

MultiModalTopicExplorer: Topic modeling for exploring multi-modal data from asynchronous online conversations

Soheil Alavi

salavis@cs.ubc.ca

Felipe Gonzalez-Pizarro

felipegp@cs.ubc.ca

Abstract

The constant increase in the volume of textual data has led to the development of various algorithms to summarize and understand this type of data. A promising solution is topic modeling, a statistical approach for extracting themes from high volumes of data. Humans who directly interact with and interpret the output of topic modeling may rely on visualization tools to better interpret the results. However, these tools still have limitations. For example, current visual representations lack support for multi-modal conversations (image and text) and do not consider how topics may change over time. We plan to design and develop MultiModalTopicExplorer, a web-based interactive visualization of topic modeling algorithm results to address these limitations. There are two key innovations in this work. First, MultiModalTopicExplorer identifies the most relevant images to each topic to help users during topic interpretation. Second, it allows NLP-experts to do a qualitative analysis of topic modeling algorithms. To examine MultiModalTopicExplorer, we will implement two different topic modeling algorithms: (1) LDA and (2) BERTopic. Based on time availability, we plan to conduct a user study to compare the quality of two topic modeling algorithms and identify the users' workload during topic interpretation using our proposed tool.

1 Introduction

The constant increase of the volume of textual data has led to the development of various algorithms intended to summarize and understand unstructured textual data (Peter et al., 2015). A promising solution to this problem is topic modeling, a robust statistical approach for extracting core themes or *topics* from large text corpora. Thus, when a topic modeling algorithm is applied to a large corpus of documents, such as a collection of news articles,

the results will include a list of topics, such as politics, economy, or sports. Each topic is defined by a set of descriptive words ranked according to their importance for the topic and by its distribution over the corpus documents (El-Assady et al., 2018).

Although powerful, topic models do not interpret themselves; therefore, humans must be involved (Chang et al., 2009; Dou et al., 2011). Visual text analytics researchers have designed algorithms and visual representations to support topic sense-making and interpretation, making probabilistic topic results legible and exploratory to a broader audience. (Dou et al., 2013). Topic modeling visualization tools help in understanding topic models output and issues in modeling (Kherwa and Bansal, 2019); however, they still have limitations. Finding mechanisms to improve these visual representations is still an open challenge (Jelodar et al., 2019).

Current topic modeling visualization tools lack support for multi-modal conversations. With the proliferation of web-based social media, there has been an exponential growth of asynchronous online conversations discussing a large variety of popular issues such as “US 2016 Election”, or “Samsung watch release” (Hoque and Carenini, 2015). To discuss these and other topics, social media users post textual and image data. To the best of our knowledge, non of the current topic modeling visualization tools support image representation of topics.

Additionally, traditional topic modeling visualization tools do not take into account how topics may change over time. The content discussed on social media websites is diverse, and users often react to current events as they happen, so the content is constantly evolving (Smith et al., 2015). Topic modeling visualization end-users can be interested in analyzing the evolution of main discovered topics over time.

To address these limitations, we plan to design and develop MultiModalTopicExplorer, a web-based interactive visualization of topic modeling-generated topics. This tool aims to answer the following questions:

- What are the most relevant topics of the corpus?
- How do these topics evolve over time (month by month)?
- What is the meaning of each topic?

The remainder of the manuscript is organized as follows. Section 2 discusses two popular topic modeling algorithms. Section 3 summarizes related work about visual representations of topic models, discussing the current limitations and positioning this proposal. Section 4 describes the dataset we plan to use. Section 5 describes tasks abstractions. Section 6 introduces our proposal. Section 7 describes a user study that we plan to conduct to evaluate our tool. Section 8 indicates the number of estimated hours per milestone. Finally, Section 9 describes our work in progress.

Personal expertise

Seyed Hossein Alavi (Soheil): His research interests are: Conversational AI, Common Sense reasoning, and applied NLP. He has been working on a variety of applied NLP research projects for the past few years. Recently Soheil started working on a Question Answering research project that would benefit from both topic modeling and its visualization.

Felipe González-Pizarro: in his MSc. thesis, he designed, developed, and evaluated TopicVisExplorer, a set of web-based interactive visualizations of LDA-generated topics using D3.js and Python. His work is available here: <http://topicvisexplorer.tk/singlecorpus>, and <http://topicvisexplorer.tk/multicorpora>.

This project is for the courses: CPSC 547-Information Visualization and CPSC503-Computational Linguistics.

2 Topic modeling algorithms

This section briefly goes over two popular topic modeling algorithms we will use in this project: LDA and BERTopic.

2.1 LDA

A large number of techniques have been proposed for the extraction and tracking of relevant topics over a large amount of text, where Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is one of the most traditional and popular methods (Meeks and Weingart, 2012; Qiang et al., 2020). The LDA model is based on the assumption that document collections have latent topics in the form of a multinomial distribution of words, which is typically presented to users via its *top-N* highest probability words (Lau et al., 2014).

