

Projet informatique

ÉTAT D'AVANCEMENT



Twelec



MAXIME DA SILVA

ENSC- Quorum

Table des matières

Le projet	3
État d'avancement	4
Phase 1 : Récupération des données	4
Phase 2 : Sauvegarde des tweets	4
Phase 3 : Agrégation des informations	4
Les problèmes rencontrés	6
Perspective et suite du projet	6
Annexes	6
Annexe1	6

Le projet

Le but de l'application est de connaître la réputation des hommes politiques sur Twitter en vue des élections présidentielles françaises de 2017.

La finalité du projet est d'avoir une carte interactive de la France pour connaître la répartition des tweets. Il y aura également une série de statistiques pour mieux comprendre et mettre en perspective les informations. Il y aura 3 niveaux de précision : la région, le département, et la ville qui changeront selon le zoom sur la carte.

L'outil proposera deux moyens de visualiser les données sur la carte :

- Un mode comparaison : Dans chaque zone (ville, département ou région) la couleur du candidat le plus populaire ressortira.
- Un mode candidat unique : pour chaque candidat on pourra connaître la répartition de ses tweets grâce à une échelle de couleurs.

Pour récupérer toutes ces données l'application « écoute » le réseau social twitter, et dès qu'un tweet est envoyé en rapport avec un ou plusieurs candidats à la présidentielle, l'application sauvegarde le candidat concerné, la ville d'où a été envoyé le tweet, et l'horodatage du post.

```
Hamon,Paris,2017-03-24 21:31:08
Melenchon,Paris,2017-03-24 21:31:08
Le pen,Paris,2017-03-24 21:31:08
fillon,Paris,2017-03-24 21:31:39
Macron,Ramonville-Saint-Agne,2017-03-24 21:32:12
Macron,Saint-Aubin-Celloville,2017-03-24 21:33:41
fillon,Porcheville,2017-03-24 21:33:56
Hamon,Dreux,2017-03-24 21:34:18
fillon,France,2017-03-24 21:35:07
```

Partie du fichier où sont stockés tous les tweets.

Mouvance politique	Parti politique/mouvement	Candidat	Mots clés
Droite & extrême droite	Les républicains	François Fillon	#LR, #lesrepublicains, #Fillon, #FF2017, #fillon2017
Gauche et extrême gauche	Parti Socialiste	Benoît Hamon	#hamon, #PS, #PartiSocialiste

Extrait du tableau des mots clés utilisés

Ici la réputation n'est pas prise en compte comme un facteur qualitatif mais seulement comme un facteur quantitatif. Même si le tweet est négatif sur un homme politique, il sera pris en compte.

État d'avancement

Phase 1 : Récupération des données

Cette phase était essentielle puisque les données sont la base de l'application. Pour cette phase, l'API twitter et la librairie python qui a implémenté cette API : « tweepy » ont été utilisés. Ils sont utilisés en mode « streaming », c'est-à-dire que les tweets sont récupérés en temps réels et le script ne s'arrête pas tant qu'il n'y a pas eu de demande d'arrêt émise.

Ce script se focalise sur les tweets intéressants (donc en rapport avec les élections présidentielles), c'est pourquoi il a été mis en place les filtres comme montrés dans le tableau des mots clés utilisés dont une partie est dans la partie 1.

L'application est également concentrée sur les tweets provenant de France. Cette concentration géographique a deux effets.

- Évidemment les tweets provenant d'autres pays sont ignorés
- Mais les tweets qui ne possèdent pas de géolocalisation sont ignorés également

Phase 2 : Sauvegarde des tweets

Quand un tweet est récupéré, un fichier Json nous est envoyé. Ce fichier Json est intéressant, mais il possède beaucoup trop d'informations que nécessaire, comme par exemple le pseudo de l'utilisateur, la couleur de son profil, un booléen pour savoir si c'est une réponse ou pas etc. Un exemple du fichier Json est disponible en annexe 1.

