Mining Workflow Models

from

Web Application

IT303 SOFTWARE ENGINEERING

Submitted By:
Ayush Bhandari (181IT209)
Abhishek Kaswan (181IT201)
Siddharth Pokharna (181IT146)

INTRODUCTION

The ubiquity of web browsers and advancements in web technologies has resulted in web applications becoming a dominant client for enterprise software.
The availability of network bandwidth enables applications to be operated by the vendor and provided as services to customers.
Operating applications on the vendor side removes some of the complexity and costs of the traditional software release and update process; while this enables shorter, more efficient and frequent release cycles with a smaller number of features, it puts more pressure on software development and requires paying close attention to operational aspects, continuous quality assurance (QA) and testing.

PROBLEM STATEMENT

Build a tool to mine behavior models from web applications that support multi-user workflows by extracting the relevant data from webpage with the crawling of particular webpage and arranging the data in structured manner and visualise it and analysis of data.

LITERATURE SURVEY

S.No	Research Paper	Author	Details
1	Mining workflow model from web applications (IEEE)	Matthias Schur, Andreas Roth, Andreas Zeller	i)Main focus was automatic exploration of webpage application ii)Testing done on real life applications and tool used was ProCrawl.
2	Mining Most Specific Workflow Models from Event-Based Data	Guido Schimm	i)Mining of data from event based data . ii)Block structured workflow model and extraction procedure were described
3	Mining workflow model from web applications (IJSPR)	Miss. Puja Shankar Salunkhe, Prof. Dr. D.R. Ingle	Increased model accuracy using active learning i.e it observes each iteration of execution actions and observe UI changes

TOOLS AND TECHNOLOGIES

Important modules used:

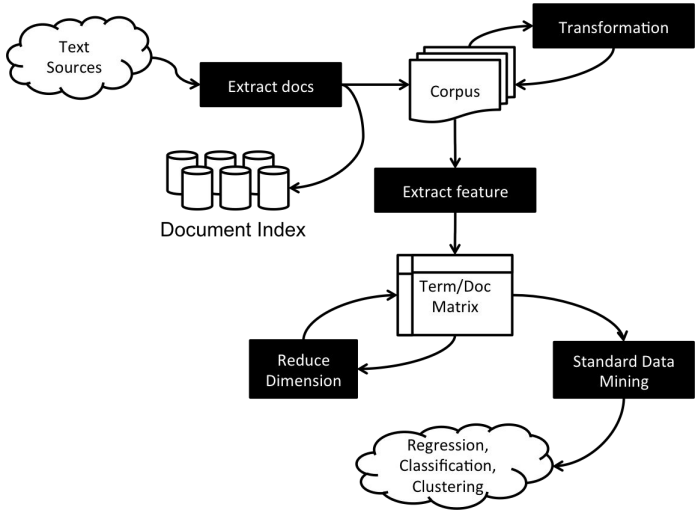
- Tweepy: An easy-to-use Python library for accessing the Twitter API.
- Nltk: NLTK is a leading platform for building Python programs to work with human language data.
- YAML: it`s a data-serialization language commonly used for configuration files.
- NLP libraries such as:
    - Textblob for sentiment analysis which is a python library for processing textual data
    - word cloud which is used for data visualization

Other:

- Dataframe and other packages from Pandas, numpy , matplotlib as plotting library

METHODOLOGY

Workflow Model



Process Model

- Data understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

Implementation

1)Extract the data from the file or scrape the website and gather the data.

2)Differentiate between structured and unstructured data

3)Data

- i)convert files into matrix format to better visualize the data.
- ii)import pandas, numpy, sklearn, seaborn , matplotlib functions

