

PROJET : MY FIRST CHATBOT

Membres de l'équipe :

-SCORDIA Matthieu

-ANGLARD Loïc

Groupe de TP :

-PMP L1-Plus

Année scolaire :

-2023/2024

Introduction :

Ce projet consiste à concevoir un chatbot capable d'analyser des textes (plus précisément 8 discours de présidents) afin de répondre à des questions précises. Pour réaliser ce projet nous avons utilisé le système d'exploitation pycharm et avons codé l'entièreté de notre programme en python. Ce projet se fait en binôme, c'est pourquoi nous avons utilisé Git afin d'assurer une bonne cohésion.

Le projet vise donc à :

- 1) Analyser les discours de présidents français.
- 2) Extraire des informations sur les présidents à partir des fichiers texte.
- 3) Utiliser la méthode TF-IDF pour représenter les mots sous forme de vecteurs numériques (score).
- 4) Développer des fonctions pour analyser la fréquence et l'importance des mots dans le corpus.
- 5) Créer un programme principal offrant des fonctionnalités d'analyse et un mode Chatbot pour répondre aux questions basé sur les discours analysés.

Les différentes fonctionnalités proposées à l'utilisateur :

L'utilisateur possède une interface utilisateur dans laquelle il peut interagir. Au départ il lui est offert 6 possibilités de choix , chacune d'elles répondant à un besoin spécifique.

```
Si vous voulez afficher la liste des mots les moins importants dans le corpus de documents entrez 1. \n "  
Si vous voulez afficher le(s) mot(s) ayant le score TD-IDF le plus élevé entrez 2. \n "  
Si vous voulez indiquer le(s) mot(s) le(s) plus répété(s) par le président Chirac entrez 3. \n "  
Si vous voulez indiquer le(s) nom(s) du (des) président(s) qui a (ont) parlé de la « Nation » et celui qui l'a répété le plus de fois entrez 4. \n "  
Si vous voulez indiquer le premier président à parler du climat et/ou de l'écologie entrez 5. \n "  
Hormis les mots dits « non importants », Si vous voulez le(s) mot(s) que tous les présidents ont évoqués entrez 6. \n "  
Si vous n'avez plus de demandes entrez 7.''))
```

Ensuite, toujours avec la même interface utilisateur, il peut, s'il le souhaite, rédiger sa propre question. Celle-ci sera traitée et le chatbot établira une réponse en conséquence.

```
Si vous voulez poser n'importe quelle autres questions entrez 7  
Si vous n'avez plus de demandes entrez 8.?  
Posez n'importe quelle question:
```

Ce projet est fonctionnel car il offre des solutions concrètes aux demandes spécifiques énoncées. Il permet d'extraire des informations pertinentes à partir des discours des présidents français, de les traiter efficacement (nettoyage, analyse de mots clés, calculs TF-IDF), et enfin de fournir des réponses automatiques basées sur ces analyses pour répondre à des questions posées. Son architecture est conçue pour fonctionner de manière itérative et efficace, offrant ainsi une fonctionnalité réelle et utilisable dans le cadre défini.

Comment notre programme réalise les différentes fonctionnalités proposées:

1/ Tout d'abord dans une première étape nous avons dû **extraire les informations des discours présidentiels**. Pour cela il nous a fallu lire les fichiers et associer à chaque nom de président le prénom correspondant.

2/ Ensuite nous avons dû **nettoyer les textes**. C'est-à-dire que nous avons converti chaque caractère en caractères minuscules, supprimer les signes de ponctuation et traité les caractères spéciaux comme les accents.

3/ Puis nous avons établi des **scores pour chaque caractères de chaque textes**, ce sont les TF-IDF. Pour cela il a tout d'abord fallu compter les occurrences de chaque mots de chaque textes puis les comparer entre eux. Cela nous a permis de répondre aux demandes de la première partie du projet soit le fait de pouvoir répondre à des demandes spécifiques par rapport aux textes, comme les mots les plus importants (TF-IDF le plus élevé) ou bien le nombre d'occurrence d'un mot dans un texte.

4/ Pour réaliser la deuxième partie du projet nous avons dû réaliser un **nouveau TF-IDF**, mais cette fois-ci pour la question posée, afin de trouver le mot le plus important. Cela nous a ensuite permis de réaliser une **analyse de similarité** afin de renvoyer une **réponse pertinente**, dans ce cas-là, la phrase d'un des discours de président qui était la plus approprié.

Justification des choix de structure :

1/ Pour faciliter le travail en binôme nous avons réalisé un **tableau excel** comportant chacune de nos variables, nos fonctions et nos dictionnaires afin de ne pas utiliser une variable, fonction, dictionnaire, déjà existant. Cela n'était pas spécifié dans les consignes du projet.

2/ Utilisation de **variables** tout au long du programme.

3/ Utilisation de **listes** afin de mieux stocker et manipuler les données (textes, réponses, questions) et de faciliter les calculs (TF-IDF, Analyse de similarité).

4/ Utilisation de **dictionnaires** afin de stocker les listes ou encore les TFIDF (TF-IDF_clean/rep/etc..)

Principales difficultés :

La première difficulté a été la manipulation de git, nous avons eu du mal au début du projet à bien prendre en main ce système. Pour répondre à cette difficulté nous nous sommes renseigné sur des forums et auprès de nos professeurs afin d'apprendre à manipuler le système. Ensuite l'une des principales difficultés concernait les formules données.

En effet, ces formules ont été changées à de nombreuses reprises ce qui a impacté la réalisation du programme. Pour répondre à ce second problème nous avons cherché des formules par nous-même et avons continué le programme en prenant en compte ces failles.

