



**HAL**  
open science

## **Visualisation de données issues des réseaux sociaux : une plateforme de type Business Intelligence**

Maxime Masson, Christian Sallaberry, Marie-Noëlle Bessagnet, Annig Le Parc-Lacayrelle, Philippe Roose, Rodrigo Agerri

### ► **To cite this version:**

Maxime Masson, Christian Sallaberry, Marie-Noëlle Bessagnet, Annig Le Parc-Lacayrelle, Philippe Roose, et al.. Visualisation de données issues des réseaux sociaux : une plateforme de type Business Intelligence. M@ppemonde, 2024, 138, <10.4000/12jjl>. <hal-04747034>

**HAL Id: hal-04747034**

**<https://hal.science/hal-04747034v1>**

Submitted on 21 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

## Mappemonde

Revue trimestrielle sur l'image géographique et les formes du territoire

138 | 2024

Varia

---

### Visualisation de données issues des réseaux sociaux : une plateforme de type *Business Intelligence*

*Data visualization from social media: a Business Intelligence-like platform*

*Visualización de datos provenientes de redes sociales: una plataforma de tipo Business Intelligence*

**Maxime Masson, Christian Sallaberry, Marie-Noëlle Bessagnet, Annig Le Parc Lacayrelle, Philippe Roose et Rodrigo Agerri**

---



#### Édition électronique

URL : <https://journals.openedition.org/mappemonde/9780>

ISSN : 1769-7298

#### Éditeur

UMR ESPACE

#### Référence électronique

Maxime Masson, Christian Sallaberry, Marie-Noëlle Bessagnet, Annig Le Parc Lacayrelle, Philippe Roose et Rodrigo Agerri, « Visualisation de données issues des réseaux sociaux : une plateforme de type *Business Intelligence* », *Mappemonde* [En ligne], 138 | 2024, mis en ligne le 21 octobre 2024, consulté le 21 octobre 2024. URL : <http://journals.openedition.org/mappemonde/9780>

---

Ce document a été généré automatiquement le 21 octobre 2024.



Le texte seul est utilisable sous licence CC BY-NC-SA 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.

---

# Visualisation de données issues des réseaux sociaux : une plateforme de type *Business Intelligence*

*Data visualization from social media: a Business Intelligence-like platform*

*Visualización de datos provenientes de redes sociales: una plataforma de tipo Business Intelligence*

**Maxime Masson, Christian Sallaberry, Marie-Noëlle Bessagnet, Annig Le Parc Lacayrelle, Philippe Roose et Rodrigo Agerri**

---

## Introduction

- 1 Dans le contexte actuel, dominé par le big data et la croissance exponentielle des sources de données, la capacité à traiter, analyser et synthétiser efficacement d'importantes quantités d'informations est cruciale (Leung *et al.*, 2013). Ce besoin est accentué par les flux constants de contenus générés par les utilisateurs (UGC) disponibles en ligne, notamment issus des plateformes de réseaux sociaux, sources prolifiques de données textuelles. Le rôle du traitement automatique du langage naturel (TALN) est ici prépondérant, permettant la transformation de vastes volumes de données textuelles non structurées en des ensembles d'informations structurées (Souili *et al.*, 2015) via des annotations. Toutefois, face à l'ampleur des données disponibles sur les plateformes sociales et à la multiplicité des annotations possibles, l'extraction d'informations pertinentes de ces corpus devient une tâche ardue. De ce fait, il est impératif de disposer d'outils adaptés permettant une visualisation et une interprétation optimisées, spécifiquement pour des utilisateurs tels que les décideurs d'un domaine donné (par exemple : le tourisme, les politiques publiques, etc.).
- 2 Cet article présente TextBI, notre plateforme multimodale et générique qui permet de visualiser, de manière simple ou agrégée, les annotations multidimensionnelles présentes dans des corpus extraits des réseaux sociaux. Cette dernière est multisource

et indépendante du domaine. Ainsi, TextBI repose sur un modèle de données (Masson *et al.*, 2023) capable de représenter n'importe quel corpus annoté selon nos dimensions d'analyse. TextBI peut également fonctionner avec tout domaine préalablement modélisé sous la forme d'une ressource sémantique hiérarchisée (dictionnaire, thésaurus, ontologie, etc.).