2.2 BERTopic

BERTopic (Grootendorst, 2020) is a topic modeling technique that leverages Transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions.

3 Topic model representation

The raw output of such topic modeling algorithms might be so complex that it can be difficult and time-consuming for non-expert users to understand it (Smith et al., 2014; Chaney and Blei, 2012; Murdock and Allen, 2015). To address this need and add analytic value, previous work has explored different visual representation approaches to support a human interpretation of topic models. Current topic modeling visualizations tools vary in topic keywords representations, documents view, and global views of topics.

3.1 Most relevant keywords

The most common output of topic modeling algorithms is the ranked list of the top terms of each particular topic (Kherwa and Bansal, 2019). They can be represented through different topic visualization techniques: (1) word lists; (2) word lists with bars; (3) word clouds; and (4) network graphs of terms (Smith et al., 2017). Among these alternatives, simple visualizations such as word lists or word lists with bars allow users to understand topics quicker (Smith et al., 2017).

Usually, the top keywords are shown to users as a ranked list of the most frequent terms for each particular topic (Chaney and Blei, 2012; Peter et al., 2015). In LDA, this is the same that ordering the terms by their topic-specific probability. The problem with representing topics this way is that frequent common terms in the corpus often appear

near the top of such lists for multiple topics, making it hard to users to find the differences between them (Sievert and Shirley, 2014).

To mitigate this problem, an intrinsic measure to rank terms within topics was proposed. It is called *lift* (Taddy, 2012), and it is defined as the ratio of a term’s probability within a topic to its marginal probability across the corpus. Thus, let $\phi_{k\omega}$ denote the probability of the term $\omega \in 1, \dots, V$ occurring in topic $k \in 1, \dots, K$, where V denotes the number of terms in the vocabulary and K the number of topics. Let p_ω denote the marginal probability of the term ω in the corpus. The ordering of keywords by *lift* is given by :

$$lift(\omega, k) = \left(\frac{\phi_{k\omega}}{p_\omega} \right) \quad (1)$$

This measure generally decreases the rankings of globally frequent terms, which can be helpful for topic interpretation. Nevertheless, it can be noisy in some cases by giving high rankings to very rare terms that occur in only a single topic. While such terms may contain useful topical content, if they are very infrequent, the topic may remain challenging to interpret (Sievert and Shirley, 2014).

Another intrinsic measure was proposed to mitigate *lift* limitations. It is called *relevance* (Sievert and Shirley, 2014), and it is based on both term’s frequency as well as its *exclusivity*, the degree to which its occurrences are limited to only a few topics. Thus, the *relevance* of term ω to topic k given a weight parameter λ (where $0 \leq \lambda \leq 1$) is defined as:

$$r(\omega, k|\lambda) = \lambda \log(\phi_{k\omega}) + (1 - \lambda) \log\left(\frac{\phi_{k\omega}}{p_\omega}\right) \quad (2)$$

where λ determines the weight given to the probability of term ω under topic k relative to its *lift* (measuring both on the log scale). Setting $\lambda = 1$ results in the ranking of terms in decreasing order of their topic-specific probability, and setting $\lambda = 0$ ranks terms solely by their *lift*. A user study found that the optimal value of λ for topic interpretation is 0.6 (Sievert and Shirley, 2014)

3.2 Most relevant documents

Sometimes the top keywords are not enough to identify the semantics of a topic (Hoque and Carenini, 2015). That is the case when the most relevant terms are poorly connected, or when they include disparate (Mimno et al., 2011) or generic terms (e.g., “yes”, “like”, “Mr”, “maybe”) (Lee

et al., 2017; Boyd-Graber et al., 2014). Due to that, it is better to include another level of information such as the most relevant documents to each topic to help end-users during topic interpretation (Hoque and Carenini, 2015; Yang et al., 2017). Indeed, previous research found that when topic modeling visualization tools display documents, users can read them to ensure topics’ quality and verify if they satisfy their expectation (El-Assady et al., 2018).

There is no clear consensus regarding the best method to display documents associated with particular topics. For instance, visualizations that aim to support users in exploring asynchronous conversations position the most relevant documents of a topic according to their chronological ordering (Hoque and Carenini, 2014, 2015). Another method is to display the documents according to their contribution to the topic, as (Smith et al., 2015). Thus, the most relevant documents always appear first. In LDA, this is the same as ordering the documents regarding the topic-document probability for each topic.

3.3 Global view of topics

Along with showing the most relevant keywords and the documents associated with topics, topic modeling visualization tools offer different layouts to help users get a global view of the topic model.

One alternative is to represent relevant keywords and documents from topics through a graph layout. That is the case of iVisClustering (Lee et al., 2012) where documents (graph nodes) from one topic are visualized as colored circles with the same color. The edges between nodes represent the similarity between documents based on cosine similarity. Controlling a slider makes edges with higher values than the slider value appear, and those with smaller values disappear. For each cluster, there is a color-bordered rectangle with the most representative keywords (see Figure 1 (a)).

