

Comment mettre des tweets en cube ?

jerome.darmont@univ-lyon2.fr et sabine.loudcher@univ-lyon2.fr

Contexte

Le stage se déroule dans le cadre du projet Identités Numériques Urbaines (IDENUM) qui porte sur les représentations de la ville à travers les réseaux sociaux. La recherche est focalisée sur des corpus complexes constitués par les documents photographiques produits et diffusés par des usagers de la ville, ainsi que par des acteurs collectifs (collectivités territoriales, institutions patrimoniales, associations, réseau de transport public...) et collectés dans les réseaux sociaux. Dans ce cadre, près de 500 000 tweets ont été collectés sur le thème de la Fête des Lumières à Lyon en 2015 et 2016.

Objectif

L'objectif du stage est d'implémenter l'entrepôt de tweets dont le modèle conceptuel est donné dans la Figure 1 (Haddad, 2018). Contrairement aux entrepôts de données numériques, l'originalité de l'approche est de placer les tweets (avec leur contenu textuel et leur structure comme par exemple les retweets ou leur géolocalisation) comme objet principal d'analyse. Il s'agit ensuite, à partir des cubes de données issus de l'entrepôt, de proposer des analyses en ligne (OLAP) et des visualisations.

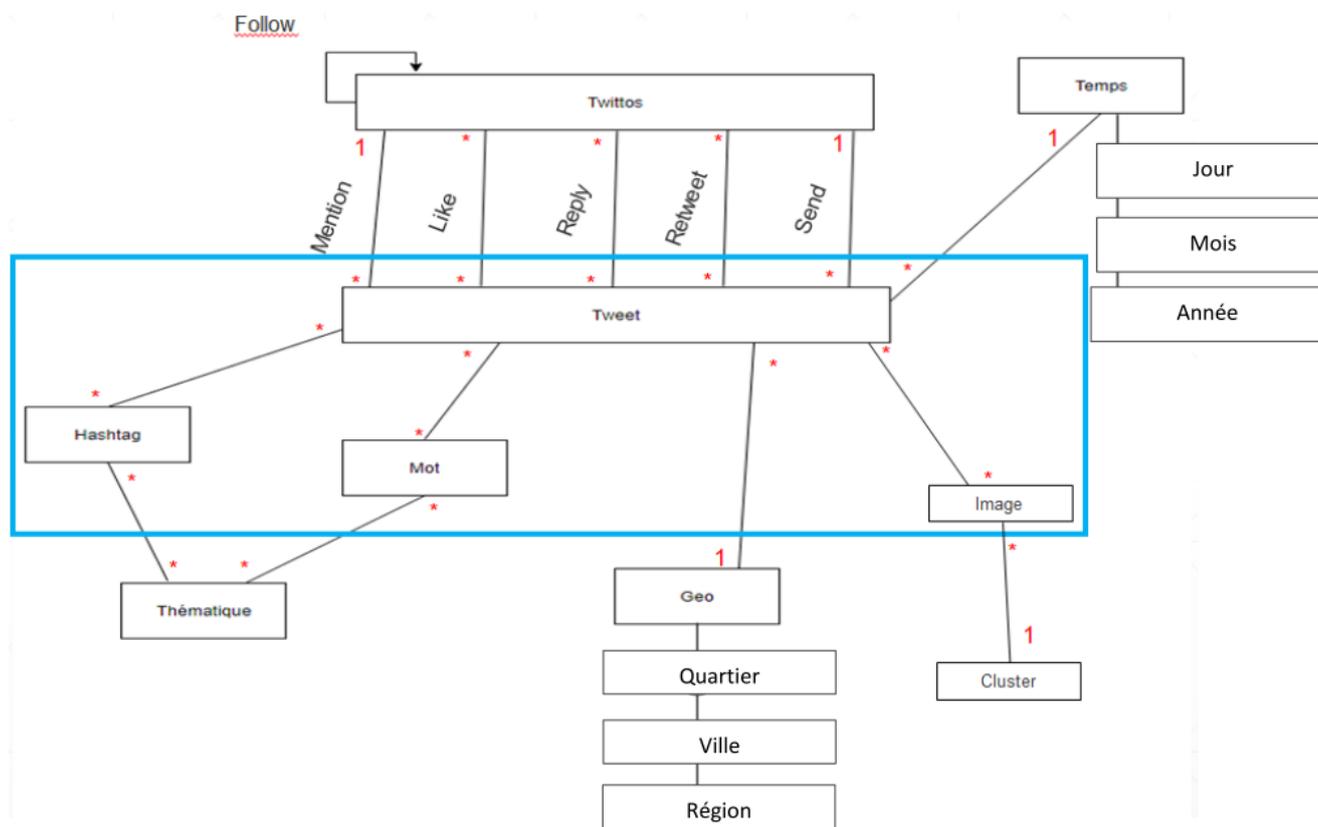


Figure 1 : Modèle conceptuel de l'entrepôt de données de tweets

Tâches à effectuer

1. Lire le mémoire de master d'Elie Haddad dont la référence est donnée ci-dessous, ainsi que les articles qu'il référence. Trouver d'autres références, si possible.
2. Implémenter l'entrepôt de données dans un SGBD relationnel comme PostgreSQL¹ ou MariaDB².

¹ <https://www.postgresql.org/>

² <https://mariadb.org/>

3. Implémenter les analyses OLAP avec notamment le développement d'opérateurs d'agrégation adaptés au contenu textuel des tweets. L'utilisateur doit pouvoir analyser les tweets selon les hashtags ou les mots utilisés, selon la thématique mais aussi en fonction de la localisation des tweets ou de la classe (*cluster*) des images associées aux tweets.
4. Écrire un article d'une dizaine de page maximum, de préférence à l'aide de LaTeX³, en français ou en anglais, qui présente votre travail. Plan de l'article :
 - a. Introduction (contexte, problématique, contribution, annonce du plan de l'article)
 - b. État de l'art sur les entrepôts de tweets
 - c. Description de l'implémentation de l'entrepôt (modèle logique relationnel, tâches d'ETL)
 - d. Analyses OLAP, visualisations et résultats
 - e. Conclusion

Première piste de lecture

Haddad, E. (2018). Comment mettre des tweets en cube ? Mémoire de master, Université de Bourgogne.
Encadrants : Darmont, J., Loudcher, S.

http://eric.univ-lyon2.fr/jdarmont/idsm/Haddad_Elie_RapportBDIA2018_LaboERIC.pdf

³ <https://www.latex-project.org>