Election Campaign Dynamics

Updates, Website

Anahita Poulad
Emre Canbazer
Amir Jamali
Melina Chioutakou

February 3, 2022

Outline

Updates

UPDATES

- Refining the Twitter Corpus
  - Requests from clients
    - New candidate: Koenig
    - Retreated/Eliminated candidates: Montebourg, Larrouturou, Bertrand
    - New features: Separate date and hours

emojicons	time	time_Y_M_D	time_H_M_S
None	2022-02-02 16:35:18	2022-02-02	16:35:18
:green_square:, :scroll:, :green_square:, :gre...	2022-02-02 11:04:25	2022-02-02	11:04:25
:green_square:, :notebook:, :green_square:, :s...	2022-02-02 10:53:58	2022-02-02	10:53:58
:France:, :megaphone:, :green_square:, :scroll...	2022-02-02 10:49:05	2022-02-02	10:49:05
	2022-		

Candidate Topic Relation over time

```
similarity = {'dates':[],'sim':[]}
for ind in candid_1_topics_r.index:
    candid_1_t = candid_1_topics_r.topic[ind]
    candid_2_t = candid_2_topics_r.topic[ind]
    doc1 = nlp(candid_1_t)
    doc2 = nlp(candid_2_t)
    similarity['dates'].append(ind.date())
    similarity['sim'].append('{:.3}'.format(doc1.similarity(doc2)))
similarity_df = pd.DataFrame.from_dict(similarity,dtype=float)
```

Figure 1: Get the semantic similarity based on time

	dates	sim
0	2022-01-16	0.790
1	2022-01-17	0.836
2	2022-01-18	0.669
3	2022-01-19	0.923
4	2022-01-20	0.681
5	2022-01-21	0.727
6	2022-01-22	0.800

