TextBI: An Interactive Dashboard for Visualizing Multidimensional NLP Annotations in Social Media Data

Maxime Masson[1,2], Christian Sallaberry[1], Marie-Noelle Bessagnet[1], Annig Le Parc Lacayrelle[1], Philippe Roose[1], and Rodrigo Agerri[2]

[1]LIUPPA, E2S, University of Pau and Pays Adour (UPPA), Pau, France
[2]HiTZ Center - Ixa, University of the Basque Country UPV/EHU

Abstract

In this paper we introduce *TextBI*, a multimodal generic dashboard designed to present multidimensional text annotations on large volumes of multilingual social media data. This tool focuses on four core dimensions: spatial, temporal, thematic, and personal, and also supports additional enrichment data such as sentiment and engagement. Multiple visualization modes are offered, including frequency, movement, and association. This dashboard addresses the challenge of facilitating the interpretation of NLP annotations by visualizing them in a user-friendly, interactive interface catering to two categories of users: (1) domain stakeholders and (2) NLP researchers. We conducted experiments within the domain of tourism leveraging data from X (formerly Twitter) and incorporating requirements from tourism offices. Our approach, *TextBI*, represents a significant advancement in the field of visualizing NLP annotations by integrating and blending features from a variety of Business Intelligence, Geographical Information Systems and NLP tools. A demonstration video is also provided.[1]

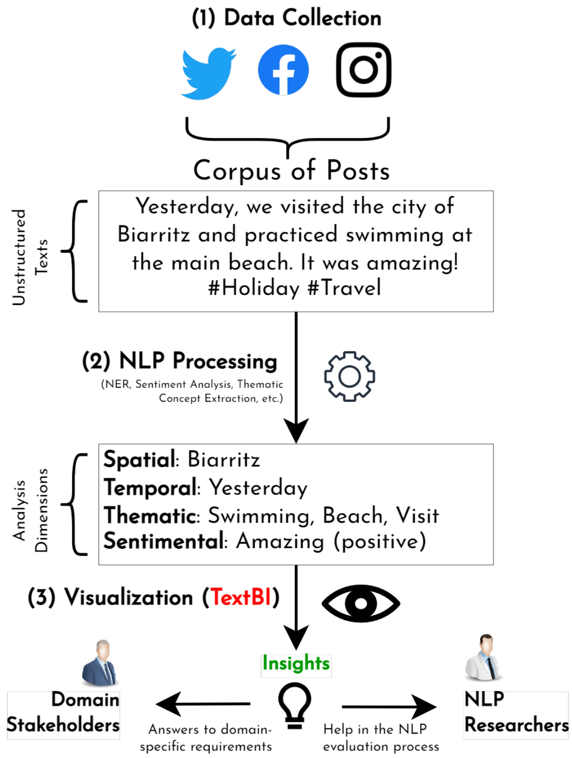

Figure 1: TextBI's role in the interaction with the NLP pipeline

1 Introduction

In today's data-driven era, the ability to quickly analyze, interpret, and make decisions based on vast amounts of data has become crucial (Leung et al., 2013). This is particularly true for social media platforms, which generate extensive amounts of textual data. Natural Language Processing (NLP) plays a crucial role in transforming unstructured text data, like social media posts, into structured knowledge (Souili et al., 2015). However, the enormous quantity of social media information and the wide range of potential automatic annotations can make it challenging to efficiently extract insights from annotated social media data. Consequently, there is a need for tools that can facilitate the interpretation and understanding of automatic annotations within a social media corpus for people who are not necessarily domain-experts. This paper introduces *TextBI*, a generic, multi-modal dashboard that enables comprehensive visualizations of multidimensional annotations extracted from social media (see Fig. 1).

TextBI offers a user-friendly and intuitive interface that presents text annotations visually. *TextBI* goes further than existing NLP visualization tools by incorporating interactions, multidimensional combined filtering, and visual synchronization, commonly found in Business Intelligence (BI) tools. Raw multilingual text annotations are converted into interactive visuals, making them more

[1]https://youtu.be/x714RKvo9Cg

Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics System Demonstrations, pages 1–9
March 17-22, 2024

easily understandable thereby facilitating their interpretation. This dashboard operates on four generic dimensions: spatial (*location data*), temporal (*time stamps*), thematic (*data semantics linked with a domain resource*), and personal (*user-related data and profiles*). Furthermore, it supports optional and extensible enrichment data, including sentiment and engagement metrics (e.g., *likes, shares*). *TextBI* also comes with a generic data model (APs Model; Masson et al., 2023b) capable of modeling any annotated corpus from social media along the aforementioned dimensions.

