**CR – D4G – Validalab 07/04/21**

Présents (6) :

Corentin

Yann

Jean-Marc

Serge

David

Anastasia

**Réflexion sur la crédibilité d’une source :**

Plutôt par catégories

**Réflexion sur la base SQL :**

Ça dépend des viz qu’on veut avoir, parfois il faudra repasser en SQL. Même si pour certaines on peut exploiter le graph avec cypher, ça peut être couteux au niveau de la recherche. Choix architectural 🡪 il faut réfléchir ce qu’on va mettre, faire la maquette produit.

**Ce qui a été fait cette semaine :**

Jean-Marc : a assisté Yann dans la mise en place des données

Yann : cf suite

Anastasia : xmind

**Outils mis en place par Yann :**

* Serveur accesssible via ssh, accès shell et des mdp créés
* Dataiku : permet de mettre des projets data, prototyper facilement des flows de données, intégration avec git, permet de pooler facilement. Community edition 3 utilisateurs 🡪 voir qui a quel compte dans le fichier validalab-info

**Tracks principales :**

* industrialiser le scraping (HYPHE tourne bien, on pourrait le mettre dans un docker pour qu’il tourne en continu…)
* donner la possibilité à l’utilisateur de rajouter un nouveau site
* enrichissement de la base de données
* app à mettre en place (Flask qui pour l’instant tourne que sur la machine de Jean-Marc)

**Qui veut faire quoi ?**

* Serge : plutôt industrialisation et mise en place du site, transformation des données Graph en bases SQL
* David : architecture et industrialisation
* Corentin : enrichissement de la base de données
* Anastasia : enrichissement de la base de données, data engineering

**Pour la semaine prochaine :**

* Jean-Marc : maquette produit en draft
* David et Serge : regardent en détail le code pour proposer des choses en industrialisation
* Corentin et Anastasia : faire une liste des plateformes de fact-checking, voir s’il y a des APIs et comment on pourrait les scraper et les intégrer dans le Graph.

Prochaine réunion : mercredi 14 avril à 19h30 heure de Paris