Healthy Search With Ai

Mozilla Hackathon 2019

Team: Winter Coders

Members:

Rajashekhar.G

BhanuPrasad.T

Pranavi.V

Contact Email:

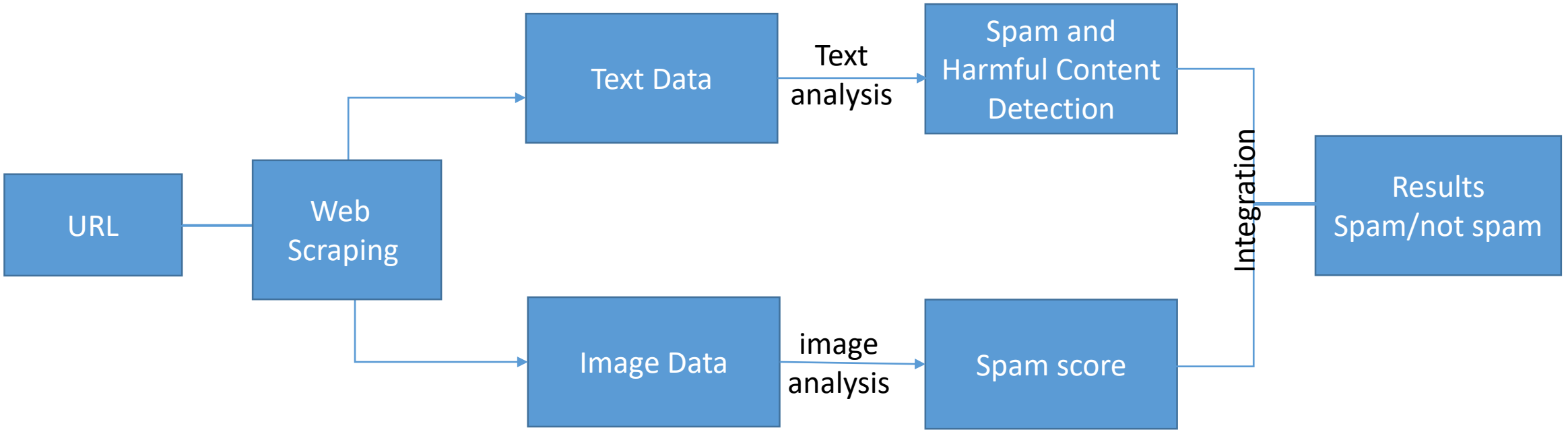
rajharry418@gmail.com

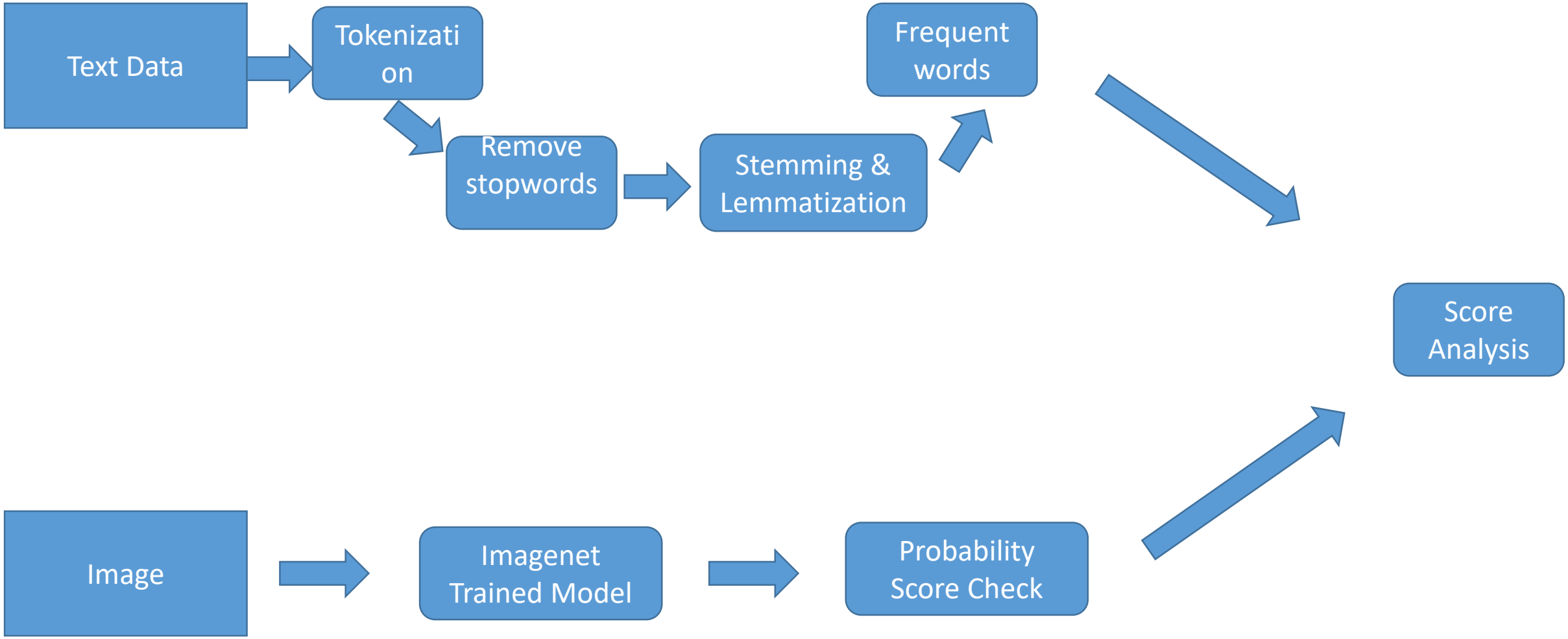
Our Theme [Healthy Search With Ai]

- Integrate Artificial Intelligence to the web browsers

- Develop a spam website detector using text summarization and image analysis.

Our Approach




Text Data
Tokenizati on
Remove stopwords
Stemming & Lemmatization
Frequent words
Image
Imagenet Trained Model
Probability Score Check
Score Analysis

Text Analysis

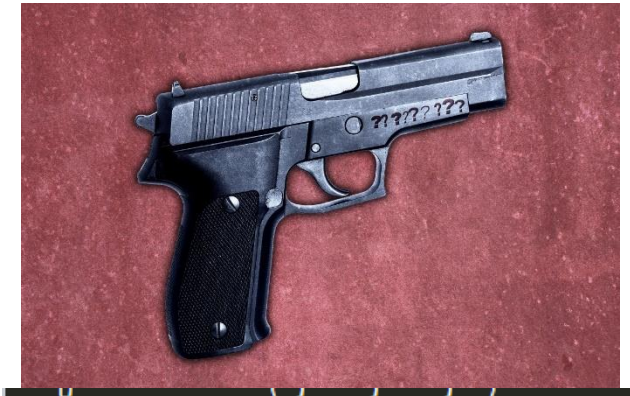
1) Fetching text from web page
2) Preprocessing
    - Text
    - Tokenization
    - Remove stop words
    - Stemming
    - Lemmatization
    - Most Frequent words
3) Spam Analysis

Image Analysis

```
[[('n02123597', 'Siamese_cat', 0.30934158), ('n01877812', 'wallaby', 0.080341235), ('n02326432', 'hare', 0.07509846), (
0.05053066), ('n03223299', 'doormat', 0.048173606)]]
```

- Fetching images from web page
- Neural network model
- Probability scores
- Spam Analysis

```
[[('n03937543', 'pill_bottle', 0.794242), ('n04116512', 'rubber_eraser', 0.18120602), ('n03729826', 'matchstick', 0.004279587), ('n04423845',
'thimble', 0.0027179786), ('n02783161', 'ballpoint', 0.0020048076)]]
```



```
[[('n03527444', 'holster', 0.70728076), ('n04086273', 'revolver', 0.2470884), ('n02749479', 'assault_rifle', 0.035497267), ('n04090263', 'rifle',
0.010082758), ('n02879718', 'bow', 1.4817832e-05)]]
```