Data collection and NLP must be undertaken before its use to produce an annotated corpus of social media posts. *TextBI* is designed to address the requirements of two distinct categories of users:

- **Domain Stakeholders**: seeking specific insights related to their field. For example, in the tourism industry, tourism offices might find it beneficial to analyze certain types of information. This could include identifying the most popular tourist activities, determining which cities are often visited together, and understanding the emotions or opinions of visitors regarding their experiences. This information can assist them in making informed decisions.
- **NLP Researchers** in need of a tool to evaluate NLP processes and models. This could involve observing the distribution of various types of annotations to better diagnose recurring issues.

Both categories of users may have different end goals, but there might be common interests in terms of what they want to observe. We believe that both groups are likely to share an interest in identifying frequencies, associations, and sequences of annotations. Additionally, they may want to conduct cross-dimensional analyses involving different types of annotations. Thus, *TextBI* could meet the requirements for both categories of end-users.

Previous work on Business Intelligence (BI; Datig and Whiting, 2018; Orlovskyi and Kopp, 2020; Vashisht and Dharia, 2020; Desai et al., 2021), Geographical Information Systems (GIS; Kurt Menke et al., 2016) and NLP tools (Chantrapornchai and Tunsakul, 2021; Rajaonarivo et al., 2022) have addressed the visualization of NLP annotations in an independent manner. In this work we take into account previous approaches while also addressing their respective limitations. From BI tools, we borrow the interactive design, user-friendly interfaces, and visual synchronization, adapting these features for their use with annotated text data as opposed to traditional numerical data. We adopt detailed spatial views of GIS tools, acknowledging that social media data often contains a spatial aspect. However, unlike traditional GIS tools, we aim for a multidimensional approach, not just a spatial one. From NLP tools, we take their analytical strength, such as co-occurrence and frequency analysis, but go beyond their usual focus on text and basic words to include dimensional entities like thematic concepts, locations, and time periods. By merging these elements, TextBI aims to provide an inclusive visualization of NLP annotations that benefits both researchers and domain-specific stakeholders.

Summarizing, we believe that *TextBI* represents a significant advancement in the field of visualizing NLP annotations by integrating and blending features of a variety of tools.

2 Related Work

Geographical Information Systems (GIS) such as QGIS (Kurt Menke et al., 2016) and ArcGIS (Booth et al., 2001) offer a wide variety of functionalities to meet the varying needs of different users. While these systems are useful for visualizing geospatial data in depth, their primary focus is on the spatial aspect of data, which limits their usefulness in non-spatial data contexts. Although GIS tools are capable of displaying thematic data (Murthy et al., 2003), they must be associated with a specific spatial area.

In the realm of **Business Intelligence (BI) tools**, Tableau (Datig and Whiting, 2018), Power BI (Ferrari and Russo, 2016), and QlikView (Shukla and Dhir, 2016) are recognized for their ability to empower decision-making processes (Hansoti, 2010; Orlovskyi and Kopp, 2020) via their interactive data exploration and user-friendly dashboards. However, these tools are primarily designed to handle numerical and well-structured data, resulting in significant challenges when working with text data. While certain efforts have been made to incorporate NLP processes into BI tools (Vashisht and Dharia, 2020; Desai et al., 2021), they struggle to present sequential data (*e.g., trajectories*) and draw connections across text annotations from various dimensions. Additionally, these tools lack comprehensive support for multilingual data.