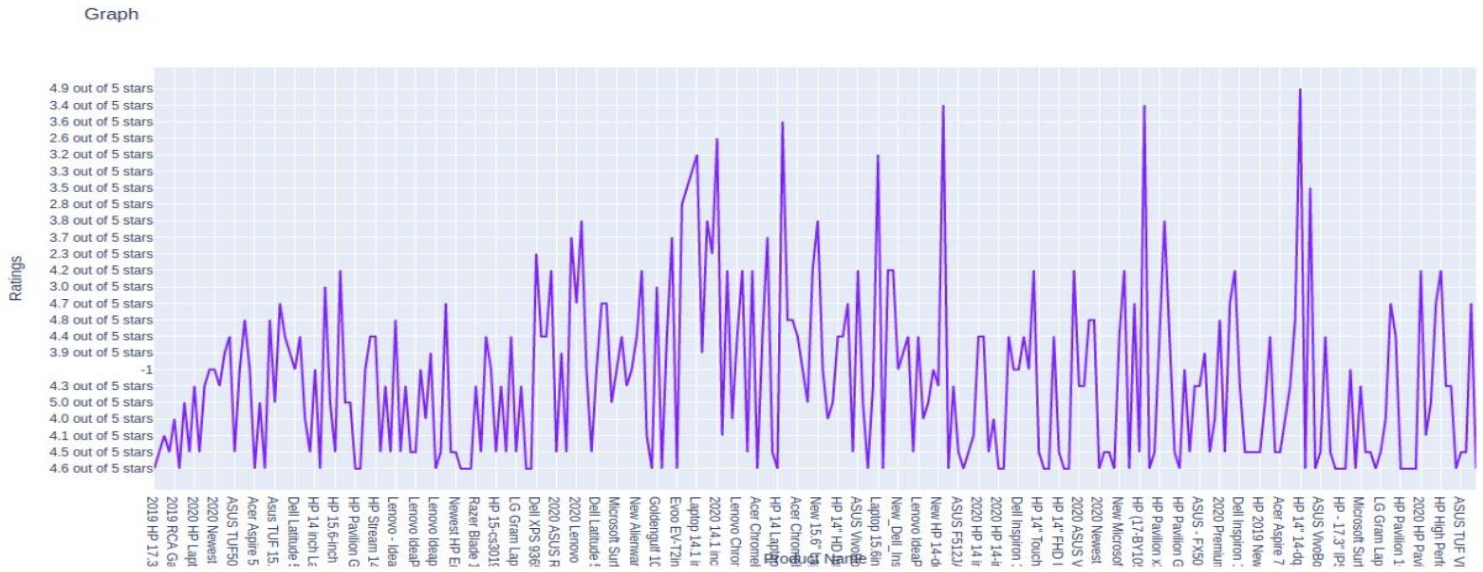
5) Visualistation

- i)Plotting of the data
- ii) Analysis of the result

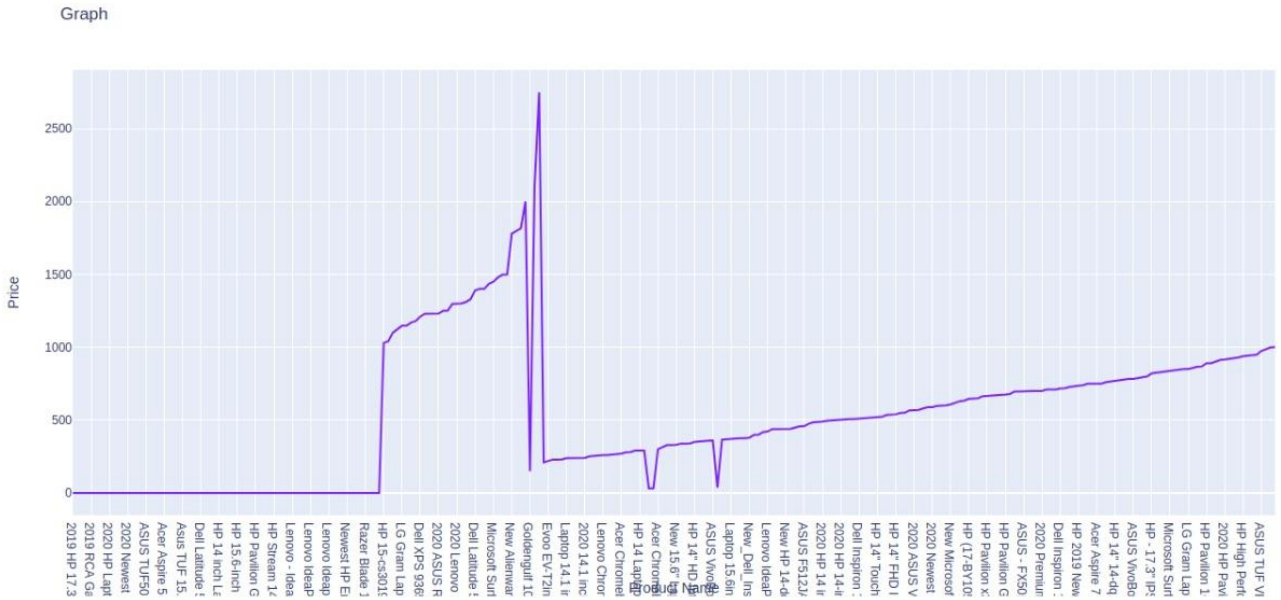
RESULT AND ANALYSIS

Testing was done on two websites :Amazon ,Twitter

A)Amazon webpage of laptop purchase using crawling

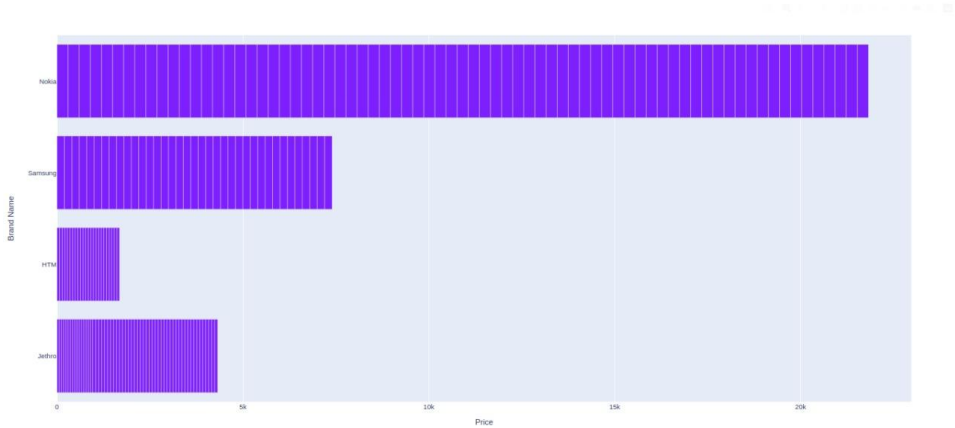


Rating review given by the users to on different laptop till that day



Price of each laptop on a particular day

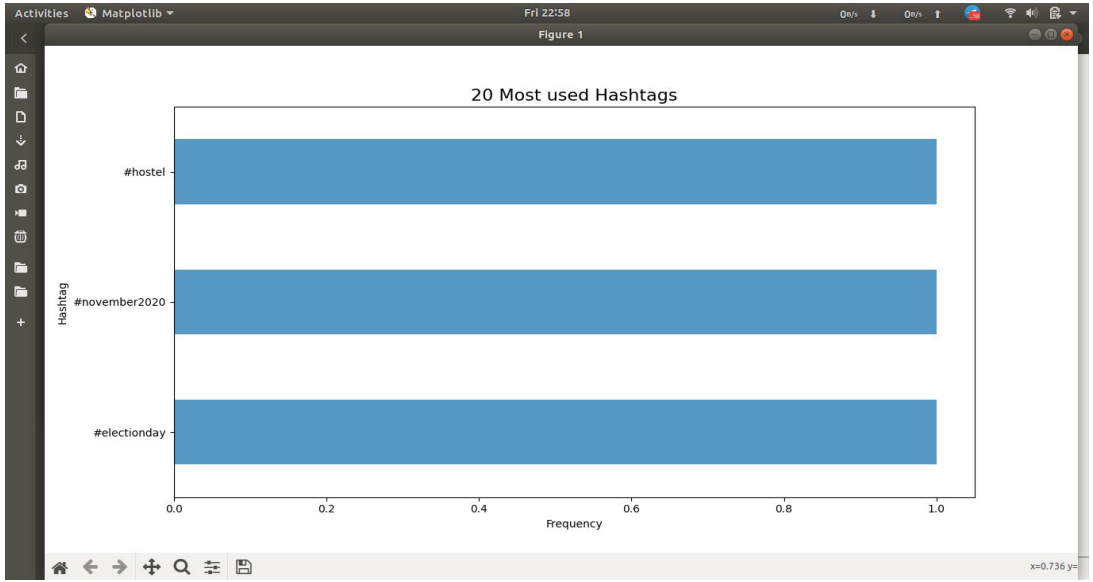
Amazon dataset on smartphone



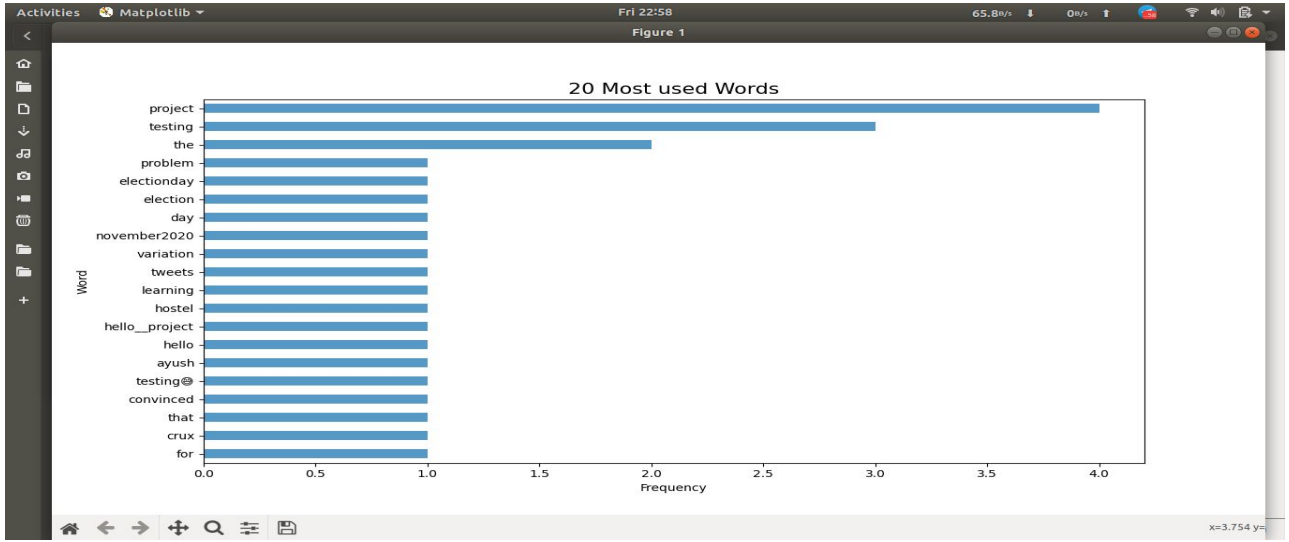
Price of some top selling smartphone brands

B) Twitter account testing

User:



Top mentioned accounts on twitter by user



Top 20 most frequent words used by the user

Most frequent hashtags used by user

FUTURE WORK

The project in future work can be focused on the following topics:

1. **Integrating existing tests:** many Web applications come with existing unit and system tests,so exploring means to integrate and adapt these tests into automated crawling, using their data for input provisioning, and their interaction flows for even better coverage of data size.
2. **Richer models.**: At this point, the base of the model is purely state-based. On adding leveraging context-free and context-sensitive grammars would lead to expressing much more complex interactions and dependencies.
3. **Alternative platforms:** Besides Web applications, one could also perform mining on generic GUI-driven applications, providing model extraction and subsequent model-based testing on a wide range of platforms and programs.

CONCLUSION

In this project we tried to mine behavior models from web applications that support multi-user workflows by extracting the relevant data from webpage with the crawling of particular webpage and arranging the data in structured manner and visualise it and analysis of data.

As our evaluation on several real-world web applications shows, the models mined are adequate in size, accurate, cover most of the workflow-relevant actions. The models mined provide an excellent starting point for manual refinement. With these different mining techniques , the project lays the path for future model mining tools.

REFERENCES

[1]Mining workflow from web applications ieee report by Matthias Schur,Andreas Roth,Andreas Zeller

[2] Mining Most Specific Workflow Models from Event-Based Data by Guido Schimm

[3] Mining Workflow Models from Web Application(IJSPR) by Miss. Puja Shankar Salunkhe, Prof. Dr. D. R. Ingle

[4] V. Dallmeier, N. Knopp, C. Mallon, S. Hack, and A. Zeller. Generating test cases for specification mining. In *ISSTA*, pages 85–96, New York, USA, 2010. ACM.

[5] C. Di Francescomarino, A. Marchetto, and P. Tonella. Reverse engineering of business processes exposed as web applications. *Proceedings of the European Conference on Software Maintenance and Reengineering, CSMR*, pages 139–148, 2009.

THANK YOU