Integration

- The spam scores of text and images are integrated
- The web page will be categorized as spam or not spam based on threshold spam score
- The spam message will prompted to user

Outputs

```
Enter Url:https://www.facebook.com/pages/category/Scientist/Dangerous-Khiladi-VK-936997396426201/
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.4332268238067627
** No spam in this website.. :)**

>> Image Analysis
/usr/local/lib/python3.6/dist-packages/PIL/Image.py:914: UserWarning: Palette images with Transparency   expressed in bytes should be converted to RGBA images
  'to RGBA images')
Illegal weapons count:  0
** You are good to go!!! **
time elapsed for *Image Analysis*:  1.5957505702972412
```

```
url = input("Enter Url:")
img_text_model(url)
```

```
Enter Url:https://www.thedelite.com/worlds-most-dangerous-dog-breeds/
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.19484519958496094
** No spam in this website.. :)**

>> Image Analysis
Illegal weapons count 0 out of 0
** You are good to go!!! **
time elapsed for *Image Analysis*:  0.21002531051635742
```

```
[53] url = input("Enter Url:")
     img_text_model(url)
```

```
Enter Url:https://www.expressvpn.com/unrestricted-4
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.3072023391723633
**Webpage Warning (Spam Content Detected)**
```

```
url = input("Enter Url:")
img_text_model(url)
```

```
Enter Url:https://listotop.com/top-10-most-dangerous-guns-in-the-world/
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  2.8095827102661133
** No spam in this website.. :)**

>> Image Analysis
Illegal weapons count 11 out of 36
** WebPage Blocked (Illegal Content) **
time elapsed for *Image Analysis*:  9.403199672698975
```

Outputs

```
url = input("Enter Url:")
img_text_model(url)
```

```
Enter Url:https://www.facebook.com/pages/category/Scientist/Dangerous-Khiladi-VK-936997396426201/
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.4485604763031006
** No spam in this website.. :)**

>> Image Analysis
/usr/local/lib/python3.6/dist-packages/PIL/Image.py:914: UserWarning: Palette images with Transparency   expressed in bytes should be converted to RGBA images
  'to RGBA images')
Illegal weapons count 0 out of 9
** You are good to go!!! **
time elapsed for *Image Analysis*:  1.5955278873443604
```

```
url = input("Enter Url:")
img_text_model(url)
```

```
Enter Url:https://www.expressvpn.com/unrestricted-4
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.3072023391723633
**Webpage Warning (Spam Content Detected)**
```

```
[54] url = input("Enter Url:")
     img_text_model(url)
```

```
Enter Url:https://www.surveycompare.co.in/?cid=525bed4a6b55f&source=google&medium=cpc&campaign=920833155&adgroup=44453224245&targetid=kwd-302595485974&keyword=e
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.4898972511291504
WebPage Blocked (Spam Content Detected)
```


```
url = input("Enter Url:")
img_text_model(url)
```

```
Enter Url:https://www.expressvpn.com/unrestricted-4
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.3072023391723633
**Webpage Warning (Spam Content Detected)**
```

```
[54] url = input("Enter Url:")
     img_text_model(url)
```

```
Enter Url:https://www.surveycompare.co.in/?cid=525bed4a6b55f&source=google&medium=cpc&campaign=920833155&adgroup=44453224245&targetid=kwd-302595485974&keyword=e
>> analysis started !!!
>> Text Analysis
time elapsed for *Text Analysis*:  0.4898972511291504
WebPage Blocked (Spam Content Detected)
```

Conclusion and Future work

- User can browse in safe internet
- Develop a add-on for a browser.
- Ai model for Text summarization
- Ai model For Illegal object detection.