Les champs qui nous intéressent sont donc le texte (champ text, pour savoir quel candidat est visé), la ville (champ place>name), et la date (champ created_at).

Ces informations sont ensuite exportées vers un fichier csv (Comma-separated-values) grâce à la bibliothèque csv de python. Un extrait est également présent dans la partie 1

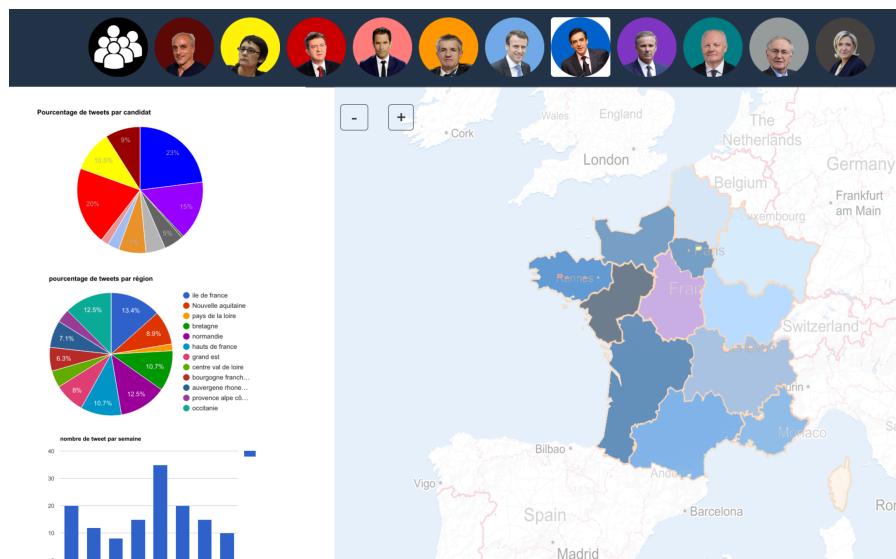
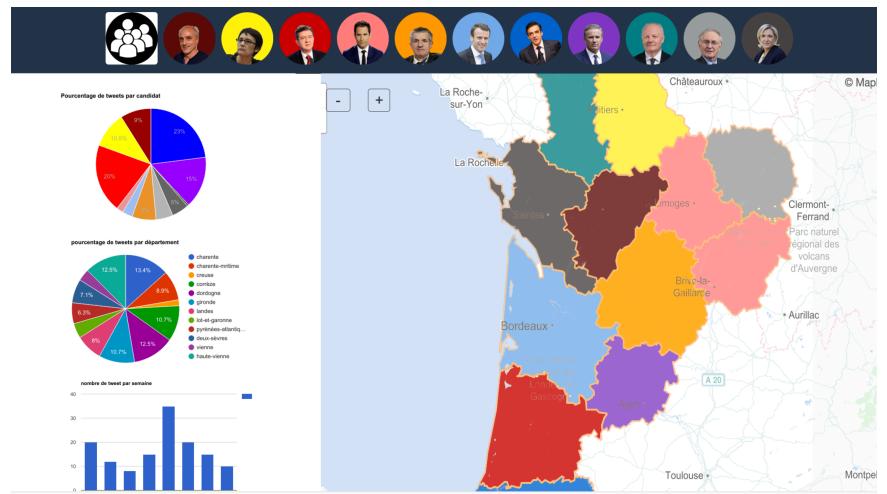
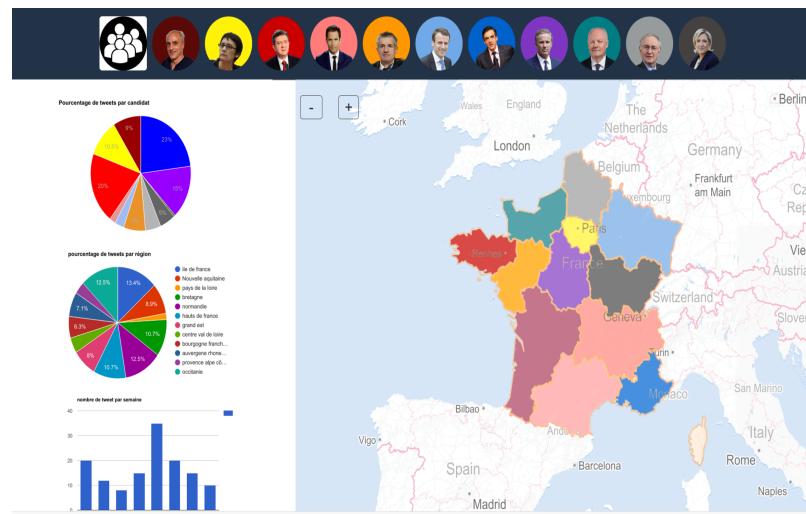
Phase 3 : Agrégation des informations

