

Le Traitement Automatique du Langage au service de la détection des levées de fonds dans les corpus de presse



VINET Emmanuel, evinet@idaia.group, Groupe IDAIA, Bordeaux

CATEGORIE 1 : Récupération des données : Tracking, crawling, consolidation, parsing,...

Résumé : (en 6 lignes maximum)

Pour aider les entreprises à identifier de réelles opportunités de business ou à réaliser de la veille concurrentielle et/ou sectorielle, il est nécessaire de combiner une véritable expertise en technologies de Data Science et une très bonne connaissance métier. Pour cela, par sa cellule de R&D « BIG DATA LAB », le Groupe IDAIA a utilisé l'intelligence artificielle et des algorithmes et a développé un robot de traitement automatique du langage qui détecte les levées de fonds dans les corpus de presse.

Mots clés : (maximum 6)

IA, NLP, algorithme, analyse sémantique, parsing, text mining

1. Introduction

Pour aider les entreprises à identifier de réelles opportunités de business ou à réaliser de la veille concurrentielle et/ou sectorielle, il est nécessaire de combiner une véritable expertise en technologies de Data Science et une très bonne connaissance métier. Des besoins sur lesquels s'applique le BIG DATA LAB, la cellule R&D du Groupe IDAIA (ex-Groupe CARTÉGIE), qui a élaboré des algorithmes permettant de réaliser notamment une analyse sémantique complète d'articles de presse économique.

2. Méthodologie

Au cours de cette conférence, nous illustrerons leur travail d'analyse de contenus afin de détecter un certain nombre d'indicateurs et d'identifier précisément les entreprises qui sont citées. Seront notamment abordées les techniques d'étiquetage morpho-syntaxique, de parsing, de coréférence et de reconnaissance avancée d'entités. A partir des bibliothèques NLP du groupe de recherche de l'Université de Stanford, le BIG DATA LAB a développé une suite d'algorithmes de traitement du langage en langue française et bâti un ensemble de règles pour effectuer des détections d'objets sémantiques métiers. Une plateforme écrite en Java et Scala sous Spark permet d'effectuer ces traitements d'analyse sur des corpus de grands volumes à des fréquences rapides. Des dashboard de datavisualisation actualisés en temps réel complètent ces traitements.

3. Originalité / perspective

Grâce à ce savoir-faire, le Groupe IDAIA (ex-Groupe CARTÉGIE) est capable de proposer à ses clients de nouveaux services et des outils de décisions performants pour qu'ils puissent toujours mieux connaître leurs propres clients et mener à bien leurs projets.