

Compte rendu de réunion

Le 09/11/2021 à 16h-17h à l'université de Paul Valéry

Objet : Première exploration des données

Ordre du jour :

- ! Présenter nos avancées
- ! Difficultés rencontrées
- ! Plan des actions à réaliser

Légende du tableau ci-dessous : P = Présent, A = Absent, E = Excusé, NC = non concerné par la réunion

Liste des participants au projet			
Nom	Email	Fonction	Présent
Sandra BRINGAY	sandra.bringay@univ-montp3.fr	Encadrante pédagogique	P
Florian LOMBARDO	florian.lombardo@univ-montp3.fr	Autre encadrant pédagogique	P
Laura SENECAILLE	laura.senecaille@etu.univ-montp3.fr	Etudiante	P
Anamé ROUMY	aname.roumy@etu.univ-montp3.fr	Etudiante	P
Lisa BETEILLE	lisa.beteille@etu.univ-montp3.fr	Etudiante	P
Matéo CALSACY	mateo.calsacy@etu.univ-montp3.fr	Etudiant	P
Jean CHABANOL	jean.chabanol@etu.univ-montp3.fr	Etudiant	P
Célia TEYSSIER	celia.teyssier@etu.univ-montp3.fr	Etudiante	P

Sommaire

Avancées.....	2
Difficultés rencontrées.....	2
Plan d'organisation.....	2
Actions à entreprendre.....	2

Avancées

- Les lectures en anglais des documents nécessaires à l'état de l'art sont en cours. Vous nous avez conseillé de les lire en diagonale et de les comprendre, pour nous permettre de nous positionner par rapport à ce qui a déjà été fait.
- Le site internet pour présenter nos résultats. Chaque onglet sera dédié à une catégorie de données (Twitter, Reddit, Wikipédia) avec des explorations de visualisations, les résultats des algorithmes de décisions...

Difficultés rencontrées

- Nous avons rencontré des erreurs sur l'algorithme de Wikipédia qui nous empêchaient de télécharger les données et donc de travailler dessus.

Plan d'organisation

- Pour plus d'efficacité, nous nous sommes réparti le travail :
 - o L'état de l'art et les résumés des articles sont pris en charge par Matéo et Jean.
 - o Laura s'occupera de créer le site internet qui représente nos résultats.
 - o Anamé s'occupera de la partie Visualisation des données.
 - o Célia et Lisa de la partie Nettoyage des données.
 - o Anamé, Laura, Célia et Lisa de la partie Exploration des données sur Iramuteq.
 - o La partie sur la classification sera prise en charge par tout le monde.

Actions à entreprendre

- La partie visualisation des données ne peut pas être démarrée pour le moment car les données ne sont pas encore entièrement nettoyées. Nous pourrions visualiser les données sur python même si elles ont été nettoyées sur R, c'est nous qui pouvons décider.
- Pour la partie nettoyage des données, nous avons pour le moment seulement le dossier Reddit complet (avec données relatives au « suicide » et d'autres à la « joie »).
 - o Sandra Bringay va nous envoyer les mots clés qu'elle a utilisés pour exporter les données.

Pour les données sur Wikipédia, comme pour les données Twitter, nous avons seulement les données concernant le suicide. Mais l'algorithme pour Wikipédia ne fonctionne plus, nous devons prendre contact avec Mathéo Daly pour qu'il nous aide à régler cette erreur.

- Les données Reddit étant complètes, nous devons commencer à faire de la classification dessus.