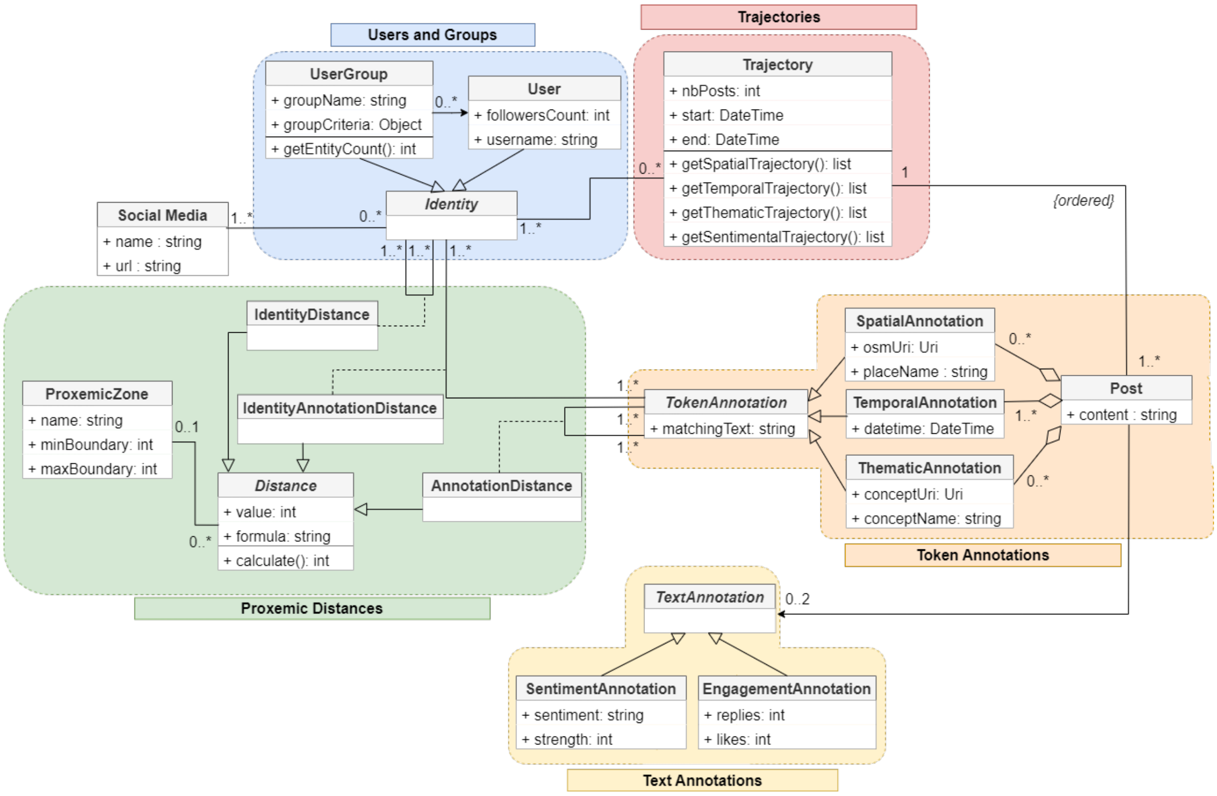

Figure 2: The generic APs data model using proxemics to model social media data (Masson et al., 2023b).

When it comes to **NLP tools**, they can be divided into two main categories. Some of them, such as SpaCy (Chantrapornchai and Tunsakul, 2021), TextRazor (Rajaonarivo et al., 2022), GATE (Maynard et al., 2000), and Gensim (Rehurek and Sojka, 2011), primarily focus on data processing, offering limited visualization capabilities such as word clouds, semantic graphs, and text-based annotation views. Other tools, such as IRaMuTeQ (Loubère and Ratinaud, 2014), Voyant (Welsh, 2014), VOSviewer (Van Eck and Waltman, 2013), and SentimentViz (Healey and Ramaswamy, 2022), offer a broader range of visualization options, but are generally focused on a single dimension (like *sentiment*) or word-based statistical analyses. However, some of these tools can be challenging for non-computer scientists due to their complexity.

3 Generic Data Model

The *TextBI* dashboard is based on a data model called the *APs Model* with the aim of modeling in a generic way any kind of annotated social media corpus. This data model consists of 5 dimensions. The class diagram of the data model is shown in Fig. 2.

The **Users and Groups** dimension allows modeling the studied population: individual users and user groups featuring common characteristics or traits.

The **Trajectory** dimension provides the ordered sequence of posts belonging to a given user. It gives a comprehensive view of an user's activities on the chosen social media and allows linking posts together. It can be broken down into several sub-trajectories (*spatial, thematic, spatio-thematic, etc.*).