Figure 2: Resulting similarity dataframe

Candidate Topic Relation over time

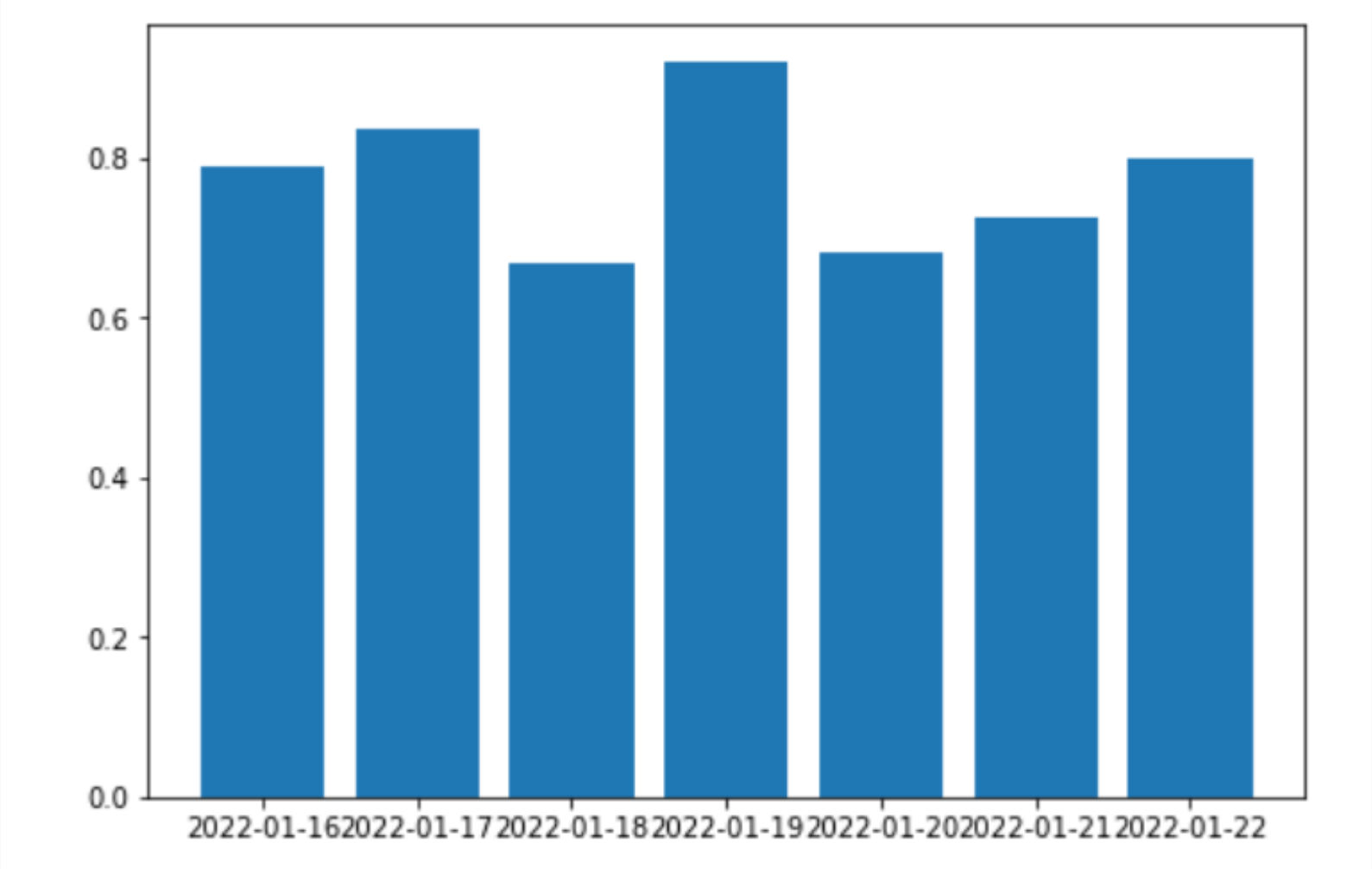
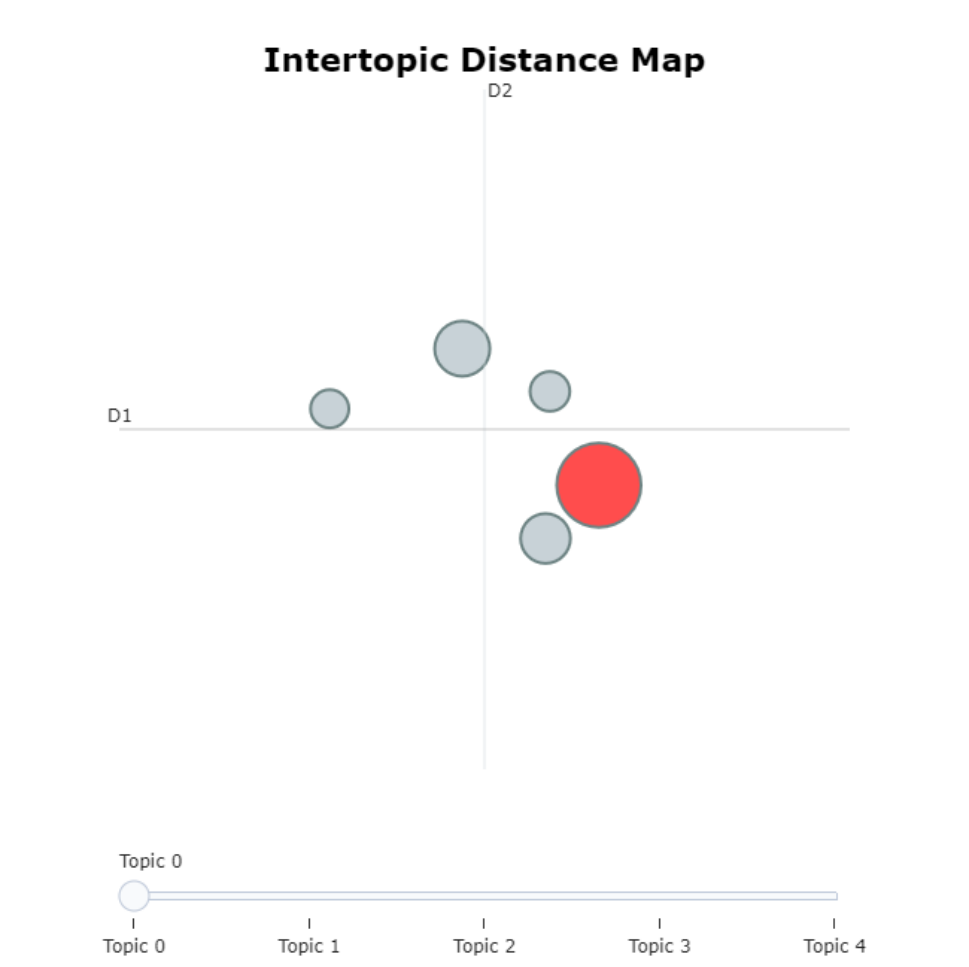