A second approach consists in displaying the term-topic distributions through a matrix layout. In this approach, proposed in Termite (Chuang et al., 2012), the rows correspond to terms and the columns to topics. It uses circular areas to encode term probabilities. Thus, the most frequent terms are represented by circles with a larger area (see Figure 1 (b)).

A third alternative consists in projecting the similarity between topics into a two-dimensional space.

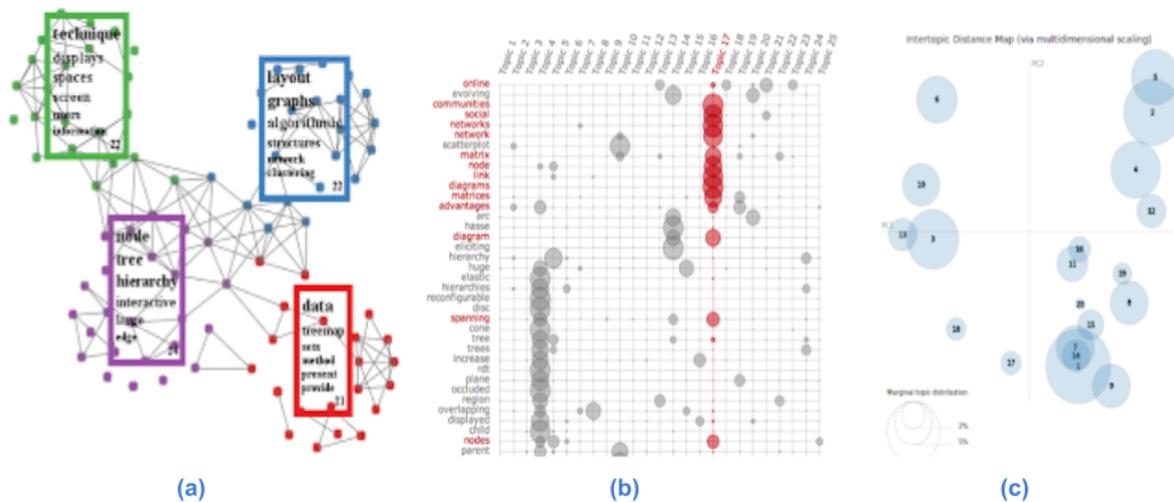


Figure 1: Layout of global view of topics in: (a) iVisClustering (b) Termite (c) LDAvis

In this approach, proposed in LDAvis (Sievert and Shirley, 2014) (see Figure 1 (c)), the topics are represented as circles. Their centers are determined by computing the distance between topics and then using multidimensional scaling to project the inter-topic distances onto two dimensions. In this layout, each topic’s overall prevalence is encoded using the areas of the circles, such that a more extensive area indicates a higher prevalence. This layout provides a global view of the topics, via their prevalences and similarities to each other, in a compact space.

3.4 Relevant commercial tools

As far as we know, there are not any relevant commercial tools or efforts from practitioners related to multi-modal topic modeling visualization.

4 Datasets

For this research effort, we focus on unstructured multimodal conversational data gathered from users’ interactions on famous social media forums such as 4Chan.

4.1 4Chan dataset

4chan is an imageboard website with virtually no moderation. An *Original Poster* (OP) creates a thread by posting an image and a message. Content is organized in subcommunities, called boards with various topics of interest. Other users can post in the OP’s thread with a message or an image. On 4chan, users do not need a registered account to post content. In this project, we will focus on the politically incorrect board (pol), which has been

shown to include a high volume of racist, xenophobic, and hateful content (Mittos et al., 2020)

We will use a random sample of the 4chan dataset gathered by (Papasavva et al., 2020), which contains over 3.3M threads and 134.5M posts from the Politically incorrect board (/pol/), posted over a period of almost 3.5 years (June 2016–November 2019) (Papasavva et al., 2020). We plan to evaluate how many threads can be incorporated into the visualization tool. The fact that users embed images in their posts can be useful for our novel multi-modal topic modeling visualization of conversational texts approach. An example of a typical pol thread is given in Figure 2.

Why this dataset?

Social media sites such as Twitter, Facebook, and 4chan allow users to instantly share their ideas and opinions. However, there are several ill consequences, such as online harassment, trolling, cyberbullying, fake news, and hate speech. We believe that an exploration of these conversations could help understand how these communities interact on these platforms. Moreover, it is the first step before creating automated hate speech detection and mitigation systems.

4.2 Other datasets

While we will focus on the 4chan dataset in this project, we expect that MultiModalTopicExplorer can support other social media data from different sources such as Twitter. Evaluating our tool with other datasets can provide insights regarding the quality of topic modeling algorithms.



Figure 2: Example of a typical pol thread

5 Task Abstraction

As explained in Section 4, we will work with multi-modal unstructured conversational data. As explained before, this tool aims to answer the questions: What are the most relevant topics of the corpus?, How these topics evolves over time?, and What is the meaning of each topic?