The **Token Annotation** dimension models the posts along with their associated *token annotations*. A given post can have several token annotations. Those can be spatial (*toponyms*), temporal (*temporal entities*) or thematic (*domain concepts*). Thematic annotations are resolved according to the studied domain description (domain specific ontology, thesaurus or dictionary). These semantic resources provide additional hierarchy information. When it comes to spatial annotations, those are associated with a unique identifier linked to a spatial database. This allows for places to feature hierarchical relationships (e.g., *a city is within a region, itself within a country*).

The **Text Annotation** dimension contains *Sentiment Annotation* which models the overall sentiment of the associated post and *Engagement Annotation* which models the engagement associated with a given post (*based on the number of replies, likes and quotes*).

The classes for both text and token annotations are designed to be extensible, thereby empowering the end user with the flexibility to incorporate new annotation types as desired. Annotations can be instantiated by running NLP modules on the extracted social media posts (*e.g., NER, concept extraction*) or by parsing post metadata (*e.g., engagement metrics, geotags, etc.*).

Lastly, the **Proxemic Distances** dimension allows calculating and storing distances across spatial, temporal, thematic or individual entities. Such distances can be computed within a unique dimension or through two distinct dimensions.

4 Annotation Setup

The first step (see Fig. 1) in generating our visualization with *TextBI* consists of initializing the data model using an annotated corpus of data tied to a domain-specific knowledge base. In our case, we make use of the *Thesaurus on Tourism and Leisure Activities* (World Tourism Organization, 2002), an extensive multilingual terminology with over 2500 touristic concepts. With respect to the corpus, we collected 3,293 tweets issued by tourists during the summer of 2019 from the social media X (formerly Twitter) [2]. The data collection approach is outlined in Masson et al. (2022). The data model was instantiated using two types of data: metadata and tweet content. Metadata-based instantiation was a straightforward process, encompassing elements like engagement metrics, profile features, timestamps, and geotags, among others. The textual content of the tweets was processed using three deep learning-based NLP modules to generate automatic annotations at the token level for (1) locations and (2) thematic concepts, and at the text level for (3) sentiment. We divided the dataset into three parts with a split of 60% for training, 20% for validation, and 20% for testing, ensuring a uniform distribution of languages and users across these splits. The main objective of these experiments was to establish the optimal method while keeping the amount of manually labelled data to a minimum. This meant testing various few-shot and fine-tuning techniques for each of the three NLP tasks mentioned above (Masson et al., 2023a).

Thus, with respect to sentiment analysis (1), we experimented with various techniques such as using pre-trained models fine-tuned on our tweet corpus, and employing a few-shot method such as Pattern-Exploiting Training (PET; Schick and Schütze, 2020). However, due to space constraints and the paper's focus, we cannot elaborate on a comprehensive evaluation. Ultimately, we found that the fine-tuning the XLM-T model for sentiment (Barbieri et al., 2022) was the most effective approach, achieving a high accuracy value of 0.939, compared to 0.877 for PET with the language model.

We adopted a similar methodology for Named Entity Recognition (NER) (2), which involved a single class (LOC) that encompassed a broad array of 995 place names. Here, the challenge was the low frequency of the label words in the annotated data. Despite this, fine-tuning multilingual BERT (mBERT; Devlin et al., 2018) demonstrated superior performance with an F1-micro (Tjong Kim Sang, 2002) score of 0.848, compared to 0.788 with the few-shot technique for sequence labelling implemented by EntLM (Ma et al., 2021).

Lastly, for the task of thematic concept extraction (3), which included a highly diverse set of classes (315 instantiated touristic concepts from the thesaurus) with few label words, we found that prompt-based few-shot learning with EntLM was significantly more effective, achieving an F1-micro score of 0.840, compared to 0.241 for the fine-tuning approach. For extensive details of the experiments briefly outlined in this section please check Masson et al. (2023a).

In the next section, we will describe how to visualize these automatically generated annotations using *TextBI*.

5 The TextBI Dashboard

The *TextBI* dashboard (Fig. 3) is a web-based tool that allows interactive visualization of NLP annotations for social media data. It is designed to be adaptable to any social media and domain, as long as the data adheres to the specified data model (see Section 3). The dashboard offers a variety of features to facilitate the analysis of multilingual social media data across four dimensions: spatial, temporal, thematic, and personal. In our context, the domain is tourism, implying potential stakeholders could include tourism offices or Destination Marketing Organizations (Gretzel et al., 2006).

