Indian Institute of Information Technology, Allahabad
Prayagraj

Event Extraction From Emails Using Natural Language Processing

Natural Language Processing - Project

Presented By:

Anand Geed
IDS2022005
M.Tech (DSA)

Submitted to:
Dr. Muneendra Ojha

Contents

“Abstract

Abstract

Event extraction from emails involves identifying and extracting relevant information about events mentioned in the email. This can include the event's name, location, date, time, duration, and any other relevant details. The process often involves using natural language processing (NLP) techniques, such as named entity recognition (NER) and information extraction (IE), to identify and extract the relevant information from the email text. The extracted information can then be used to populate event calendars or other scheduling tools, enabling users to keep track of their upcoming events more easily.

“Introduction

Introduction

In today's digital age, email is one of the primary modes of communication, and people often receive a large number of emails every day. These emails can contain a wide range of information, including invitations to events such as meetings, conferences, and social gatherings. Keeping track of these events can be challenging, especially when they are spread across multiple emails or when the event details are not clearly stated in the email text.

Introduction

Event extraction from email aims to automate the process of identifying and extracting relevant information about events mentioned in emails. By automatically extracting the key details about an event, such as the event name, location, date, and time, users can easily add the event to their calendars or other scheduling tools, saving them time and effort. This can be particularly helpful for busy professionals who need to manage multiple events and commitments.

Introduction

The process of event extraction from email typically involves using natural language processing (NLP) techniques, such as named entity recognition (NER) and information extraction (IE), to identify and extract the relevant information from the email text. This information can then be used to populate event calendars or other scheduling tools, enabling users to keep track of their upcoming events more easily

Proposed methodology in detail

Proposed methodology in detail

Event extraction from emails involves automatically identifying and extracting information about events mentioned in email messages. Here is a proposed methodology for event extraction from email:

- Data Collection: The first step in the methodology is to collect a dataset of Email. This dataset will be used to train the machine learning models used in subsequent steps. The dataset should be diverse.
- Data Preprocessing: The first step is to preprocess the email data to extract the relevant content, such as subject line, body text, and any attachments. This can be done using libraries such as NLTK or spaCy, which can help tokenize and tag the email content.
- Named Entity Recognition: The next step is to use named entity recognition (NER) to identify and tag named entities in the email content. This includes identifying and tagging entities such as people, organizations, locations, and dates/times.

Proposed methodology in detail

- Rule-Based Approach: A rule-based approach can be used to identify specific keywords or patterns in the email content that indicate the mention of an event. For example, keywords such as "meeting," "appointment," "conference," "party," etc. can be used as indicators of an event.
- Machine Learning Approach: A machine learning approach can be used to train a classifier on a labeled dataset of emails to identify events. The classifier can use features such as the presence of specific words, named entities, and patterns in the email content to predict whether an email is about an event.
- Event Extraction: Once an email has been identified as containing information about an event, the next step is to extract the relevant information about the event, such as the event name, date/time, location, participants, and description. This can be done using techniques such as information extraction, relation extraction, and sentiment analysis.

“Implementation

Implementation

```
data = pd.read_csv("/content/drive/MyDrive/M.Tech(DSA)/emails.csv")
data.head()
```

	file	message
0	allen-p/_sent_mail/1.	Message-ID: <18782981.1075855378110.JavaMail.e...
1	allen-p/_sent_mail/10.	Message-ID: <15464986.1075855378456.JavaMail.e...
2	allen-p/_sent_mail/100.	Message-ID: <24216240.1075855687451.JavaMail.e...
3	allen-p/_sent_mail/1000.	Message-ID: <13505866.1075863688222.JavaMail.e...
4	allen-p/_sent_mail/1001.	Message-ID: <30922949.1075863688243.JavaMail.e...

```
[5] data.shape
```

```
(517401, 2)
```

Implementation

```
[32] data.head()
```

	subject	X-Folder	body
1	Re:	'sent mail	Traveling to have a business meeting takes the...
2	Re: test	'sent mail	test successful. way to go!!!
4	Re: Hello	'sent mail	Let's shoot for Tuesday at 11:45.
5	Re: Hello	'sent mail	Greg,\n\n How about either next Tuesday or Thu...
7	Re: PRC review - phone calls	'sent mail	any morning between 10 and 11:30

```
data.body.loc[7]
```

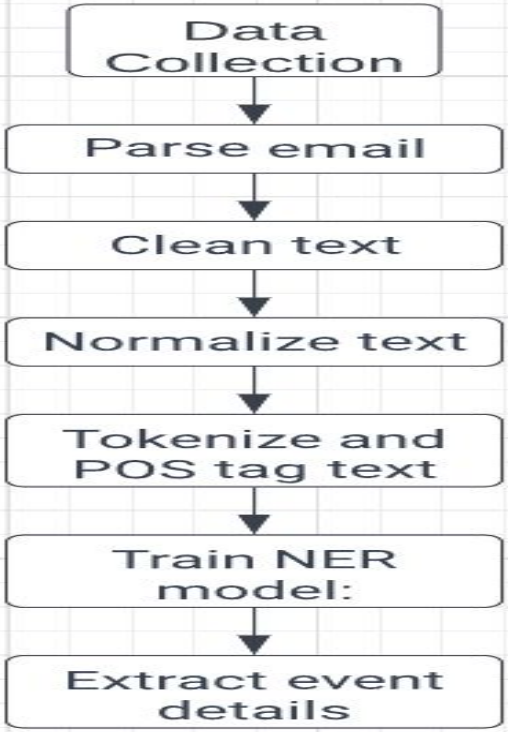
```
'any morning between 10 and 11:30'
```

```
[34] data.body.iloc[0]
```

```
'Traveling to have a business meeting takes the fun out of the trip.  Especially if you have to prepare a presentation.  I would suggest holding the business plan meetin
gs here then take a trip without any formal business meetings.  I would even try and get some honest opinions on whether a trip is even desired or necessary.\n\nAs far a
s the business meetings, I think it would be more productive to try and stimulate discussions across the different groups about what is working and what is not.  Too oft
en the presenter speaks and the others are quiet just waiting for their turn.   The meetings might be better if held in a round table discussion format.  \n\nMy suggesti
on for where to go is Austin.  Play golf and rent a ski boat and jet ski's.  Flying somewhere takes too much time.\n'
```

1m 58s completed at 11:22 PM

Implementation



Results

```
Time: 03:03:11
Date: Nov 2000
Location: http://profiles.msn.com
---
Time: 4:45 PM
Date: Wednesday, November 15, 2000
Location: Yipppeee
---
Time: 4:45 PM
Date: Wednesday, November 15, 2000
Location: None
---
Time: 12:01 PM
Date: Thursday, November 16, 2000
Location: Sanchez
---
Time: earlier today
Date: earlier in the year
Location: NEW YORK
---
Time: 2:00-2:30
Date: Friday, December 15th
Location: Desktop
```

CONCLUSION

Extracting events from emails can be a challenging task, as emails often contain unstructured and informal language. However, by using techniques such as natural language processing (NLP) and machine learning (ML), it is possible to automatically extract events from emails with a high degree of accuracy.

The extracted events can be used for a variety of purposes, such as scheduling meetings, tracking project progress, and analyzing customer interactions. Additionally, the extracted events can be integrated with other systems, such as calendars and project management tools, to facilitate efficient workflows.

Overall, event extraction from emails has the potential to streamline communication and improve productivity in various domains, making it a valuable application of NLP and ML technology.

Thank You !!