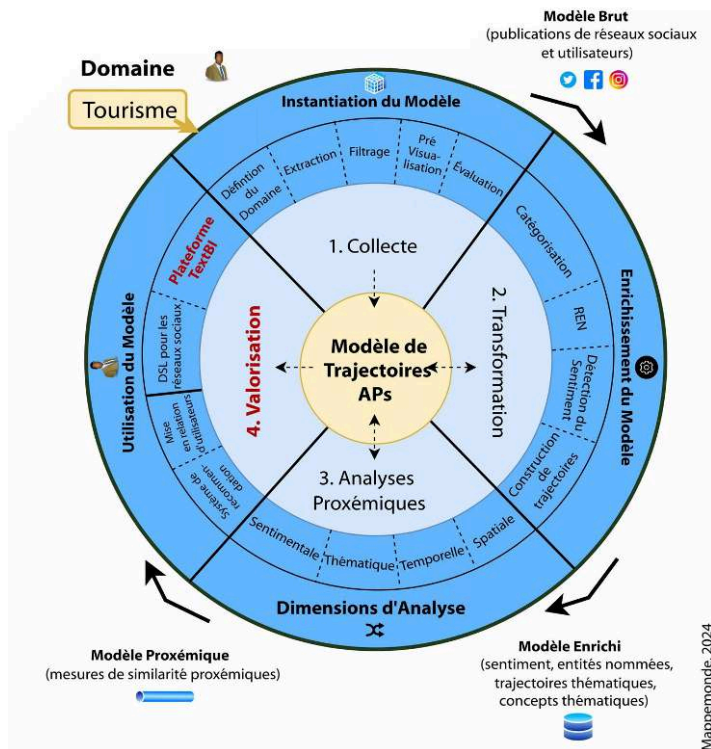
- 3 La plateforme TextBI prend la forme d'un tableau de bord reposant sur quatre dimensions centrales d'analyse : la dimension spatiale (données de localisation présentes dans le corpus), temporelle (horodatages et périodes temporelles), thématique (phénomènes décrits dans une ressource sémantique) et personnelle (données et profils associés à l'utilisateur). TextBI prend également en charge des données contextuelles d'enrichissement, telles que les informations liées au sentiment et à l'engagement (par exemple, *likes* et partages). Offrant une interface conviviale et intuitive, il intègre des fonctionnalités typiques des logiciels de *Business Intelligence* (BI), telles que le filtrage multidimensionnel combiné<sup>1</sup> et la synchronisation des visuels. Les annotations multidimensionnelles brutes sont transformées en visuels interactifs, rendant ainsi leur compréhension plus aisée et facilitant le processus d'interprétation pour les décideurs. Il est important de noter que la collecte des données à partir des réseaux sociaux et la création des annotations (via des processus de traitement automatique du langage naturel - TALN) doivent être réalisées avant l'utilisation de la plateforme pour produire un corpus annoté de publications. Dans cet article, nous nous concentrons principalement sur la composante visualisation.
- 4 Comme mentionné précédemment, cette plateforme s'adresse aux spécialistes d'un domaine d'activité donné en situation de recherche d'informations. Par exemple, dans le secteur du tourisme, les offices de tourisme pourraient vouloir identifier les principales activités touristiques dans une région d'intérêt, les villes fréquemment associées et les sentiments moyens exprimés par les visiteurs. Ces informations peuvent les assister dans la prise de décisions éclairées. Nous faisons l'hypothèse que des spécialistes de domaines d'activité différents peuvent avoir des besoins similaires comme, par exemple, identifier les fréquences, les associations et les séquences d'annotations dans des corpus de réseaux sociaux.
- 5 Cet article est organisé de la manière suivante : nous débutons en évoquant le contexte du projet et nos motivations. Ensuite, nous abordons les catégories de plateformes existantes en mettant l'accent sur leurs atouts et leurs faiblesses relatifs à ce contexte. Puis, une section détaillée présente les principales caractéristiques du tableau de bord. Ce dernier est indépendant du domaine d'application, mais pour nos expérimentations nous avons pris en exemple le domaine du tourisme. Enfin, nous concluons par une discussion sur les limites et les perspectives de notre plateforme.

## Contexte et motivations

- 6 Ces travaux de recherche sont mis à l'épreuve dans le cadre du projet APs, un acronyme pour « *Augmented Proxemic Services* ». Il s'agit d'un projet transfrontalier et multilingue, traitant de textes en anglais, français et espagnol. Il se focalise sur l'application de la théorie de la proxémique (Hall, 1966) aux données des réseaux sociaux pour répondre à divers besoins liés au tourisme, aux flux de visiteurs et à l'attention portée sur le patrimoine culturel dans la région du Pays basque, une région hautement touristique qui s'étend à la fois en France et en Espagne.

- 7 Les réseaux sociaux offrent une opportunité inégalée pour les acteurs du tourisme (par exemple, offices de tourisme, conseils municipaux, etc.) de se détourner des bases de données commerciales coûteuses, comme celles fournies par les entreprises de télécommunications ou les organisations de marketing de destination (DMOs). Les avantages de l'utilisation de données issues des réseaux sociaux incluent leur coût abordable, les mises à jour en temps réel, un volume massif et une large portée couvrant divers aspects liés au tourisme tels que les activités, l'hospitalité et le transport. Au-delà du tourisme, les données des réseaux sociaux peuvent être utilisées pour analyser un large éventail d'autres domaines (par exemple : la mode, les politiques publiques, etc.). Cependant, l'analyse des données des réseaux sociaux présente plusieurs défis. En effet, les réseaux sociaux contiennent des textes courts, non structurés et informels, souvent parsemés d'erreurs grammaticales. De plus, le grand volume de données disponibles peut rendre difficile l'extraction d'informations pertinentes et nécessite une puissance de calcul importante.
- 8 Ainsi, pour aider les professionnels à analyser les données des réseaux sociaux, nous avons proposé un cadre de travail générique et multilingue, appelé le Framework APs (Masson *et al.*, 2023). Il est composé de quatre modules et, en entrée, accepte la description sémantique d'un domaine d'intérêt (ici le tourisme, mais cela pourrait être n'importe quel autre domaine) sous la forme, par exemple, d'un dictionnaire, thésaurus ou encore d'une ontologie. Le cycle de vie du Framework APs est illustré dans la **figure 1**, il est inspiré du concept de *WaterWheel* (Bucher *et al.*, 2021). Dans cet article, nous nous concentrerons sur la phase 4 et nous n'irons donc pas dans le détail des phases précédentes.
- 9 Collecte (1) : Cette étape extrait des données des réseaux sociaux pour créer un corpus pertinent de publications, médias et utilisateurs. L'approche utilisée est générique et affinée à travers un processus itératif pour garantir la pertinence des données. Pour plus de détail sur cette méthodologie, voir Maxime *et al.* (2022).
- 10 Transformation (2) : L'objectif ici, est de convertir les données brutes, principalement des textes non structurés, en informations structurées (annotations). Cela se fait en utilisant divers outils de traitement automatique du langage naturel (TALN) pour extraire des entités, les annoter et aligner ces annotations sur des ressources sémantiques (dictionnaire, thésaurus, ontologie) ou géographiques (OpenStreetMap, IGN). Les données sont ensuite enrichies avec des dimensions comme le sentiment ou l'engagement.
- 11 Analyses Proxémiques (3) : basée sur la théorie de la proxémique, cette étape adapte les concepts traditionnellement physiques de proxémique aux interactions virtuelles dans les réseaux sociaux. La proxémique (Hall, 1966) est la science qui étudie l'impact de l'espace et l'effet des distances sur les comportements. Cette phase introduit des mesures spécifiques, telles que la similarité d'activités menées par différents acteurs, ou encore la similarité d'activités pratiquées à une période donnée de la journée, pour dégager des indicateurs de comportements. Dans cet article, nous n'entrerons pas dans le détail des calculs de similarité, mais seulement dans la manière dont les résultats sont visualisés (voir module « Valorisation »).