[2] https://twitter.com/

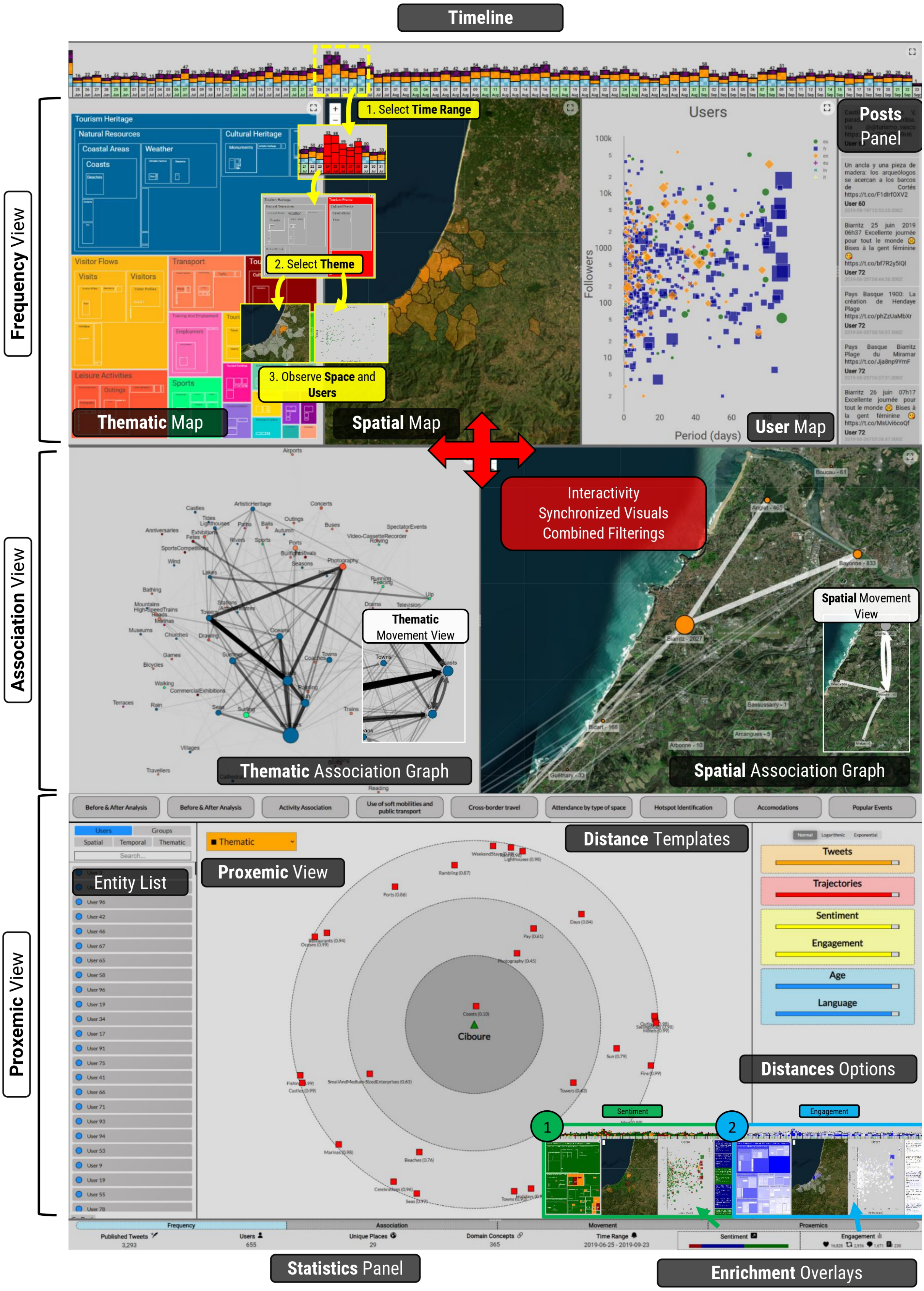

Figure 3: Overview of the TextBI platform. Live demonstration is available: https://youtu.be/x714RKvo9Cg

5.1 Frequency View

The **Frequency** view (Fig. 3) acts as the main interface, highlighting spatial, temporal, and thematic frequencies through four major visualizations.

