-based modeling and its application to email analysis." ACM Transactions on Internet Technology (TOIT) 6, no.2 (2006): 187-221.

[8]. Saleem, Ozair, Latif, Seemab. "Information Extraction from Research Papers by Data Integration and Data Validation from Multiple Header Extraction Sources." WCECS 2012,October 24-26, 2012, San Francisco, USA.