Figure 1. Cycle de vie du Framework APs



La plateforme TextBI se positionne dans la phase de valorisation. Cette représentation circulaire est inspirée de Bucher *et al.*, 2021.

- 12 Valorisation (4) : Ici, les données annotées et les indicateurs sont présentés de manière accessible, notamment via des tableaux de bord. Ils sont conçus pour être interprétés et analysés par des utilisateurs finaux, dans notre cas d'expérimentation, les acteurs du tourisme. Ces mesures pourraient aussi alimenter des systèmes intelligents, par exemple pour générer des recommandations personnalisées ou mettre en lien des individus. C'est dans ce module que se positionne la plateforme TextBI.

## Travaux connexes

- 13 TextBI représente une avancée significative dans le domaine de la visualisation des annotations TALN et des indicateurs proximiques en intégrant et combinant les fonctionnalités de diverses catégories d'outils.
- 14 Les Systèmes d'information géographique (SIG) tels que QGIS (Menke *et al.*, 2016) et ArcGIS (Booth *et al.*, 2001) offrent une grande variété de fonctionnalités pour répondre aux différents besoins des utilisateurs géographes. Bien que ces systèmes soient utiles pour visualiser en profondeur les données géospatiales, leur principal intérêt réside dans l'aspect spatial des données, ce qui limite leur utilité dans des contextes de données non spatiales. Bien que les outils SIG soient capables d'afficher des données thématiques (Murthy *et al.*, 2003), elles doivent être associées à une zone spatiale spécifique.
- 15 Dans le domaine des outils de *Business Intelligence* (BI), Tableau Public (Datig et Whiting, 2018), Power BI (Ferrari et Russo, 2016) et QlikView (Shukla et Dhir, 2016) sont

reconnus pour leur capacité à renforcer les processus de prise de décision (Hansoti, 2010 ; Orlovskiy et Kopp, 2020) grâce à leur exploration interactive des données et leurs tableaux de bord conviviaux. Cependant, ces outils sont principalement conçus pour gérer des données numériques bien structurées, ce qui pose des problèmes lorsqu'il s'agit de travailler avec des données nativement textuelles. Bien que certains efforts aient été faits pour intégrer des processus de TALN dans les outils BI (Vashisht et Dharia, 2020 ; Desai *et al.*, 2021), ils manquent d'efficacité pour présenter des données séquentielles (par exemple, des trajectoires) et pour établir des connexions entre les annotations textuelles de différentes dimensions. De plus, ces outils manquent de support complet pour les données multilingues.

- 16 Concernant les outils de traitement automatique de la langue (TALN), ils peuvent être divisés en deux principales catégories. Certains d'entre eux, comme SpaCy (Chantrapornchai et Tunsakul, 2021), TextRazor (Rajaonarivo *et al.*, 2022), GATE (Maynard *et al.*, 2000), et Gensim (Rehurek et Sojka, 2011), se concentrent principalement sur le traitement des données, offrant des capacités de visualisation limitées comme les nuages de mots, les graphes sémantiques et les vues d'annotations basées sur le texte. D'autres outils, tels que IRaMuTeQ (Loubère et Ratinaud, 2014), Voyant (Welsh, 2014), VOSviewer (Van Eck et Waltman, 2013) et SentimentViz (Healey et Ramaswamy, 2022), offrent une gamme plus large d'options de visualisation, mais se concentrent généralement sur une seule dimension (comme le sentiment) ou les analyses statistiques basées sur les mots. Cependant, certains de ces outils peuvent être très complexes pour les non-informaticiens.
- 17 Notre proposition combine des fonctionnalités des outils BI, SIG et TALN (voir **tableau 1**) au service de la visualisation de corpus annotés issus de réseaux sociaux. Des outils BI, nous empruntons le design interactif, les interfaces conviviales et la synchronisation visuelle, en adaptant ces fonctionnalités pour leur utilisation avec des données textuelles annotées plutôt qu'avec des données numériques traditionnelles.