The **Thematic Map** treemap (Fig. 3, *Thematic Map*) visualizes the hierarchical structure and frequency of thematic concepts from a semantic resource, such as a dictionary, thesaurus, or ontology. For instance, mapping tweet concepts to the *Thesaurus on Tourism and Leisure Activities* (World Tourism Organization, 2002), we find tourism heritage-related concepts are most frequent, accounting for roughly 40% of discovered concepts, with many linked to natural resources such as the coast and sea.

The **Spatial Map** (Fig. 3, *Spatial Map*) is another visualization showcasing the frequency of places mentioned in posts. Users can set the spatial granularity, which can range from broad categories such as countries to more specific ones like Points of Interest (POIs). Places are aggregated depending on the chosen granularity. The map uses a linear gradient for representation, with more transparent areas signifying fewer originating posts. Here, we focus on the French Basque Coast region at the city level and observe a hotspot of tweets in 3 cities at the northernmost part of the region. These nearby cities appear to be popular among visitors.

Next, the **User Map** (Fig. 3, *User Map*) is a scatter plot that presents the users' posting frequency on the x-axis against the count of the users' followers on the y-axis. This design helps in the rapid identification of influential users. Each user's language is represented by color and symbol, with the symbol's size corresponding to the number of posts from that user. 655 users are depicted in this example, spanning over 6 languages (*French, Spanish, English, Basque, Italian, and undetermined*).

Lastly, the **Timeline** (Fig. 3, *Timeline*) view offers a visualization of the volume of posts per day across the dataset range, divided into different times of the day such as morning, afternoon, or evening. It provides various temporal granularity options, including days, months, seasons, and years. Here, we use daily granularity for the summer of 2019. We observe a peak of tweets between the 24th and 28th of July.

5.2 Association View

In the **Association** view (Fig. 3, *Association View*), the dashboard presents visual representations that illustrate the connections between entities through their co-occurrences in posts. These connections are depicted using non-directed graphs, where the nodes represent entities such as thematic concepts or places, and the edges indicate the strength of co-occurrence between them. This allows for easy identification of heavily correlated concepts or places. As expected, *Sun*, *Beach*, *Surfing*, *Sea*, and other coastal concepts are heavily linked. This view can also display movements (Fig. 3, *Movement Views*), focusing on the sequencing of entities in user trajectories, for example, the transition from one thematic concept or place to another. This sequencing is visualized through directed graphs where edges indicate the amount of time two concepts or places are sequenced in user trajectories.

5.3 Proxemics View

TextBI's proxemic view (Fig. 3, *Proxemic View*) offers users the ability to analyze datasets via a proxemic approach (Hall et al., 1968; Greenberg et al., 2011), selecting entities such as a user, group, thematic concept, place, or time period as references. For instance, in our demonstration, we selected *Ciboure*, a touristic city, as the reference entity, and compared it with thematic concepts. The interface includes a left panel for selecting references, a central crosshair panel for results, and a right settings panel for customizing distance calculations. The system supports many combinations such as user-to-themes or place-to-users.

While specifics of the distance calculation formula are not covered in this paper, users can customize settings, for example, assigning higher weight to positive or highly-engaged tweets. The interface also features a template panel offering pre-defined distance settings for specific study domains. In the tourism domain, the *Accommodation* template, for example, limits the display to accommodation-related thematic concepts.

5.4 Enrichment Overlays

In *TextBI*, visuals support the superimposition (referred to as **overlays**) of sentiment and engagement enrichment data (Fig. 3, *Enrichment Overlays*). Sentiment is indicated through color coding (*green for positive, red for negative, and orange for mixed sentiment*), enabling a better understanding of aggregate sentiment by themes, places, time or user. Engagement is visualized using a linear gradient (*blue indicating strong engagement, white indicating low engagement*), providing insights into user engagement. We can see that most tourist concepts tend to be associated with a positive sentiment, but some, like *Transport* or *Ecology*, are more mixed.

5.5 Interactions and Visual Synchronization

The *TextBI* platform employs interactions commonly found in Business Intelligence (BI) tools, ensuring a fluid synchronization of visualizations. It accommodates multi-dimensional filtering options such as spatial-temporal, spatial-thematic, and user-temporal. When a user selects a particular location, theme, user, or time range, all subsequent visualizations adjust to display only tweets associated with the chosen filter. The system even allows for combined filtering. Within its proxemic view, users can conveniently drag and drop references onto the center of the crosshair panel.