Figure 3: Zemmour and Le Pen's topics similarity over a week

Inter-Topic Distance



TOPIC SIMILARITY

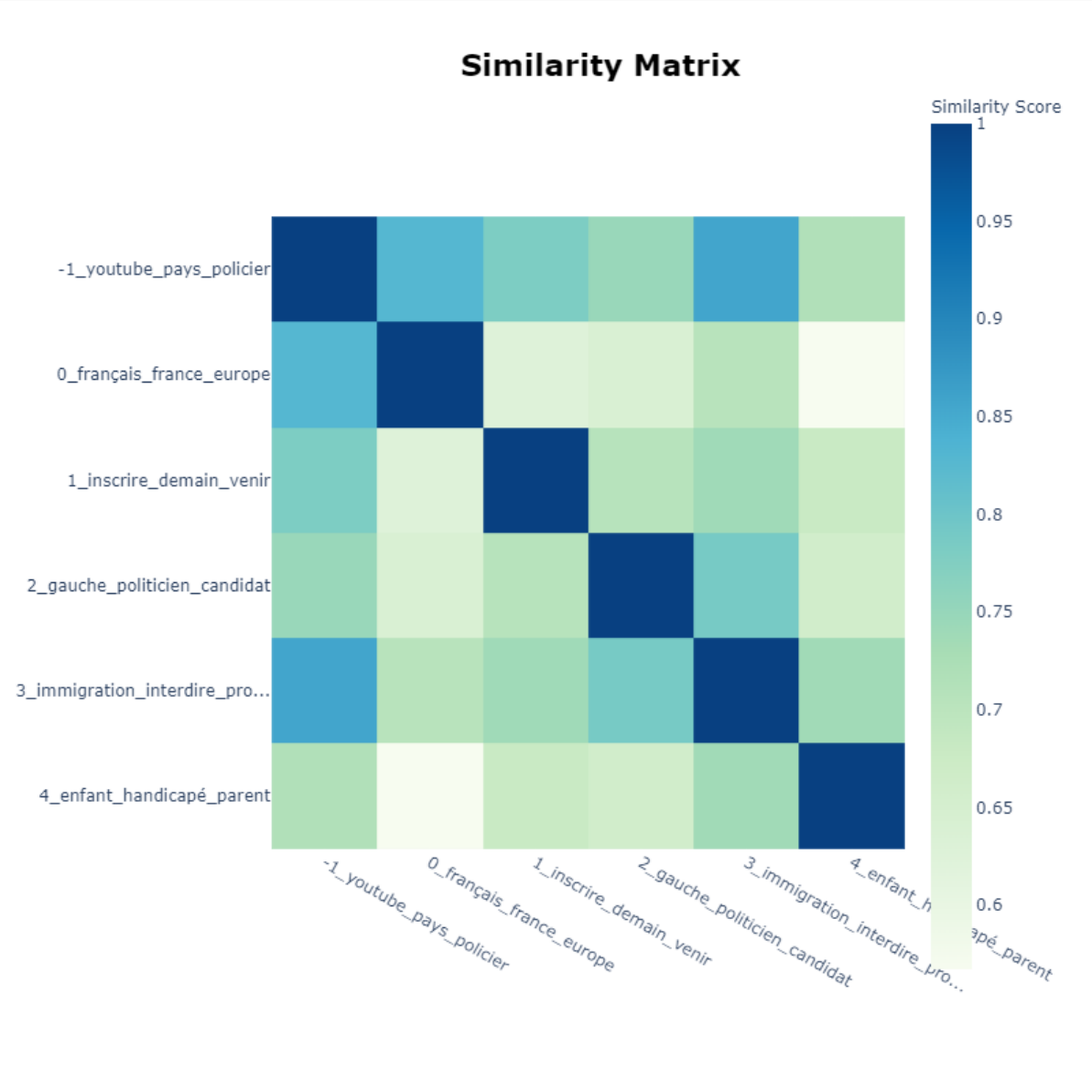
In [29]:
```
import fasttext
model = fasttext.load_model("gensim-data\\cc.fr.300.bin\\cc.fr.300.bin")
```

```
Warning : `load_model` does not return WordVectorModel or SupervisedModel any more, but a `FastTex
ar.
```