Tableau 1. Comparaison des points forts et faiblesses entre plusieurs catégories d'outils ayant une composante visualisation

|                         | Systèmes d'information géographique (SIG)                    | Outils de Business Intelligence (BI)  | Outils orientés traitement de texte (TALN)  |
|-------------------------|--|---|---|
| <b>Exemple</b>          | QGIS, ArcGIS   | Tableau Public, PowerBI, QlikView   | SpaCy, GATE, TextRazor, IRaMuTeQ, VOSViewer, SentimentViz   |
| <b>Points forts (+)</b> | Analyse spatiale poussée<br>Granularités spatiales multiples | Exploration de données multidimensionnelles<br>Simple d'utilisation<br>Interactivité inter et intra visuel<br>Filtrages combinés<br>Synchronisation des visuels | Statistiques avancées (cooccurrence, fréquences, etc.)<br>Gestion native du texte<br>Visuels plus adaptés au texte : graphes sémantiques, nuages de mot, etc. |

|                       |   |   |  |
|-----------------------|---|---|--|
| <b>Faiblesses (-)</b> | Difficile de visualiser des données sans composante spatiale<br>Peuvent être complexes pour des utilisateurs néophytes (non-géographes) | Nécessite des données structurées (e.g., base de données relationnelle)<br>Principalement pour des données numériques<br>Difficultés à visualiser des données séquentielles | Complexe, vise principalement des utilisateurs spécialistes du TALN<br>Analyse au niveau des mots ou focus sur une seule dimension (ex. : sentiment) |
|-----------------------|---|---|--|

- 18 Nous adoptons les vues spatiales détaillées des outils SIG, reconnaissant que les données des réseaux sociaux contiennent souvent un aspect spatial. Cependant, nous visons une approche intégralement multidimensionnelle. Des outils TALN, nous prenons les aspects analytiques, comme l'analyse de cooccurrence et de fréquence, mais nous allons au-delà pour intégrer, voire agréger, des entités comme des concepts thématiques, des lieux ou encore des entités temporelles.

## La plateforme TextBI

- 19 La plateforme TextBI prend la forme d'un tableau de bord permettant une visualisation interactive des annotations multidimensionnelles sur des corpus issus des réseaux sociaux. Elle est conçue pour être adaptable à n'importe quel réseau social et domaine d'application, tant que les données sont conformes à notre modèle (ce dernier est décrit dans Masson *et al.*, 2023). La plateforme offre une variété de fonctionnalités pour faciliter l'analyse de données multilingues issues des réseaux sociaux selon quatre dimensions : spatiale, temporelle, thématique et personnelle. Dans notre premier contexte expérimental, le domaine est le tourisme, ce qui implique que les parties prenantes potentielles comprennent des offices de tourisme, mairies ou organisations de marketing des destinations (Gretzel *et al.*, 2006). L'objectif est de leur permettre de manipuler directement l'interface TextBI.
- 20 Dans les exemples suivants, nous utiliserons un corpus d'environ 3 100 publications extraites du réseau social Twitter (aujourd'hui X). Ces publications ciblent la zone de la Côte Basque française durant l'été 2019 et proviennent de touristes (locaux et externes). Pour plus de détails sur la collecte et la nature du corpus de données, nous invitons le lecteur à se référer à Masson *et al.*, 2022. Nous définissons le domaine du tourisme à l'aide du *Thésaurus du Tourisme et des Loisirs*<sup>2</sup> (Organisation mondiale du tourisme, 2002), une terminologie multilingue (français, anglais et espagnol) couvrant approximativement 2 500 concepts liés au tourisme. Notre modèle de données a été instancié en utilisant deux types de données : les métadonnées et le contenu des tweets. L'instanciation basée sur les métadonnées est un processus direct, englobant des éléments tels que les mesures d'engagement, les caractéristiques du profil, l'horodatage et les géotags associés aux tweets, entre autres. Le contenu des tweets a été annoté à l'aide de trois modules de traitement automatique du langage naturel (TALN) : (1) détection des entités nommées spatiales (toponymes) en lien avec *OpenStreetMap* (2) détection du sentiment global des tweets (positif, négatif ou neutre) à l'aide du modèle de langage XLM-ROBERTa (Barbieri *et al.*, 2022) et (3) détection des concepts thématiques liés au domaine du tourisme, en se basant sur le vocabulaire hiérarchisé

du thésaurus de l'Organisation mondiale du tourisme. La **figure 2** présente une vue d'ensemble de la plateforme TextBI dont nous allons détailler chacune des parties.

