

COMMENT OPTIMISER L'EXPLOITATION DES DONNÉES DE PRESSE À PARTIR DES TECHNIQUES D'IA

URIZ GIL, Inigo, inigo.uriz-gil@u-bordeaux.fr, VIA Inno, Pessac (Orateur)

DHUIT, Inès, ines.dhuit@u-bordeaux.fr, VIA Inno, Pessac (Orateur)

FLAMAND, Marina, marina.flamand@u-bordeaux.fr, VIA Inno, Pessac

Thématique : 4 – Analyse de données / 6 – Intelligence Artificielle

Résumé : *Ce travail expose le développement d'une nouvelle méthode d'analyse visant à optimiser l'exploitation de données issues de la presse dans une démarche de compréhension des tendances d'innovation et du positionnement technologique des acteurs. La méthode d'analyse ambitionne deux phases impliquant chacune des méthodes d'intelligence artificielle dans une volonté de classification et détection des domaines des articles de presse en vue de leur exploitation.*

Mots clés : intelligence technologique ; intelligence artificielle ; analyse de données ; données de presse ; classification ; reconnaissance d'entités nommées

1. Introduction

L'intelligence technologique est une pratique visant à soutenir et à améliorer la prise de décision des acteurs de la recherche et l'innovation en leur fournissant des informations scientifiques et technologiques qui découlent de leurs problématiques. De multiples méthodes co-existent dans l'optique de fournir une image la plus représentative des tendances et positionnements d'acteur en matière d'innovation. Différentes sources de données peuvent être mobilisées à cet égard, les plus connues étant le brevet et la publication scientifique.

L'innovation ne se limitant pas aux seuls développements scientifiques et technologiques, une vision plus large des environnements innovants implique de se ressourcer auprès d'autres sources de données qui renseigneraient à la fois sur les aspects de R&D des acteurs mais aussi du développement de nouveaux services ou produits, l'évolution de la réglementation, les relations entre les acteurs, etc. Dans cet optique, les articles de presse s'illustrent comme étant une donnée particulièrement intéressante à traiter afin de compléter la vision attendue de son environnement S&T.

Le travail envisage d'exposer l'exploitation de cette source de données, les articles de presse, dont la pertinence est validée à bien des égards :

- Une donnée riche en informations de natures différenciées (positionnement technologique, relation entre les acteurs, marché, produit, etc.),
- Des informations en temps réel,
- La possibilité d'identifier des acteurs non présents dans les bases de données traditionnellement utilisées en intelligence technologique (brevets, publications).

Bien qu'intéressante, l'exploitation des articles de presse reste limitée et nécessite d'être optimisée :

- Une donnée non structurée,
- Des articles de natures très différentes > nécessité de les distinguer,
- Des documents non catégorisés > compréhension des thématiques scientifiques et technologiques difficile.

A cet effet, la plateforme VIA Inno a développé une méthode visant à optimiser l'exploitation des données d'articles de presse par la mobilisation conjointe de méthodes automatisées de classification et de détection de domaines. Basée sur une expertise développée par la plateforme VIA Inno sur des méthodes d'intelligence technologique couplée à des modèles d'intelligence artificielle, une méthodologie structurée a été mise en œuvre dans le but de réaliser une analyse pertinente à partir de l'exploitation des articles de presse.

2. Méthodologie

La méthodologie proposée se structure en 6 étapes :

- Définition des catégories des articles de presse.
- Mise en place d'un processus ETL (*Extract, transform and load*) pour l'obtention des données depuis la source (*Factiva*) et mise en forme **de ceux-ci**.
- Détection des catégories des articles de presse selon deux approches différentes :
 - Mots-clés : à l'aide de listes de mots-clés qui permettent d'identifier chacune des catégories.
 - Réseaux de neurones : à l'aide d'un dataset d'entraînement qui aura été créé manuellement par les experts, définition du réseau et entraînement de celui-ci.
- Obtention d'informations en lien avec les catégories en utilisant des modèles d'IA de type question/answering avec de listes de questions définies au préalable pour mieux identifier les détails concernant chaque article.
- Identification des acteurs mentionnés dans les articles en utilisant des modèles d'IA de type NER (*Named Entity Recognition*).
- Croisement des informations obtenues pour la réalisation de l'analyse.

3. Originalité / perspective

Cette approche est originale dans le fait qu'elle implique :

- Le développement d'une méthode d'analyse de données non structurées qui mobilise différents modèles d'intelligence artificielle,
- L'exploitation quantitative d'une donnée non structurée : les données d'articles de presse (issues de la base de données Factiva).

Prochaines étapes associées au développement des travaux :

- Répliquer la méthodologie sur des données de presse en libre accès.
- Adapter la méthodologie aux données de publications scientifiques afin d'en optimiser l'exploitation dans une démarche de compréhension des activités d'innovation.



technology intelligence

VIA Inno : est un projet universitaire créé en 2009 et situé au sein du BSE, *Bordeaux Science Économique* de l'université de Bordeaux.

Labellisée *centre d'innovation sociétale* en 2012 dans le cadre de l'IDEX de Bordeaux, VIA Inno obtient en 2019 le label *plateforme de recherche de l'université de Bordeaux* dédié à l'**intelligence technologique**.

Fondé dans la recherche publique, les missions de VIA Inno sont tournées vers la société civile. **La missions du projet VIA Inno est de diffuser la pratique de l'intelligence technologique pour les acteurs français de la recherche et de l'innovation.**

Pour réaliser sa mission, la plateforme se structure autour de 3 axes d'activité :

- *La formation et le transfert* : Transférer nos méthodes, notre vision et notre savoir-faire aux utilisateurs de l'intelligence technologique.
- *La recherche* : Participer au renouvellement des approches et s'inscrire dans la communauté de recherche sur le domaine de l'intelligence technologique.
- *La sensibilisation aux décideurs par la démonstration* : Diffuser les bénéfices de l'intelligence technologique en répondant aux besoins opérationnels des organisations.

Présentation des orateurs :



Inigo Uriz-Gil est titulaire d'une double formation, un diplôme d'ingénieur en informatique et d'un master II en économie de l'innovation. Il est actuellement analyste au sein de la plateforme VIA Inno.



Inès Dhuit est titulaire d'un master II en économie de l'innovation. Elle est actuellement analyste sénior au sein de la plateforme VIA Inno et responsable de la formation.