Consider the example depicted in Fig. 3 highlighted in yellow. If the user chooses the time range of July 24th to 27th, the thematic map updates to display a higher concentration of the *Celebration* thematic concept. Further clicking on the *Celebration* concept leads the spatial map to highlight a hotspot in the city of *Bayonne*. This coincides with the timing of the *Bayonne Celebration*, a local event that attracts over a million attendees.

The dashboard base features a *statistics panel* (Fig. 3, *Statistics Panel*) showing data related to active filters, including post, user, concept, and place counts; current post time range; total engagement level; and prevailing sentiments. Filtered posts are displayed in the *Posts Panel* on the right.

5.6 Technical Aspects and Limitations

TextBI is a web application developed using HTML, CSS, and JavaScript. It runs solely on the client-side, so it does not require a back-end web server and can be run locally. The data model has been implemented using the JSON format.

TextBI serves as a data display and aggregation tool, providing statistical analyses and calculating distances. It does not engage in any data processing tasks of its own. Data collection and transformation, including NLP, need to be completed in advance. Currently, *TextBI* does not support any analytical dimensions beyond those mentioned above. In the future, we aim to make it easily extensible through a plugin system.

6 Conclusion and Future Perspectives

We have introduced a novel dashboard called *TextBI*, designed to facilitate the visualization of automatic NLP annotations on social media data for both domain stakeholders and NLP researchers. It is powered by a generic data model, making it completely adaptable. *TextBI* focuses on four dimensions: space, time, theme, and user, and offers various viewing options. Additional enrichment data is also supported. *TextBI* provides extensive interactivity, including combined filtering, visual synchronization, aggregation, and more. Our future plans include enhancing TextBI with features such as a user interface for granularity selection and enabling it to process live data by integrating with *InfluxDB* (Ahmad and Ansari, 2017). After testing it on other domains and larger datasets to ensure its scalability, we intend to make *TextBI* open-source.

Acknowledgments

This work has been partially supported by the HiTZ center and the Basque Government (Research group funding IT-1805-22). We also acknowledge the funding from the following projects: (i) DeepKnowledge (PID2021-127777OB-C21) MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe; (ii) Disargue (TED2021-130810B-C21), MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR. Rodrigo Agerri currently holds the RYC-2017-23647 fellowship (MCIN/AEI/10.13039/501100011033 and by ESF Investing in your future). Additionally, this work was also partially supported by the urban community of Pau Béarn Pyrénées.

References

Khaleel Ahmad and Masroor Ansari. 2017. Hands-on influxdb. In *NoSQL: Database for Storage and Retrieval of Data in Cloud*, pages 341–354. Chapman and Hall/CRC.

Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. 2022. Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266.

Bob Booth, Andy Mitchell, et al. 2001. Getting started with arcgis.

Chantana Chantrapornchai and Aphisit Tunsakul. 2021. Information extraction tasks based on bert and spacy on tourism domain. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 15(1):108–122.

Ilka Datig and Paul Whiting. 2018. Telling your library story: tableau public for data visualization. *Library Hi Tech News*, 35(4):6–8.

Zeel Desai, Karishma Anklesaria, and Harish Balasubramaniam. 2021. Business intelligence visualization using deep learning based sentiment analysis on amazon review data. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7. IEEE.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Alberto Ferrari and Marco Russo. 2016. *Introducing Microsoft Power BI*. Microsoft Press.

Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. 2011. Proxemic interactions: the new ubicomp? *interactions*, 18(1):42–50.

Ulrike Gretzel, Daniel R Fesenmaier, Sandro Formica, and Joseph T O'Leary. 2006. Searching for the future: Challenges faced by destination marketing organizations. *Journal of Travel Research*, 45(2):116–126.

Edward T Hall, Ray L Birdwhistell, Bernhard Bock, Paul Bohannan, A Richard Diebold Jr, Marshall Durbin, Munro S Edmonson, JL Fischer, Dell Hymes, Solon T Kimball, et al. 1968. Proxemics [and comments and replies]. *Current anthropology*, 9(2/3):83–108.

Bhumika Hansoti. 2010. Business intelligence dashboard in decision making. *Purdue University, College of Technology Directed Projects*.

Healey and Ramaswamy. 2022. Twitter sentiment visualization. Accessed: 2023-06-30.