### Mode « Fréquence »

- 21 Le mode « Fréquence » (**figure 2**, « Mode Fréquence ») sert d'interface principale à la plateforme. Il met en évidence les fréquences spatiales, temporelles et thématiques dans les publications ainsi que la distribution des annotations à travers quatre visualisations majeures.
- 22 La carte thématique (ou *treemap*, **figure 2** « Carte thématique ») visualise la structure hiérarchique et la fréquence des concepts thématiques à partir d'une ressource sémantique, comme un dictionnaire, un thésaurus ou une ontologie. Par exemple, en cartographiant les thèmes extraits des tweets avec le *Thésaurus du Tourisme et des Loisirs* de l'OMT, nous observons que les concepts thématiques liés au patrimoine touristique sont les plus fréquents dans notre corpus, représentant environ 40 % des concepts découverts, dont beaucoup sont liés à des ressources naturelles telles que la côte et la mer.
- 23 La carte spatiale (**figure 2**, « Carte spatiale ») est une autre visualisation montrant la fréquence des lieux mentionnés dans les publications. Les utilisateurs peuvent définir la granularité spatiale, qui peut varier de catégories larges comme les pays à des catégories plus spécifiques comme les Points d'Intérêt (POIs). Les lieux sont regroupés en fonction de la granularité choisie. La carte utilise un gradient linéaire pour la représentation, avec des zones plus ou moins opaques selon la concentration des publications. Ici, nous nous concentrons sur la région de la Côte Basque française au niveau de la granularité « commune » et observons un point chaud de tweets sur 3 villes (Bayonne, Anglet et Biarritz) situées au nord de la région. Ces villes proches semblent être populaires parmi les visiteurs à cette période de l'année.
- 24 Ensuite, la carte des utilisateurs (**figure 2**, « Carte des utilisateurs ») est un nuage de points qui présente la fréquence de publication des utilisateurs sur l'axe x par rapport au nombre de followers des utilisateurs sur l'axe y. Cette visualisation aide à l'identification rapide des utilisateurs influents. La langue de chaque utilisateur est représentée par une couleur et un symbole, la taille du symbole correspondant au nombre de publications de cet utilisateur. 655 utilisateurs sont représentés dans cet exemple, couvrant plus de 6 langues (français, espagnol, anglais, basque, italien et indéterminé).
- 25 Enfin, la frise chronologique (**figure 2**, « Frise chronologique ») offre une visualisation du volume de publications par jour sur la plage temporelle du corpus de données. Chaque jour est subdivisé en différentes périodes telles que le matin (bleu), l'après-midi (orange) ou le soir (violet). Cette visualisation offre différentes options de granularité temporelle, dont les jours, les semaines, les mois, les saisons et les années. Ici, nous utilisons la granularité jour pour l'été 2019. Nous observons un pic de tweets entre le 24 et le 28 juillet 2019.

### Modes « Association » et « Mouvement »

- 26 Dans le mode « Association » (**figure 2**, « Mode Association »), le tableau de bord présente des représentations visuelles illustrant les connexions entre les entités en se

basant sur leurs cooccurrences dans les publications. Ces connexions sont dépeintes à l'aide de graphes non dirigés, où les nœuds représentent des entités telles que des concepts thématiques ou des lieux, et les arêtes indiquent la force de cooccurrence entre eux. Cela permet d'identifier facilement les concepts ou lieux fortement corrélés. Comme on pouvait s'y attendre, « soleil », « plage », « surf », « mer » et autres concepts côtiers sont fortement liés.

- 27 Ce mode peut également se transformer pour afficher les mouvements (**figure 2**, « Mouvements thématiques » et « Mouvements spatiaux »), se concentrant sur le séquençement des entités dans les trajectoires spatio-thématiques des utilisateurs, par exemple, la transition d'un concept thématique à un autre. Ces séquences sont visualisées à travers des graphes dirigés : l'épaisseur des arêtes y est proportionnelle au nombre de séquences extraites des trajectoires.

### Mode « Proxémique »

- 28 Le mode « Proxémique » de TextBI (**figure 2**, « Mode Proxémique ») offre aux utilisateurs la capacité d'analyser des corpus de données de réseaux sociaux via une approche originale de calcul d'indicateurs (Hall, 1966). Les utilisateurs peuvent alors sélectionner une entité comme référence (un utilisateur, une cohorte, un concept thématique, un lieu ou une période de temps) et observer la proximité (exprimée via une mesure de similarité proximique) d'autres entités par rapport à cette référence.
- 29 Par exemple, dans notre démonstration, nous avons sélectionné « Ciboure », une ville touristique, comme entité de référence, et l'avons comparée à des concepts thématiques. L'interface comprend un panneau pour la sélection des références (**figure 2**, « Listes d'entités »), un panneau central appelé « réticule proximique » pour visualiser les résultats (**figure 2**, « Réticule proximique »), et un panneau de paramètres (**figure 2**, « Options de calcul ») pour personnaliser les calculs de similarité proximique. Le système prend en charge de nombreuses combinaisons, telles qu'utilisateur-vers-thèmes ou lieu-vers-utilisateurs. Le manipulateur de l'interface TextBI peut personnaliser le mode de calcul de la similarité proximique, par exemple en attribuant un poids plus élevé aux tweets positifs ou ayant un engagement fort. L'interface comprend également un panneau de modèle (**figure 2**, « Modèles prédéfinis ») offrant des paramètres de distance prédéfinis pour des domaines d'étude spécifiques. Dans le domaine du tourisme, le modèle « Hébergement », par exemple, limite l'affichage aux concepts thématiques liés à l'hébergement.

### Superpositions et données contextuelles

- 30 Dans TextBI, les visuels prennent en charge des superpositions (“overlays”) pour présenter les données contextuelles de sentiment et d'engagement (**figure 2** « Sentiment » et « Engagement »).
- 31 Le sentiment est indiqué par un code couleur (vert pour positif, rouge pour négatif et orange pour un sentiment mixte) permettant une meilleure compréhension du sentiment global par thèmes, lieux, temps ou utilisateur. L'engagement (*likes*, citations, réponses, etc.) est visualisé à l'aide d'un gradient linéaire (bleu indiquant un fort engagement, blanc indiquant un faible engagement) offrant des informations sur l'engagement des utilisateurs. Nous constatons, dans cet exemple, que la plupart des

concepts touristiques ont tendance à être associés à un sentiment positif, mais certains, comme « transport » ou « écologie », sont plus nuancés.