Pour cette phase, c'est un serveur écrit en node js, grâce notamment au module expressjs, qui permet d'agrégner les informations. Il associe à chaque ville tous les tweets émis depuis celle-ci.

"10055": {"Poutou": 2, "fillon": 4}

Extrait du fichier Json

Ici la ville n'est plus représentée par son nom mais par son code INSEE, qui a été récupéré grâce à un fichier regroupant toutes les communes françaises et fournies grâce à data.gouv.fr.



Visuels de l'application sous axure

Les problèmes rencontrés

Au début du projet, l'outil devait analyser non seulement twitter mais également google trends et facebook. Cependant il n'a pas été possible de se baser sur ces deux autres médias.

Facebook filtre énormément les requêtes de ce genre pour respecter la vie privée de ses utilisateurs. Il est donc assez difficile de récupérer des informations pertinentes sur ce réseau social. Pour Google trends, il n'existe tout simplement pas d'API officielle pour récupérer les informations.

Sur twitter il a également été assez difficile de se repérer dans la documentation de l'API. En effet, sur le fichier json que renvoie l'API, certaines informations semblaient intéressantes mais ne l'étaient pas, comme par exemple les champs geo et coordinates pour récupérer la géolocalisation. Ces deux champs représentent la position du téléphone, et cette option est généralement désactivée. Il a fallu un peu de temps avant de repérer le champ place.name qui est la ville d'origine rentrée par l'utilisateur et qui est plus fréquemment utilisée.

Perspective et suite du projet

Pour la suite du projet, il ne reste « plus qu'à » mettre en place la visualisation des données via des outils comme leaflet et mapbox qui ont d'ailleurs été rajoutés au projet.

S'il reste assez de temps, il est également envisagé de séparer complètement le « front » et le « rest », avec une API qui fournit les données à visualiser selon les options données (grâce à un outil comme flask-restful notamment), et un front qui ne gère que la visualisation.

Les statistiques sont également en cours de traitement, avec l'utilisation d'outil comme « chartjs ».