We will aggregate/summarize the data over periods of time (months) to figure out how topics have evolved throughout the time. In this research effort our goal is to provide a means for topic modeling visualization tool end-users (e.g. NLP researchers) to:

1. Identify what are the most frequent topics in the given corpus.
2. Identify out of the previously identified topics, which ones people have mostly talked about in a specific month.
3. Compare the frequency of the identified topics throughout the time.
4. Identify the most important (frequent) words from each topic.
5. Identify the most relevant images from each topic.
6. Identify a sample of conversations per month.

Although we mentioned NLP researchers as an example for our target audience, since the introduced datasets are widely used in other domains (e.g., social science, social computing, etc.), MultiModalTopicExplorer can potentially be used by researchers with backgrounds other than NLP as well.

6 MultiModalTopicExplorer

We propose an interactive visualization system called MultiModalTopicExplorer to address some limitations of previous topic modeling visualization tools. Different visual components are designed for this purpose. This section provides an overview of the visual interface features

The MultiModalTopicExplorer layout is illustrated in Figure 3. It has five components to allow users to get a sense of the most frequent topics and visualize specific information about them.

6.1 Most relevant topics

We will apply LDA, a widely-used topic modeling technique (Kim et al., 2020), to identify the main themes of the dataset. Our intuition is that there would be a large number of topics in the corpus, however, since we are only interested in the most frequent ones, as one of the design decisions, we will pick the top 20 frequent ones.

Figure 3 (a) shows the most relevant topics of the corpus. Each row represents a topic. The default

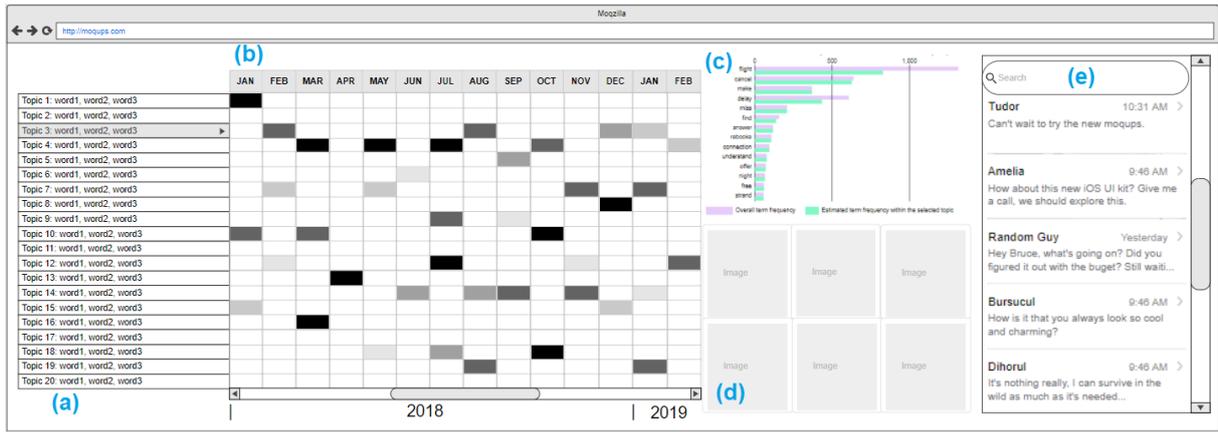


Figure 3: MultiModalTopicExplorerlayout

name of each topic is indicated by its three most relevant keywords. Users can click in each row to visualize specific information about the topic such as: its most relevant keywords (see Figure 3 (c)), and its most relevant images Figure 3 (d)). A summary of what, why, and how analysis of the component is described in Table 1.

Table 1: What-Why-How analysis of most relevant topics component

What: Data	Table: one categorical key attribute (topic name), one quantitative value attribute (topic relevance score).
What: Derived	Ordered key attribute (by topic relevance score).
How: Encode	Express value attribute with vertical position.
Why: Task	Identify and compare topics.
Scale	Items: twenty.

6.2 Evolution of topics over time

MultiModalTopicExplorer also allows users to identify the evolution of the most frequent topics over time. Figure 3 (b) shows a matrix where each column represents a period (month), and each row represents a topic. We use luminance to indicate the topic’s popularity per month (lower luminance indicates more popularity). Moreover, users can click a column to visualize a sample of conversations for that specific period (see Figure 3 (e)). A summary of what, why, and how analysis of the component is described in Table 2.

Table 2: What-Why-How analysis of evolution of topics over time component

What: Data	Table: two categorical key attributes (topic name, month), one quantitative value attribute (topic’s popularity per month).
What: Derived	Four quantitative attributes to indicate topic’s popularity
How: Encode	2D matrix alignment of area marks, luminance-map.
Why: Task	Find clusters, outliers, summarize
Scale	Categorical attribute levels: dozens to hundreds (twenty rows, hundreds of columns). Quantitative attribute levels: 4 (luminance levels).