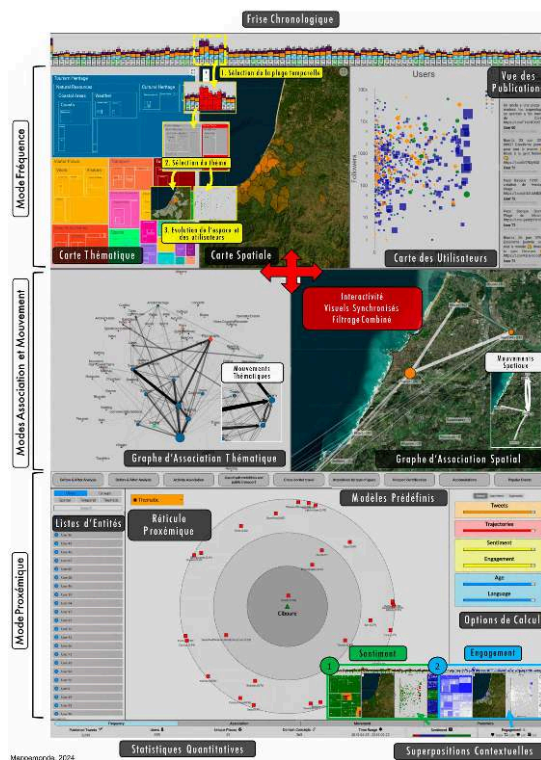
## Interactions et synchronisation visuelle

- 32 La plateforme TextBI utilise des interactions couramment exploitées dans les outils de *Business Intelligence* (BI), garantissant une synchronisation fluide des visualisations. Elle propose des combinaisons de filtrage multidimensionnelles telles que « spatial-temporel », « spatial-thématique » et « utilisateur-temporel ». Lorsque le manipulateur de l'interface sélectionne un lieu, thème, utilisateur ou une plage de temps particulière, toutes les autres visualisations s'ajustent pour afficher uniquement les publications et analyses associées au filtre choisi. Dans sa vue proxémique, le manipulateur de l'interface peut facilement glisser-déposer des références au centre du panneau de réticule.
- 33 Prenons l'exemple illustré dans la **figure 2**. Si l'on choisit la plage de temps du 24 au 27 juillet, la carte thématique se met à jour pour afficher une concentration plus élevée du concept thématique « célébration ». En cliquant davantage sur le concept « célébration », la carte spatiale met en évidence un point chaud dans la ville de « Bayonne ». Cela coïncide avec la période des fêtes de Bayonne, un événement local qui attire plus d'un million de participants.
- 34 Enfin, la plateforme dispose d'un panneau de statistiques (**figure 2** « Statistiques quantitatives ») montrant le volume de données associé aux filtres actifs, y compris les décomptes des publications, des utilisateurs, des concepts et des lieux, la plage temporelle actuelle des publications, le niveau total d'engagement, et les sentiments dominants. Les textes des publications filtrées sont affichés dans le panneau des publications, sur la droite (**figure 2**, « Vue des publications »).

## Aspects techniques

- 35 TextBI est une application web développée en HTML, CSS et JavaScript. Elle fonctionne uniquement côté client, elle ne nécessite donc ni serveur distant, ni *backend* et peut être exécutée localement. Le modèle de données a été mis en œuvre au format JSON.
- 36 TextBI sert d'outil d'affichage et d'agrégation de données, fournissant des analyses statistiques et calculant des indicateurs. Il ne s'engage dans aucune tâche de traitement de données proprement dite. La collecte et la transformation des données, y compris le traitement du langage naturel (TALN), doivent être effectuées en amont. À l'avenir, nous envisageons de le rendre facilement extensible sur d'autres dimensions.

Figure 2. Vue d'ensemble de la plateforme TextBI



Une démonstration en temps réel est disponible à l'adresse suivante : <https://maxime-masson.github.io/TextBI/>

## Conclusion

- 37 TextBI est un tableau de bord novateur conçu pour simplifier la visualisation d'indicateurs multidimensionnels issus de données des réseaux sociaux. TextBI est composé de plusieurs familles de visuels génériques, paramétrables en fonction du domaine spécifique d'application ciblé. Il s'adresse principalement aux professionnels non-techniques, experts dans leur domaine, tels que les offices de tourisme dans le cadre de notre étude sur le tourisme. TextBI est basé sur un modèle de données générique, lui permettant d'être aisément adaptable à différents domaines métiers. Il se focalise sur quatre dimensions principales : spatiale, temporelle, thématique et individuelle, tout en proposant une variété d'options de visualisation (fréquences, association, mouvement, proxémique, etc.). Dans le contexte des réseaux sociaux, il prend en charge des données d'enrichissement supplémentaires, telles que l'analyse des sentiments et le niveau d'engagement.
- 38 TextBI se distingue par sa capacité d'interactivité avancée, incluant des interactions telles que le filtrage croisé, la synchronisation visuelle, l'agrégation et bien d'autres. Pour l'avenir, nous envisageons d'enrichir TextBI avec de nouvelles fonctionnalités, notamment une interface utilisateur dédiée à la sélection de la granularité des données, ainsi que la capacité de traiter des données en temps réel en se connectant à *InfluxDB* (Ahmad et Ansari, 2017). Par exemple, nous souhaiterions expérimenter la plateforme pour suivre des événements importants en temps réel (événements sportifs, élections, etc.) sur les réseaux sociaux et, ainsi, obtenir des indicateurs évolutifs sur les tendances

et des opinions publiques. Après avoir validé sa capacité à être déployé dans divers domaines et son aptitude à gérer des corpus de données massifs, notre objectif est de rendre TextBI *open source* et accessible à tous, le mettant ainsi à la disposition de la communauté scientifique.