Annexes

Annexe1

```
{  
  "text": "RT @PostGradProblem: In preparation for the NFL lockout, I will be spending twice as much  
  time analyzing my fantasy baseball team during ...",  
  "truncated": true,  
  "in_reply_to_user_id": null,  
  "in_reply_to_status_id": null,  
  "favorited": false,  
  "source": "<a href=\"http://twitter.com/\" rel=\"nofollow\">Twitter for iPhone</a>",  
  "in_reply_to_screen_name": null,  
  "in_reply_to_status_id_str": null,  
  "id_str": "54691802283900928",  
  "entities": {  
    "user_mentions": [  
      {  
        "indices": [  
          3,  
          19  
        ],  
        "text": ""  
      }  
    ]  
  }  
}
```

```

        "screen_name": "PostGradProblem",
        "id_str": "271572434",
        "name": "PostGradProblems",
        "id": 271572434
    }
],
"urls": [],
"hashtags": []
},
"contributors": null,
"retweeted": false,
"in_reply_to_user_id_str": null,
"place": null,
"retweet_count": 4,
"created_at": "Sun Apr 03 23:48:36 +0000 2011",
"retweeted_status": {
    "text": "In preparation for the NFL lockout, I will be spending twice as much time analyzing my fantasy baseball team during company time. #PGP",
    "truncated": false,
    "in_reply_to_user_id": null,
    "in_reply_to_status_id": null,
    "favorited": false,
    "source": "<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">HootSuite</a>",
    "in_reply_to_screen_name": null,
    "in_reply_to_status_id_str": null,
    "id_str": "54640519019642881",
    "entities": {
        "user_mentions": [],
        "urls": [],
        "hashtags": [
            {
                "text": "PGP",
                "indices": [
                    130,
                    134
                ]
            }
        ]
    },
    "contributors": null,
    "retweeted": false,
    "in_reply_to_user_id_str": null,
    "place": null,
    "retweet_count": 4,
    "created_at": "Sun Apr 03 20:24:49 +0000 2011",
    "user": {
        "notifications": null,
        "profile_use_background_image": true,
        "statuses_count": 31,
        "profile_background_color": "C0DEED",
        "followers_count": 3066,
        "profile_image_url": "http://a2.twimg.com/profile_images/1285770264/PGP_normal.jpg",
        "listed_count": 6,
        "profile_background_image_url": "http://a3.twimg.com/a/1301071706/images/themes/theme1/bg.png",
        "description": "",
        "screen_name": "PostGradProblem",
        "default_profile": true,
        "verified": false,
        "time_zone": null,
        "profile_text_color": "333333",
        "is_translator": false,
        "profile_sidebar_fill_color": "DDEEF6",
    }
}
]
}

```

```

    "location": "",
    "id_str": "271572434",
    "default_profile_image": false,
    "profile_background_tile": false,
    "lang": "en",
    "friends_count": 21,
    "protected": false,
    "favourites_count": 0,
    "created_at": "Thu Mar 24 19:45:44 +0000 2011",
    "profile_link_color": "0084B4",
    "name": "PostGradProblems",
    "show_all_inline_media": false,
    "follow_request_sent": null,
    "geo_enabled": false,
    "profile_sidebar_border_color": "C0DEED",
    "url": null,
    "id": 271572434,
    "contributors_enabled": false,
    "following": null,
    "utc_offset": null
  },
  "id": 54640519019642880,
  "coordinates": null,
  "geo": null
},
"user": {
  "notifications": null,
  "profile_use_background_image": true,
  "statuses_count": 351,
  "profile_background_color": "C0DEED",
  "followers_count": 48,
  "profile_image_url":
  "http://a1.twimg.com/profile_images/455128973/gCsVUnofNqqyd6tdOGevROvk01_500_normal.jpg",
  "listed_count": 0,
  "profile_background_image_url":
  "http://a3.twimg.com/a/1300479984/images/themes/theme1/bg.png",
  "description": "watcha doin in my waters?",
  "screen_name": "OldGREG85",
  "default_profile": true,
  "verified": false,
  "time_zone": "Hawaii",
  "profile_text_color": "333333",
  "is_translator": false,
  "profile_sidebar_fill_color": "DDEEF6",
  "location": "Texas",
  "id_str": "80177619",
  "default_profile_image": false,
  "profile_background_tile": false,
  "lang": "en",
  "friends_count": 81,
  "protected": false,
  "favourites_count": 0,
  "created_at": "Tue Oct 06 01:13:17 +0000 2009",
  "profile_link_color": "0084B4",
  "name": "GG",
  "show_all_inline_media": false,
  "follow_request_sent": null,
  "geo_enabled": false,
  "profile_sidebar_border_color": "C0DEED",
  "url": null,
  "id": 80177619,
  "contributors_enabled": false,
  "following": null,
}

```

```
        "utc_offset": -36000
    },
    "id": 54691802283900930,
    "coordinates": null,
    "geo": null
}
```