6.3 Most relevant keywords

The most relevant keywords panel of MultiModal-TopicExplorerlayout (see Figure 3 (c)) depicts a horizontal bar chart for the most relevant terms to the selected topic. For each term, two bars are unfolded. Violet bars represent the corpus-wide frequency of a given term, and the green bar represents the topic-specific frequency of such terms. This kind of linked selection allows users to examine a large number of topic-term relationships compactly and supports users in topic interpretation (Sievert and Shirley, 2014). The most useful terms to a given topic are ranked according to the *relevance* score, allowing users to flexibly rank terms in order of usefulness for topic interpretation (Sievert

and Shirley, 2014). A higher relevance score designates greater importance to the frequency of terms within the selected topic (green bar). However, at the same time, it reduces the importance of their exclusivity. In other words, how rare these words are on other topics. We use a relevance score equal to 0.6, as is suggested by a prior used study (Sievert and Shirley, 2014). We plan in future versions of our project to incorporate a slider allowing users to alter the relevance score hyper-parameter to better interpret the results. A summary of what, why, and how analysis of the component is described in Table 3.

Table 3: What-Why-How analysis of most relevant keywords component

What: Data	Table: one categorical key attribute (keyword), three quantitative value attributes (keyword relevance score, overall term frequency, estimated term frequency within the selected topic)
What: Derived	Ordered key attribute (by keywords relevance score).
How: Encode	Line marks, express value attribute with aligned horizontal position, separate key attribute with vertical position.
How: Encode	Express value attribute (keyword relevance score) with vertical position
Why: Task	Lookup and compare values.
Scale	Key attribute: twenty levels.

6.4 Most relevant images

We believe that images can help users to have a complementary perspective on the meaning of each topic. Figure 3 (d) shows the six most relevant images of the topic. These images will be retrieved from the 4chan dataset using CLIP (Radford et al., 2021), a neural network model, which recognizes a wide variety of visual concepts in images and associates them with names. We will use this model to retrieve the most related images from the dataset to the top ten keywords of each topic. To the best of our knowledge, this is the first topic modeling visualization tool that supports image content. A summary of what, why, and how

analysis of the component is described in Table 4.

Table 4: What-Why-How analysis of most relevant images component

What: Data	Table: one categorical key attribute (image), one quantitative value attribute (relevance score to the selected topic)
How: Encode	Express value (relevance) with horizontal and vertical spatial position and image marks.
Why: Task	Identify and compare
Scale	Items: six to nine.

6.5 Conversations View

This proposal includes a new component to get a sample of conversations for each month (see Figure 3 (e)). In this panel, users can also search for conversations that contain specific terms. Doing so will highlight the searched term in the documents. A summary of what, why, and how analysis of the component is described in Table 5.

Table 5: What-Why-How analysis of conversations view component

What: Data	Table: two categorical key attributes (month, document),
How: Encode	Vertical position separate documents. Text marks. Color hue to highlights searched terms.
Why: Task	Part-to-whole relationship, Lookup documents and terms.
Scale	Items: dozens to hundreds

6.6 Implementation

MultiModalTopicExplorer will be written in Javascript and Python. It will support within-browser interaction capabilities using D3 (Bostock et al., 2011). The visualization can be saved to a stand-alone HTML file for easy sharing. The full code will be available online.

6.7 Ethical considerations

We will analyze publicly available data which are masked for privacy purposes. We will follow stan-

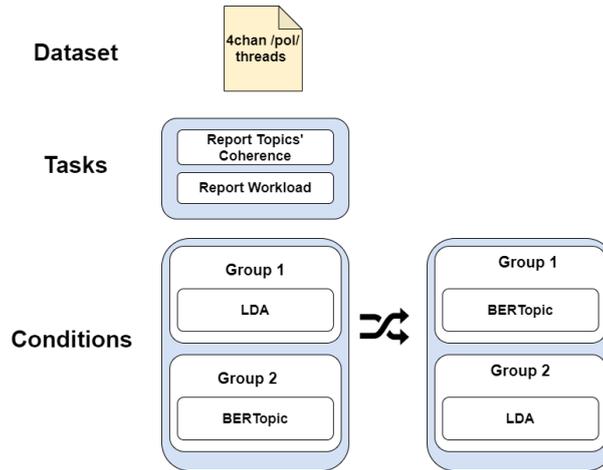


Figure 4: Dataset, tasks, and conditions in our user study

standard ethical guidelines, not attempting to track users or deanonymize them.

6.8 Scenario of use

In this section we consider 3 possible scenarios of how users can interact with MultiModalTopicExplorer.

Scenario 1

A user wants to get an overview of all the frequent topics in the corpus. MultiModalTopicExplorer provides users a list of top 20 topics in the corpus that are sorted by their importance (e.i. frequency). If they click on each topic, the horizontal bar chart (shown by (c) in the figure) gets updated showing the representative terms for that topic, sorted by their frequencies. Additionally, images in window (d) will get updated and they will be replaced by the most relevant images for that topic.