- 39 Ce projet a bénéficié d'un financement de la Communauté d'agglomération Pau Béarn Pyrénées.

---

## BIBLIOGRAPHIE

- AHMAD K., ANSARI M. (2017). "Hands-on InfluxDB". In DEKA G. C., *NoSQL: Database for Storage and Retrieval of Data in Cloud*, New-York : Chapman and Hall/CRC, p. 341-354.
- BARBIERI F., ANKE L. E., CAMACHO-COLLADOS J. (2022). "XLM-T: Multilingual language models in twitter for sentiment analysis and beyond". *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, p. 258-266.
- BOOTH B., MITCHELL A. (2001). *Getting started with ArcGIS*. Redlands : ESRI Press, 253 p.
- BUCHER B., HEIN C., RAINES D., GOUET BRUNET V. (2021). "Towards Culture-Aware Smart and Sustainable Cities: Integrating Historical Sources in Spatial Information Infrastructures". *ISPRS International Journal of Geo-Information*, vol. 10, n° 9, p. 588.
- CHANTRAPORNCHAI C., TUNSAKUL A. (2021). "Information extraction tasks based on BERT and Spacy on tourism domain". *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 15, n° 1, p. 108-122.
- DATIG I., WHITING P. (2018). "Telling your library story: tableau public for data visualization". *Library Hi Tech News*, vol. 35, n° 4, p. 6-8.
- DESAI Z., ANKLESARIA K., BALASUBRAMANIAM H. (2021). "Business intelligence visualization using deep learning based sentiment analysis on Amazon review data". *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, p. 1-7. IEEE.
- FERRARI A., RUSSO M. (2016). *Introducing Microsoft Power BI*. Redmond : Microsoft Press.
- GRETZEL U., FESENMAIER D.R., FORMICA S., O'LEARY J.T. (2006). "Searching for the future: Challenges faced by destination marketing organizations". *Journal of Travel Research*, vol. 45, n° 2, p. 116-126.
- HALL E. T. (1966). *The Hidden Dimension*. New York : Anchor Books, vol. 609, 217 p.
- HANSOTI B. (2010). *Business Intelligence Dashboard in Decision Making*. Purdue e-Pubs, Purdue University.
- HEALEY C. G., RAMASWAMY S. (2022). "Twitter Sentiment Visualization". [https://www.csc2.ncsu.edu/faculty/healey/tweet\\_viz/](https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/)
- LEUNG D., LAW R., VAN HOOF H., BUHALIS D. (2013). "Social media in tourism and hospitality: A literature review". *Journal of travel & tourism marketing*, vol. 30, n° 1-2, p. 3-22.

- LOUBÈRE L., RATINAUD P. (2014). *Documentation iramuteq 0.6 alpha 3 version 0.1*. Récupéré sur [http://www.iramuteq.org/documentation/fichiers/documentation\\_19\\_02\\_2014.pdf](http://www.iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf)
- MASSON M., ROOSE P., SALLABERRY C., AGERRI R., BESSAGNET M.N., LE PARC LACAYRELLE A. (2023). "APs: A proxemic framework for social media interactions modeling and analysis". In CRÉMILLEUX B., HESS S., NIJSSEN S., éd., *Advances in Intelligent Data Analysis XXI: 21st International Symposium on Intelligent Data Analysis, IDA 2023, Louvain-la-Neuve, Belgium, April 12-14, 2023, Proceedings*, Springer, p. 287-299.
- MASSON M., SALLABERRY C., AGERRI R., BESSAGNET M.N., ROOSE P., LE PARC LACAYRELLE A. (2022). "A domain-independent method for thematic dataset building from social media: the case of tourism on twitter". In CHBEIR R., HUANG H., SILVESTRI F., MANOLOPOULOS Y., ZHANG Y., éd., *Web Information Systems Engineering - WISE 2022*. Springer, "Lecture Notes in Computer Science", vol. 13 724, p. 11-20.
- MAYNARD D., CUNNINGHAM H., BONTCHEVA K., CATIZONE R., DEMETRIOU G., GAIZAUSKAS R., HAMZA O., HEPPLER M., HERRING P., MITCHELL B., OAKES M., PETERS W., SETZER A., STEVENSON M., TABLAN V., URSU C., WILKS Y. (2000). *A survey of uses of GATE*. Technical Report CS-00-06, Department of Computer Science, University of Sheffield.
- MENKE K., SMITH R. JR., PIRELLI L., VAN HOESEN J. (2016). *Mastering QGIS*. Birmingham : Packt Publishing Ltd, 486 p.
- MURTHY K.S. R., AMMINEDU E., VENKATESWARA RAO V. (2003). "Integration of thematic maps through GIS for identification of groundwater potential zones". *Journal of the Indian Society of Remote Sensing*, vol. 31, p. 197-210.
- ORGANISATION MONDIALE DU TOURISME. (2002). *Thésaurus du tourisme et des loisirs*, 803 p.
- ORLOVSKYI D., KOPP A. (2020). "A business intelligence dashboard design approach to improve data analytics and decision making". *CEUR Workshop Proceedings*, vol. 2833
- RAJAONARIVO L., MINE T., ARAKAWA Y. (2022). "Coupling of semantic and syntactic graphs generated via tweets to detect local events". *12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, IEEE, p. 128-133.
- REHUREK R., SOJKA P. (2011). *Gensim - statistical semantics in Python*. Poster.
- SHUKLA A., DHIR S. (2016). "Tools for data visualization in business intelligence: case study using the tool QlikView". *Information Systems Design and Intelligent Applications: Proceedings of Third International Conference INDIA 2016*, vol. 2, Springer, p. 319-326.
- SOUILI A., CAVALLUCCI D., ROUSSELOT F. (2015). "Natural language processing (NLP). A solution for knowledge extraction from patent unstructured data". *Procedia engineering*, vol. 131, p. 635-643.
- VAN ECK N. J., WALTMAN L. (2013). *VOSViewer manual*. Univeristeit Leiden, 53 p.
- VASHISHT V., DHARIA P. (2020). "Integrating chatbot application with qlik sense business intelligence (BI) tool using natural language processing (NLP)". *Micro-Electronics and Telecommunication Engineering: Proceedings of 3rd ICMETE 2019*, Springer.
- WELSH M. E. (2014). "Review of voyant tools". *Collaborative Librarianship*, vol. 6, n° 2, p. 96-98.