In [30]:
```
import numpy as np

df = pd.DataFrame({"questions":topic_list_1})

df["vecs"] = df["questions"].apply(model.get_sentence_vector)

from scipy.spatial.distance import pdist, squareform
out = pdist(np.stack(df['vecs']), metric="cosine")
cosine_similarity = squareform(out)
print(cosine_similarity)
```

```
[[0.         0.56311359 0.45381749 0.6564606  0.64648623 0.59208383]
 [0.56311359 0.         0.57012764 0.53140708 0.60120124 0.56137877]
 [0.45381749 0.57012764 0.         0.50060046 0.59027854 0.519126  ]
 [0.6564606  0.53140708 0.50060046 0.         0.62654354 0.68910435]
 [0.64648623 0.60120124 0.59027854 0.62654354 0.         0.58767631]
 [0.59208383 0.56137877 0.519126   0.68910435 0.58767631 0.        ]]
```

Topic Similarity



Evaluation

- Evaluating a topic model is a challenging task
  - unsupervised models
  - The absence of standard measures and well-established tools

◇ Normalized Point-wise Mutual Information (NPMI)

◇ It measures the topic coherence between high scoring words in the topic

◇ Ranges from [-1,1]

◇ The higher positive NMPI the better

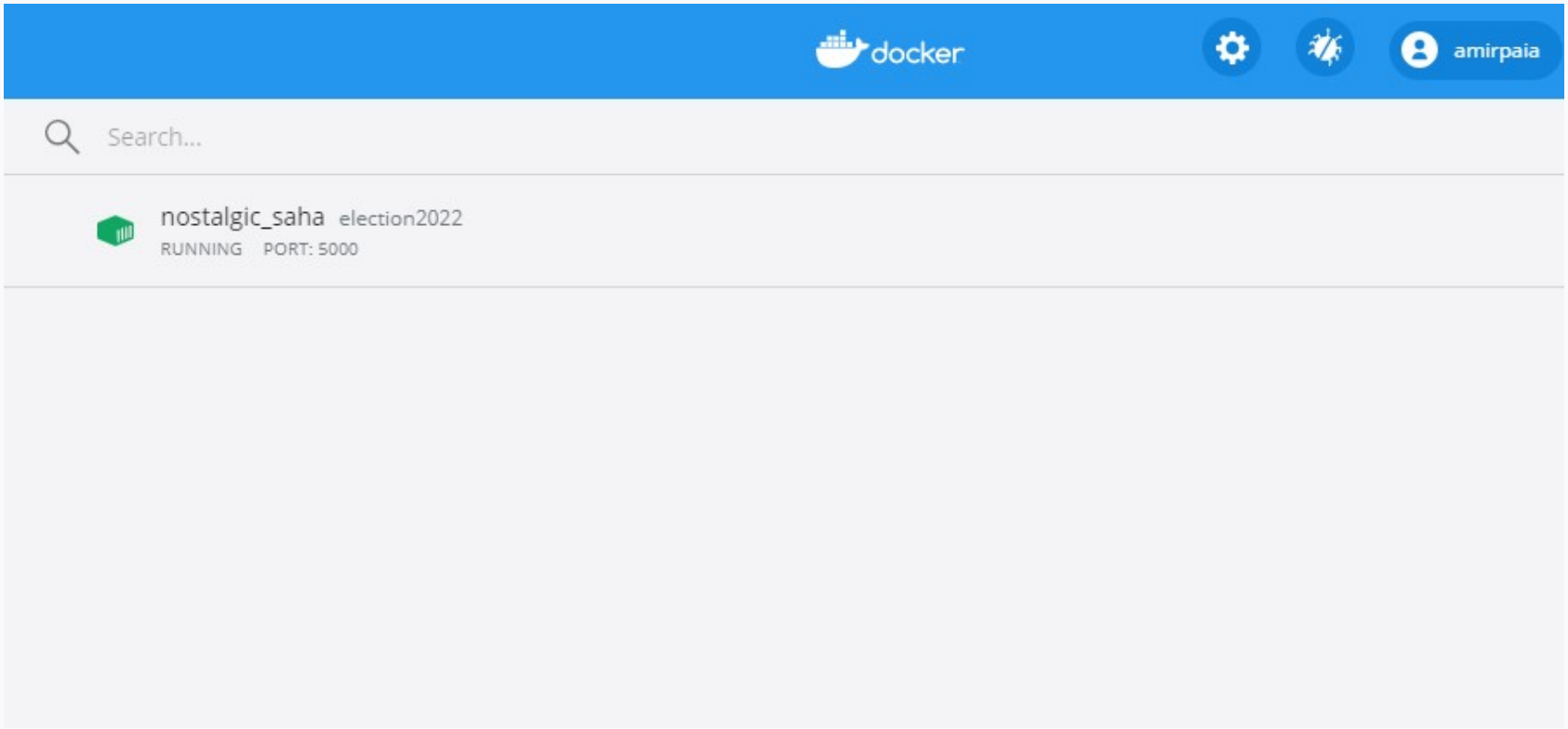
Website

Docker Image and Deploy

- Creating the dockerfile
- Building the docker image
- Tagging the docker image on ACR (Azure Container Registry)
- Pushing the tagged docker image to be deployed