Scenario 2

A user wants to see which important topics have been discussed within a specific month. The columns in the matrix in Figure 3 (shown by (b) in the figure) are aggregated threads of conversations throughout that specific month. If user clicks on the columns, a sample of conversations from that month will be shown in window (e). Each cell (i, j) in the matrix indicates if topic i was discussed in month j or not, and if it has, how frequent it has been.

Scenario 3

A user is interested to see how topics have evolved throughout the time (multiple months). They can achieve this goal by using the mentioned matrix.

For example, if for Jan 2018, the cell has lower luminance than September 2018 for topic i , this means that topic i has been discussed more on Jan 2018 in comparison to September 2018. Although users can see a window with size 12 months, they can scroll throughout the history by the horizontal scroll at the bottom of the matrix to cover more dates.

7 Evaluation

Based on time availability, we plan to conduct a user study to compare the quality of two topic modeling algorithms and identify the users' workload during topic interpretation. To capture the quality of topics, we will ask participants to rate each topic on a 5-point scale, indicating how coherent the topic is. A higher value indicates a higher coherence. Moreover, we will use the NASA Task Load Index (NASA-TLX) to allow users to self-report the workload perceived on a scale from 0 to 100. This questionnaire identifies six dimensions: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration level. The average score for these dimensions is called the unweighted NASATLX score. It is the most common method to evaluate and report the overall workload level perceived during the task (Cao et al., 2009). Figure 4 summarizes the user study set up.

8 Milestones

Table 6 shows the milestones and their estimated required time we foresee for this project. The overall estimated required time to finish this project is 220 hours.

Table 6: Number of estimated hours per task

Task	Total	Felipe	Soheil	Deadline	Progress
Literature review	6	3	3		Done
Brainstorming and mock-up design	6	4	2		Done
Write proposal	32	16	16		Done
Investigate previous similar implementations	10	5	5	Nov 9	Done
Learn D3.js	16	6	10		In Progress
Data preprocessing	10	5	5		Done
LDA implementation	12	6	6		Done
BERTopic implementation	15	0	15		Done
Integrate both algorithms and visualize the most important topics (Figure 3 (a))	10	5	5	Nov 20	In progress
Aggregate data for each month	6	3	3		Not started
Find important topics for each month	6	3	3		Not started
Visualize the aggregated data and topics (Figure 3 (b))	20	10	10	Nov 25	Not started
Find the most important words for each topic	3	1.5	1.5		In progress
Visualize the term frequencies (Figure 3 (c))	10	5	5	Nov 21	In progress
Preprocess and retrieve the most related images for each topic	10	5	5		Not started
Visualize the most related images for each topic (Figure 3 (d))	10	5	5	Nov 30	In progress
Visualize sample of conversations for each month (Figure 3 (e))	10	5	5	Dec 7	Not started
Presentation	10	5	5	Dec 15	Not started
Final Report	20	10	10	Dec 17	Not started

