Tableau de bord Lifranum

13.11

Les données sur lesquelles on veut travailler dans un premier temps : « content ». Est-il possible de travailler uniquement sur les content de wordpress ? Travailler sur toutes les données wordpress nous parait complexe, nous nous demandons donc comment couper les données ? Quels critères utilisés pour sélectionner certaines données et pas d’autres ? Peut-être pouvons-nous privilégier des données complètes, c’est-à-dire des blogs où ils ne manquent aucune information.

Nos idées :

* Utiliser BERT pour travailler sur les « content » et dégager des thématiques. A partir de ces thématiques, travailler plus finement sur le texte et apporter un regard plus SHS (pour le moment on ne peut rien dire, notre travail dépendra de ce que nous allons trouver comme thématiques)
* Travailler sur les auteurs, essayer de trouver des critères communs (nous ne sommes pas surs de la faisabilité de cette partie, avons-nous les données nécessaires ?)
* Nous aimerions aussi, récupérer les liens des blogs et faire une représentation réseau (découle du premier point et des thématiques que l’on va trouver)