



Impression à partir d'une page du site internet de l'AMF

21 avril 2021

L'AMF remercie les nombreux participants à son challenge Entity Name Matching sur Codalab

Du 13 janvier au 12 mars 2021, l'AMF a lancé un défi aux codeurs passionnés : proposer une méthode d'identification efficace des acteurs de marchés à l'aide de techniques de traitement du langage naturel. Le challenge a été remporté par Robert Stanca, étudiant en informatique à l'Université Politehnica de Bucarest.

Le 13 janvier 2021, l'Autorité des marchés financiers lançait son challenge Entity Name Matching sur Codalab, une plateforme open-source spécialisée dans les compétitions de datascience.


Le challenge invitait les participants à proposer une méthode d'identification efficace des acteurs de marché : à partir du nom d'un intervenant mentionné dans un document de façon plus ou moins approximative, il s'agissait de déterminer son LEI (Legal Entity Identifier ou Identifiant d'entité juridique), autrement dit son code unique d'identification. Les solutions proposées pourraient permettre à l'AMF de mieux croiser les différentes sources de données pour n'obtenir qu'une seule et même vue consolidée.

Le challenge s'est achevé le 12 mars 2021. L'AMF remercie l'ensemble des contributeurs et en particulier le lauréat Robert Stanca. A l'aide de techniques pour le traitement du langage naturel (NLP), Robert Stanca a proposé une approche permettant de calculer la similarité entre le nom de l'entité mentionnée dans un document et les données du répertoire mondial des identifiants d'entité juridique (LEI).

Pour calculer cette ressemblance, il faut d'abord transformer le texte (comme le nom des entités) en valeurs numériques : c'est ce qu'on appelle l'encoding. Mais cette étape ne doit pas se faire au hasard : en réalité, on transpose le texte dans un espace vectoriel de sorte à ce que celui-ci conserve les propriétés du texte. Par exemple les vecteurs associés aux mots « maison » et « appartement » doivent être assez proches l'un de l'autre parce qu'ayant un sens proche ; les mots « participer » et « participant » aussi parce que partageant une racine commune.

De cette façon, l'algorithme détecte dans le référentiel quelle serait la ligne qui contient la valeur la plus proche du nom de l'intervenant recherché, et, puisque le référentiel contient également le LEI associé, le tour est joué !

SUR LE MÊME THÈME

 S'abonner à nos alertes et flux RSS

ACTUALITÉ INNOVATION

19 avril 2022

L'AMF participe à la seconde édition de la French Fintech Week



ACTUALITÉ CRYPTO-ACTIFS

26 janvier 2022

Nouvelle étape dans l'adoption du règlement Régime pilote pour les infrastructures de marché sur blockchain



PRISE DE PAROLE

INNOVATION

25 octobre 2021

Discours de Robert Ophèle, président de l'AMF - Conférence Banque de France/Université d'Orléans : "Crypto-actifs et cybersécurité : innovations et défis - Les enjeux de la..."



Mentions légales :

Responsable de la publication : Le Directeur de la Direction de la communication de l'AMF. Contact :
Direction de la communication, Autorité des marchés financiers - 17, place de la Bourse - 75082 Paris
Cedex 02