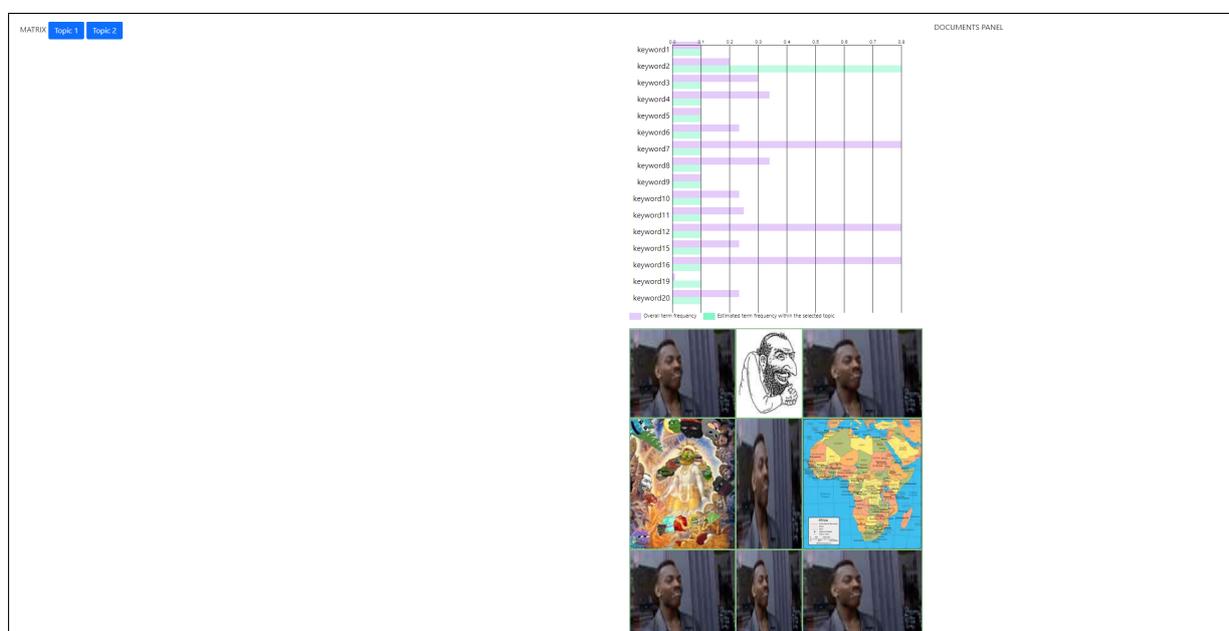


Figure 5: Current front-end status of MultiModalTopicExplorer

9 Work in progress

This section describes our current work in progress. In this stage, we have run the analysis in a small portion of our dataset (1000 documents) to avoid long execution times.

9.1 Preprocessing

Before applying topic modeling algorithms, in our corpus, we removed stopwords and transformed the text to lowercase. Also, we identified bigrams and trigrams. Moreover, we used lemmatization as is suggested by (González et al., 2019).

9.2 BERTopic initial implementation

We have implemented BERTopic following the instructions from <https://github.com/MaartenGr/BERTopic>. In order to create a topic representation, we took the top 20 words per topic based on their c-TF-IDF scores. The higher the score, the more representative it should be of its topic as the score is a proxy of information density.

We used visualization techniques to get an initial idea of the quality of the results. Figure 6 shows the most relevant keywords for some topics in our model. Moreover, Figure 7 shows the similarity between topics in our BERTopic model.

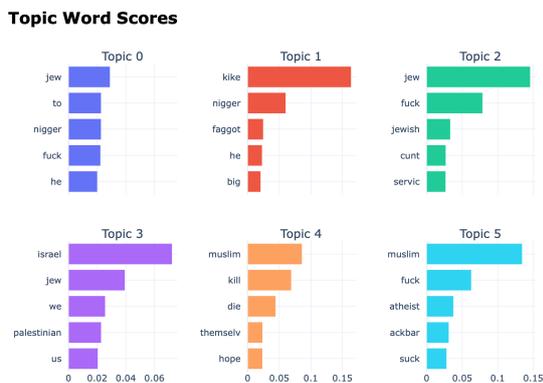


Figure 6: Topics' keywords in our BERTopic model

9.3 LDA initial implementation

We have implemented LDA following the instructions from <https://radimrehurek.com/gensim/models/ldamodel.html>. We used LDAvis (Sievert and Shirley, 2014) to get an initial idea of the quality of the results (see Figure 8).

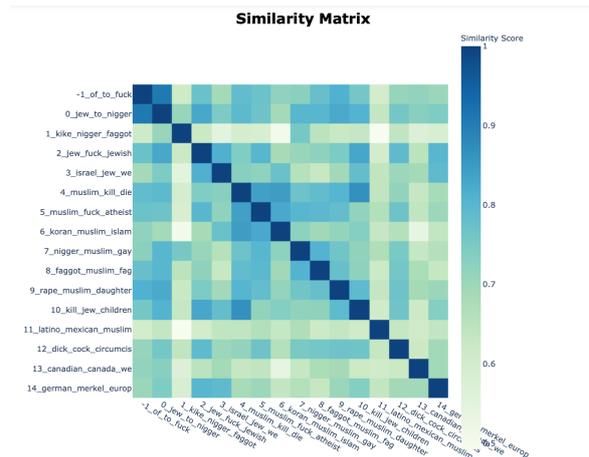


Figure 7: Similarity between topics in our BERTopic model

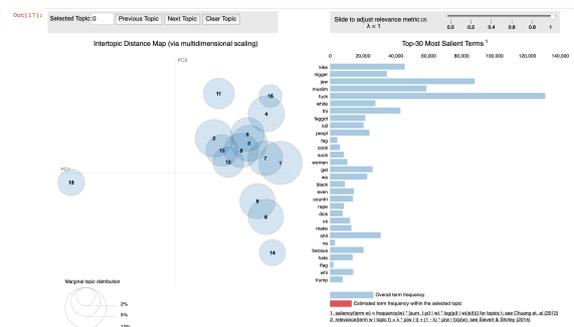


Figure 8: Topics in our LDA model

9.3.1 MultiModalTopicExplorer initial implementation

Figure 5 shows the current status of the MultiModalTopicExplorer layout. We have implemented the most relevant keywords component (see Figure 3 (c)) and the most relevant images component (see Figure 3 (d)). Currently, the visualization shows dummy data. We plan in the future connect these components with information retrieved from BERTopic and LDA models.

References

- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022.
- Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309.
- Jordan Boyd-Graber, David Mimno, and David Newman. 2014. Care and feeding of topic models: Problems, diagnostics, and improvements.