La dernière difficulté était le manque de temps ce qui nous a empêché de concevoir une interface utilisateur plus intuitive. Afin de répondre à ce problème nous nous sommes répartis équitablement la charge de travail afin d'avancer efficacement.

Présentation des résultats :

Partie 1/

Les résultats sont très concluants, le programme fonctionne parfaitement. Voici ci-dessous plusieurs illustrations de ces résultats :

```
Si vous voulez afficher la liste des mots les moins importants dans le corpus de documents entrez 1.
Si vous voulez afficher le(s) mot(s) ayant le score TD-IDF le plus élevé entrez 2.
Si vous voulez indiquer le(s) mot(s) le(s) plus répété(s) par le président Chirac entrez 3.
Si vous voulez indiquer le(s) nom(s) du (des) président(s) qui a (ont) parlé de la « Nation » et celui qui l'a répété le plus de fois entrez 4.
Si vous voulez indiquer le premier président à parler du climat et/ou de l'écologie entrez 5.
Hormis les mots dits « non importants », Si vous voulez le(s) mot(s) que tous les présidents ont évoqués entrez 6.
Si vous voulez poser n'importe quelle autre question entrez 7
Si vous n'avez plus de demandes entrez 8.2
Les mots les plus forts sont : ['doit']

Si vous voulez afficher la liste des mots les moins importants dans le corpus de documents entrez 1.
Si vous voulez afficher le(s) mot(s) ayant le score TD-IDF le plus élevé entrez 2.
Si vous voulez indiquer le(s) mot(s) le(s) plus répété(s) par le président Chirac entrez 3.
Si vous voulez indiquer le(s) nom(s) du (des) président(s) qui a (ont) parlé de la « Nation » et celui qui l'a répété le plus de fois entrez 4.
Si vous voulez indiquer le premier président à parler du climat et/ou de l'écologie entrez 5.
Hormis les mots dits « non importants », Si vous voulez le(s) mot(s) que tous les présidents ont évoqués entrez 6.
Si vous voulez poser n'importe quelle autre question entrez 7
Si vous n'avez plus de demandes entrez 8.4
Le premier président à parler de la Nation est François Mitterrand et il en parle 1 fois.
```

Partie 2/

Pour ce qui est de la partie 2, le programme marche correctement. Toutefois les fonctionnalités de la partie 3 que nous n'avons pas pu coder en vue d'un manque de temps aurait pu enrichir les réponses du chatbot.

Illustration des resultats de la partie 2 :

```
Posez n'importe quelle question:  
Comment être écologique ?  
Après analyse, il est temps d'engager la transition énergétique et écologique.  
|
```

```
Posez n'importe quelle question:  
Pour ça on peut comment une nation peut elle prendre soin du climat ?  
Oui, bien sûr! et je songe bien sûr à François Hollande, faisant oeuvre de précurseur avec l'accord de Paris sur le climat et protégeant les Français dans un monde frappé par le terrorisme.
```

```
Posez n'importe quelle question:  
Comment une nation peut elle prendre soin du climat ?  
Après analyse, et je songe bien sûr à François Hollande, faisant oeuvre de précurseur avec l'accord de Paris sur le climat et protégeant les Français dans un monde frappé par le terrorisme.  
|
```

Conclusion : Bilan et enseignements retenus

1/ Plan technique :

Ce projet a été une excellente source d'apprentissage. Tout d'abord sur le plan technique, ce projet nous a permis de maîtriser le langage de programmation Python. En effet les consignes demandées nous ont permis d'apprendre à réaliser un chatbot mais aussi plus précisément à gérer et traiter des fichiers, réaliser des fonctions efficaces ainsi que savoir les utiliser et à concevoir un programme, d'une assez grande ampleur, et de le faire fonctionner correctement. D'autre part l'autonomie nous a permis de gérer nos propres erreurs ce qui nous a d'autant plus appris de nouvelles choses par nous-même. (Par exemple nous avons appris à situer une erreur dans le code et à la réparer).

2/ Plan relationnel :

Pour ce qui est de l'organisation du travail et la communication au seins de l'équipe, tout s'est bien déroulé. Comme dit précédemment, nous avons mis au point un tableau excel contenant nos variables, fonctions et autre afin de s'assurer de ne pas utiliser une variable déjà défini. D'autre part, l'utilisation du système git et git-hub nous a permis de concentrer notre travail en un seul point et de pouvoir programmer simultanément sur pycharm. Ce projet nous a appris à travailler en équipe, à savoir répartir une quantité de travail équitable et surtout a savoir rassembler clairement nos idées et les appliquer.

3/ Gestion du temps :

En ce qui concerne la gestion du temps, nous avons bénéficié des heures dédiés à la gestion du projet, mais en vue de la quantité de travail a fournir pour finaliser ce projet nous avons dû travailler davantage en dehors des cours. Etant en classe de prépa plus, nous avons dû répartir nos heures de travail sans impacter les autres enseignements (le week-end ou en fin d'après-midi). Cela nous a appris a gérer efficacement notre temps afin de répondre aux attentes du projet (date de rendu des parties) tout en ne laissant pas de côté les autres matières.

4/ Conclusion :

En conclusion ce projet nous a été une source d'apprentissage en tout termes. La réalisation d'un chatbot est un sujet qui nous plonge dans l'actualité et nous illustre les taches qui peuvent nous être administrées en tant qu'ingénieur informatique. Ce fu un plaisir de réaliser ce projet.