

MLQDA

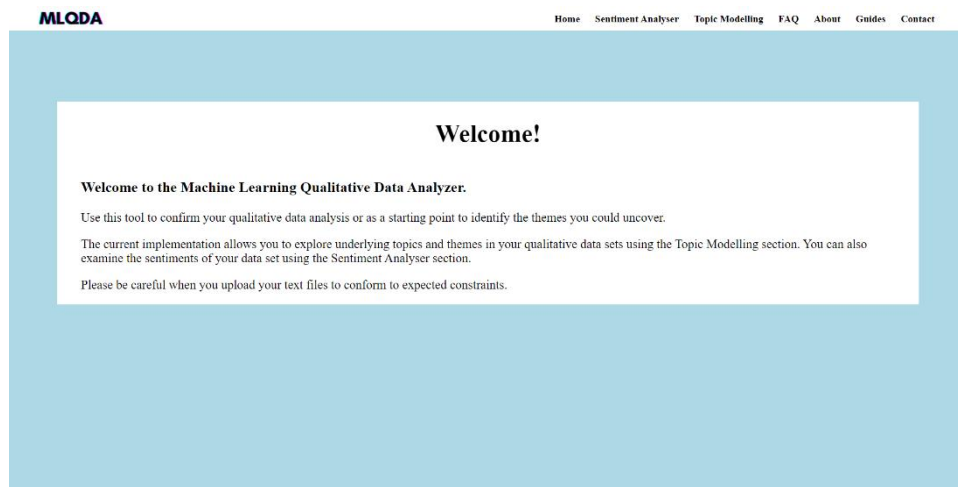
# Machine Learning for Qualitative Data Analysis

User Guide

|                                     |    |
|-------------------------------------|----|
| Home.....                           | 3  |
| FAQ.....                            | 4  |
| About .....                         | 5  |
| Guides .....                        | 6  |
| Contact.....                        | 7  |
| Topic Modelling .....               | 8  |
| Topic Modelling starting page ..... | 8  |
| Topic Modelling landing page.....   | 8  |
| Topic Modelling results.....        | 9  |
| Sentiment Analysis.....             | 11 |
| Sentiment Analysis start .....      | 11 |
| Sentiment Analysis finish .....     | 11 |
| Sentiment Analysis results.....     | 12 |

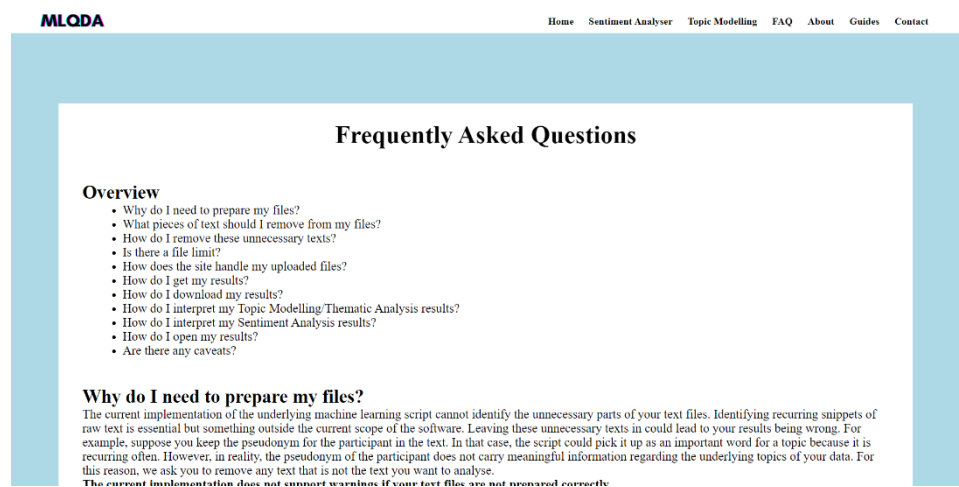
## Home

The home page welcomes the users and highlights some of the most essential information about using the site. The navigation bar allows users to access the rest of the site.



## FAQ

The Frequently Asked Questions section contains explanations regarding the most common issues that may arise while using the site.



The screenshot shows the MLQDA website's 'Frequently Asked Questions' page. The page has a light blue header with the MLQDA logo on the left and a navigation menu on the right containing links for Home, Sentiment Analyser, Topic Modelling, FAQ, About, Guides, and Contact. The main content area is white and features a large blue border. At the top of this area is the title 'Frequently Asked Questions' in bold. Below the title is a section titled 'Overview' which contains a bulleted list of ten common questions. The first question is 'Why do I need to prepare my files?'. Below the list is a section titled 'Why do I need to prepare my files?' which provides a detailed explanation of why text preparation is necessary for the machine learning script. The text explains that the script cannot identify unnecessary parts of text files and that leaving such text in could lead to incorrect results. It gives an example of a participant's pseudonym being incorrectly identified as an important word for a topic. It concludes by asking the user to remove any text that is not the data they want to analyse. At the very bottom of the page, there is a small line of italicized text: 'The current implementation does not support warnings if your text files are not prepared correctly'.