- Handbook of mixed membership models and their applications*, 225255.
- Alex Cao, Keshav K Chintamani, Abhilash K Pandya, and R Darin Ellis. 2009. Nasa tlx: Software for assessing subjective mental workload. *Behavior research methods*, 41(1):113–117.
- A. J. Chaney and David M. Blei. 2012. Visualizing topic models. In *ICWSM*.
- Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, pages 288–296.
- Jason Chuang, Christopher D Manning, and Jeffrey Heer. 2012. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the international working conference on advanced visual interfaces*, pages 74–77. ACM.
- Wenwen Dou, Xiaoyu Wang, Remco Chang, and William Ribarsky. 2011. Paralleltopics: A probabilistic approach to exploring document collections. In *2011 IEEE conference on visual analytics science and technology (VAST)*, pages 231–240. IEEE.
- Wenwen Dou, Li Yu, Xiaoyu Wang, Zhiqiang Ma, and William Ribarsky. 2013. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011.
- Mennatallah El-Assady, Fabian Sperrle, Rita Sevastjanova, Michael Sedlmair, and Daniel Keim. 2018. Ltma: Layered topic matching for the comparative exploration, evaluation, and refinement of topic modeling results. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pages 1–10. IEEE.
- Felipe González, Yihan Yu, Andrea Figueroa, Claudia López, and Cecilia Aragon. 2019. Global reactions to the cambridge analytica scandal: A cross-language social media study. In *Companion Proceedings of the 2019 world wide web conference*, pages 799–806.
- Maarten Grootendorst. 2020. [Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics](#).
- Enamul Hoque and Giuseppe Carenini. 2014. Convis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum*, volume 33, pages 221–230. Wiley Online Library.
- Enamul Hoque and Giuseppe Carenini. 2015. Convisit: Interactive topic modeling for exploring asynchronous online conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 169–180. ACM.
- Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211.
- Pooja Kherwa and Poonam Bansal. 2019. Topic modeling: A comprehensive review.
- Hannah Kim, Barry Drake, Alex Endert, and Haesun Park. 2020. Architext: Interactive hierarchical topic modeling. *IEEE transactions on visualization and computer graphics*.
- Jey Han Lau, David Newman, and Timothy Baldwin. 2014. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 530–539.
- Hanseung Lee, Jaeyeon Kihm, Jaegul Choo, John Stasko, and Haesun Park. 2012. ivisclustering: An interactive visual document clustering via topic modeling. In *Computer graphics forum*, volume 31, pages 1155–1164. Wiley Online Library.
- Tak Yeon Lee, Alison Smith, Kevin Seppi, Niklas Elmqvist, Jordan Boyd-Graber, and Leah Findlater. 2017. The human touch: How non-expert users perceive, interpret, and fix topic models. *International Journal of Human-Computer Studies*, 105:28–42.
- Elijah Meeks and Scott B Weingart. 2012. The digital humanities contribution to topic modeling. *Journal of Digital Humanities*, 2(1):1–6.
- David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models.

- In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 262–272.
- Alexandros Mittos, Savvas Zannettou, Jeremy Blackburn, and Emiliano De Cristofaro. 2020. “and we will fight for our race!” a measurement study of genetic testing conversations on reddit and 4chan. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 452–463.
- Jaimie Murdock and Colin Allen. 2015. Visualization techniques for topic model checking. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Antonios Pappasavva, Savvas Zannettou, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. 2020. Raiders of the lost kek: 3.5 years of augmented 4chan posts from the politically incorrect board. *14th International AAAI Conference On Web And Social Media (ICWSM) 2020*.
- Jessica Peter, Steve Szigeti, Ana Jofre, and Sara Diamond. 2015. Topicks: Visualizing complex topic models for user comprehension. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 207–208. IEEE.
- Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, and Xindong Wu. 2020. Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering*.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*.
- Carson Sievert and Kenneth Shirley. 2014. Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70.
- Alison Smith, Timothy Hawes, and Meredith Myers. 2014. Hierarchy: Visualization for hierarchical topic models. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 71–78.
- Alison Smith, Tak Yeon Lee, Forough Poursabzi-Sangdeh, Jordan Boyd-Graber, Niklas Elmqvist, and Leah Findlater. 2017. Evaluating visual representations for topic understanding and their effects on manually generated topic labels. *Transactions of the Association for Computational Linguistics*, 5:1–16.
- Alison Smith, Sana Malik, and Ben Shneiderman. 2015. Visual analysis of topical evolution in unstructured text: Design and evaluation of topicflow. In *Applications of Social Media and Social Network Analysis*, pages 159–175. Springer.
- Matt Taddy. 2012. On estimation and selection for topic models. In *Artificial Intelligence and Statistics*, pages 1184–1193.
- Yi Yang, Quanming Yao, and Huamin Qu. 2017. Vistopic: A visual analytics system for making sense of large document collections using hierarchical topic modeling. *Visual Informatics*, 1(1):40–47.