## NOTES

1. Le filtrage multidimensionnel combiné permet d'appliquer plusieurs filtres sur différentes dimensions de données en même temps. Ces filtres s'exécutent simultanément pour restreindre l'analyse aux données pertinentes selon les critères choisis.

2. <https://www.unwto.org/fr>

---

## RÉSUMÉS

TextBI est un tableau de bord interactif destiné à visualiser des indicateurs multidimensionnels sur de grandes quantités de données multilingues issues des réseaux sociaux. Il cible quatre dimensions principales d'analyse : spatiale, temporelle, thématique et personnelle, tout en intégrant des données contextuelles comme le sentiment et l'engagement. Offrant plusieurs modes de visualisation, cet outil s'insère dans un cadre plus large visant à guider les diverses étapes de traitement de données des réseaux sociaux. Bien qu'il soit riche en fonctionnalités, il est conçu pour être intuitif, même pour des utilisateurs non informaticiens. Son application a été testée dans le domaine du tourisme en utilisant des données de Twitter (aujourd'hui X), mais il a été conçu pour être générique et adaptable à de multiples domaines. Une vidéo de démonstration est accessible au lien suivant : <https://maxime-masson.github.io/TextBI/>

TextBI is an interactive dashboard designed to visualize multidimensional indicators over large volumes of multilingual data from social media. It targets four main dimensions of analysis: spatial, temporal, thematic, and personal, while integrating contextual data such as sentiment and engagement. Offering multiple visualization modes, this tool fits into a broader framework aimed at guiding various stages of social network data processing. Although feature-rich, it is designed to be intuitive, even for non-computer literate users. Its application has been tested in the field of tourism using Twitter data (today known as X) but it is designed to be generic and adaptable to multiple domains. A demonstration video can be accessed at the following link: <https://maxime-masson.github.io/TextBI/>

TextBI es un cuadro de mandos interactivo diseñado para visualizar indicadores multidimensionales a partir de grandes volúmenes de datos multilingües procedentes de redes sociales. Considera cuatro dimensiones de análisis: espacial, temporal, temática y personal. Además, integra datos contextuales como el sentimiento y la participación. Mediante diferentes modos de visualización, esta herramienta facilita y guía al usuario en las diversas etapas del tratamiento de los datos de las redes sociales. A pesar de sus amplias funcionalidades, está diseñado para ser intuitivo, incluso para usuarios sin conocimientos informáticos. Su aplicación ha sido testeada en el campo del turismo utilizando datos de Twitter (hoy conocido como X) pero está diseñado para ser genérico y adaptable a múltiples dominios. Se puede acceder a una demostración en el siguiente enlace: <https://maxime-masson.github.io/TextBI/>

## INDEX

**Mots-clés** : visualisation, réseaux sociaux, tourisme, multidimensionnel, dimension spatiale, dimension temporelle, dimension thématique, dimension personnelle

**Palabras claves** : visualización, redes sociales, turismo, multidimensional, dimensión espacial, dimensión temporal, dimensión temática, dimensión personal

**Keywords** : visualization, social networks, tourism, multidimensional, dimension spatial, dimension temporal, dimension thematic, dimension personal

## AUTEURS

### **MAXIME MASSON**

LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA), Pau, France et Centre HiTZ - Ixa, Université du Pays basque UPV/EHU, Donostia-San Sebastian, Espagne

### **CHRISTIAN SALLABERRY**

LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA), Pau, France

### **MARIE-NOËLLE BESSAGNET**

LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA), Pau, France

### **ANNIG LE PARC LACAYRELLE**

LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA), Pau, France

### **PHILIPPE ROOSE**

LIUPPA, E2S, Université de Pau et des Pays de l'Adour (UPPA), Pau, France

### **RODRIGO AGERRI**

Centre HiTZ - Ixa, Université du Pays basque UPV/EHU, Donostia-San Sebastian, Espagne