

Catégorisez automatiquement des questions

# Introduction

Poser une question sur Stack Overflow est très simple, plus dur est de trouver les bons tags permettant de catégoriser votre question. Le but de ce projet est de construire un système de suggestion de tags utilisant en entrée le titre et le corps de la question que vous posez. L’idée est donc de récupérer dans un premier temps des données réelles, disponibles sur Stack Exchange, de les analyser, de construire un modèle permettant de générer des tags, et enfin de les inclure dans une IHM dédiée que l’on mettra à disposition.

# Visualisation des données

De nombreux champs sont disponibles

# Nettoyage

# Word occurrence

# Méthode non supervisée

LDA + TF

NMF + TF IDF

# Méthode supervisée

Sparses matrices

Multi output – multi label

Predict vs predict proba

# Comparaison des résultats

# Evaluation