Docker Image has already been built

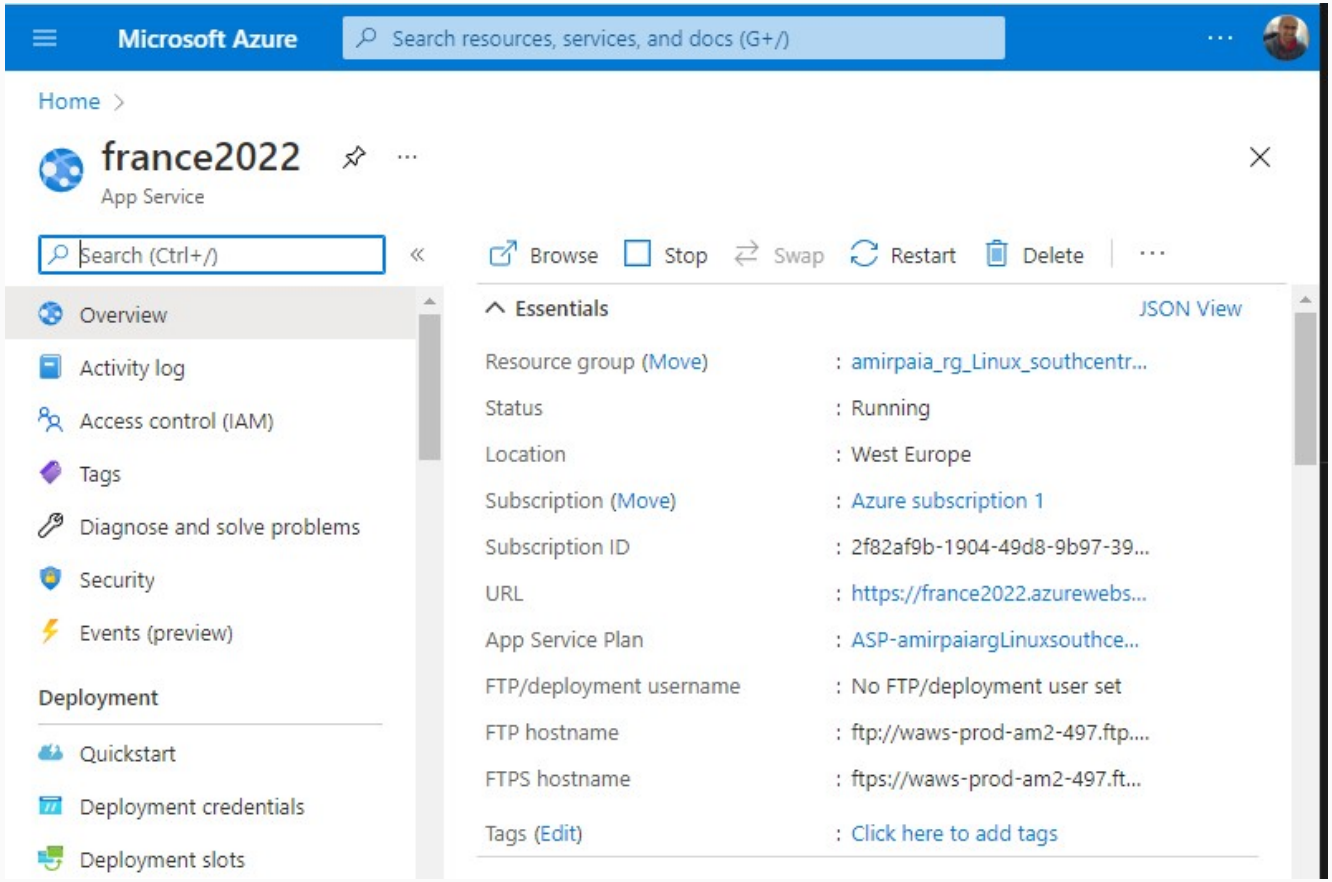
A docker image has been successfully built



Election2022

Deployment on Azure

Docker image can be easily pushed and deployed into Azure



VISUALIZASION

Bringing the abilities of Matplotlib and Plotly into the browser by creating an HTML file (Click Here)

In the next presentation, we will connect this possibility to the candidate names

Thank you!