GISP Kurt Menke, Richard Smith Jr, Luigi Pirelli, GISP John Van Hoesen, et al. 2016. *Mastering QGIS*. Packt Publishing Ltd.

Daniel Leung, Rob Law, Hubert Van Hoof, and Dimitrios Buhalis. 2013. Social media in tourism and hospitality: A literature review. *Journal of travel & tourism marketing*, 30(1-2):3–22.

Lucie Loubère and Pierre Ratinaud. 2014. Documentation iramuteq 0.6 alpha 3 version 0.1. *http://www.iramuteq.org*.

Ruotian Ma, Xin Zhou, Tao Gui, Yiding Tan, Linyang Li, Qi Zhang, and Xuanjing Huang. 2021. Template-free prompt tuning for few-shot ner. *arXiv preprint arXiv:2109.13532*.

Maxime Masson, Rodrigo Agerri, Christian Sallaberry, Marie-Noelle Bessagnet, Annig Le Parc Lacayrelle, and Philippe Roose. 2023a. Optimal strategies to perform multilingual analysis of social content for a novel dataset in the tourism domain. *arXiv preprint arXiv:2311.14727*.

Maxime Masson, Philippe Roose, Christian Sallaberry, Rodrigo Agerri, Marie-Noelle Bessagnet, and Annig Le Parc Lacayrelle. 2023b. Aps: A proxemic framework for social media interactions modeling and analysis. In *International Symposium on Intelligent Data Analysis*, pages 287–299. Springer.

Maxime Masson, Christian Sallaberry, Rodrigo Agerri, Marie-Noelle Bessagnet, Philippe Roose, and Annig Le Parc Lacayrelle. 2022. A domain-independent method for thematic dataset building from social media: the case of tourism on twitter. In *International Conference on Web Information Systems Engineering*, pages 11–20. Springer.

Diana Maynard, Hamish Cunningham, Kalina Bontcheva, Roberta Catizone, George Demetriou, Robert Gaizauskas, Oana Hamza, Mark Hepple, Patrick Herring, Brian Mitchell, et al. 2000. A survey of uses of gate. Technical report, Technical Report CS–00–06, Department of Computer Science, University of Sheffield.

KSR Murthy, E Amminedu, and V Venkateswara Rao. 2003. Integration of thematic maps through gis for identification of groundwater potential zones. *Journal of the Indian Society of Remote Sensing*, 31:197–210.

Dmytro Orlovskyi and Andrii Kopp. 2020. A business intelligence dashboard design approach to improve data analytics and decision making. In *Selected Papers of the 7th International Conference "Information Technology and Interactions" (IT&I-2020). Conferece Proceedings, Kyiv, Ukraine, December 02-03, 2020*, volume 2833 of *CEUR Workshop Proceedings*, pages 48–59. CEUR-WS.org.

Landy Rajaonarivo, Tsunenori Mine, and Yutaka Arakawa. 2022. Coupling of semantic and syntactic graphs generated via tweets to detect local events. In *2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 128–133. IEEE.

Radim Rehurek and Petr Sojka. 2011. Gensim—statistical semantics in python. *Retrieved from genism. org*.

Timo Schick and Hinrich Schütze. 2020. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint arXiv:2001.07676*.

Ayushi Shukla and Saru Dhir. 2016. Tools for data visualization in business intelligence: case study using the tool qlikview. In *Information Systems Design and Intelligent Applications: Proceedings of Third International Conference INDIA 2016, Volume 2*, pages 319–326. Springer.

Achille Souili, Denis Cavallucci, and François Rousselot. 2015. Natural language processing (nlp)–a solution for knowledge extraction from patent unstructured data. *Procedia engineering*, 131:635–643.

Erik F. Tjong Kim Sang. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

Nees Jan Van Eck and Ludo Waltman. 2013. Vosviewer manual. *Leiden: Univeristeit Leiden*, 1(1):1–53.

Vipul Vashisht and Pankaj Dharia. 2020. Integrating chatbot application with qlik sense business intelligence (bi) tool using natural language processing (nlp). In *Micro-Electronics and Telecommunication Engineering: Proceedings of 3rd ICMETE 2019*, pages 683–692. Springer.

Megan E Welsh. 2014. Review of voyant tools. *Collaborative Librarianship*, 6(2):96–98.

World Tourism Organization. 2002. *Thesaurus on Tourism and Leisure Activities*. World Tourism Organization.