**MLQDA** Home Sentiment Analyser Topic Modelling FAQ About Guides Contact

### Frequently Asked Questions

#### Overview

- Why do I need to prepare my files?
- What pieces of text should I remove from my files?
- How do I remove these unnecessary texts?
- Is there a file limit?
- How does the site handle my uploaded files?
- How do I get my results?
- How do I download my results?
- How do I interpret my Topic Modelling/Thematic Analysis results?
- How do I interpret my Sentiment Analysis results?
- How do I open my results?
- Are there any caveats?

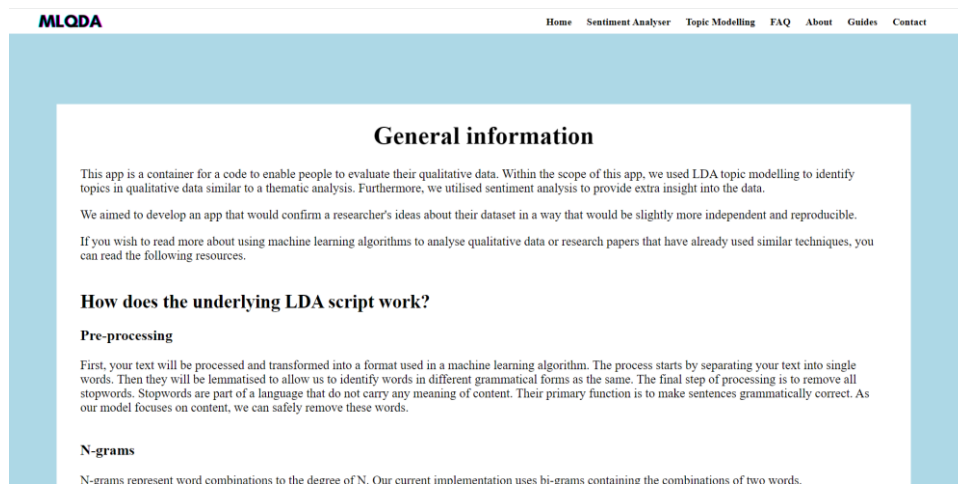
#### Why do I need to prepare my files?

The current implementation of the underlying machine learning script cannot identify the unnecessary parts of your text files. Identifying recurring snippets of raw text is essential but something outside the current scope of the software. Leaving these unnecessary texts in could lead to your results being wrong. For example, suppose you keep the pseudonym for the participant in the text. In that case, the script could pick it up as an important word for a topic because it is recurring often. However, in reality, the pseudonym of the participant does not carry meaningful information regarding the underlying topics of your data. For this reason, we ask you to remove any text that is not the text you want to analyse.

*The current implementation does not support warnings if your text files are not prepared correctly*

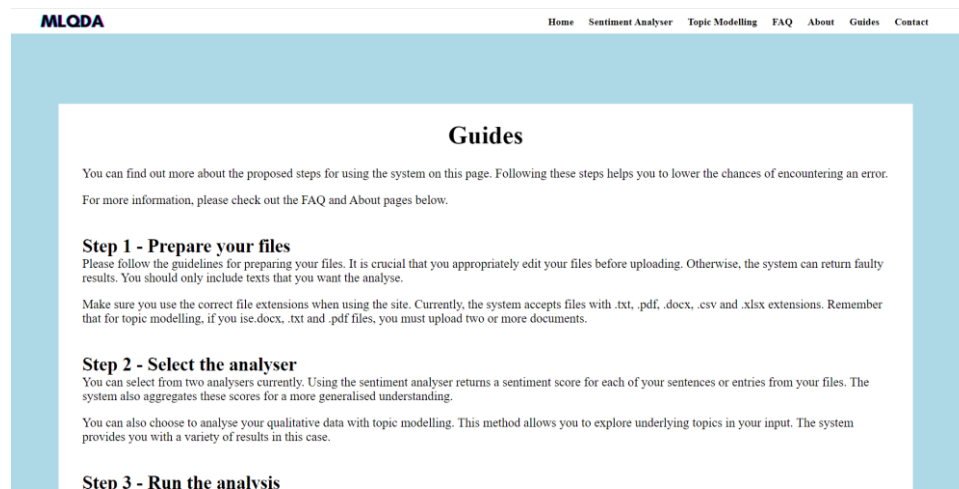
## About

The About page explains the how the underlying machine learning scripts work for anyone who is interested in the process.



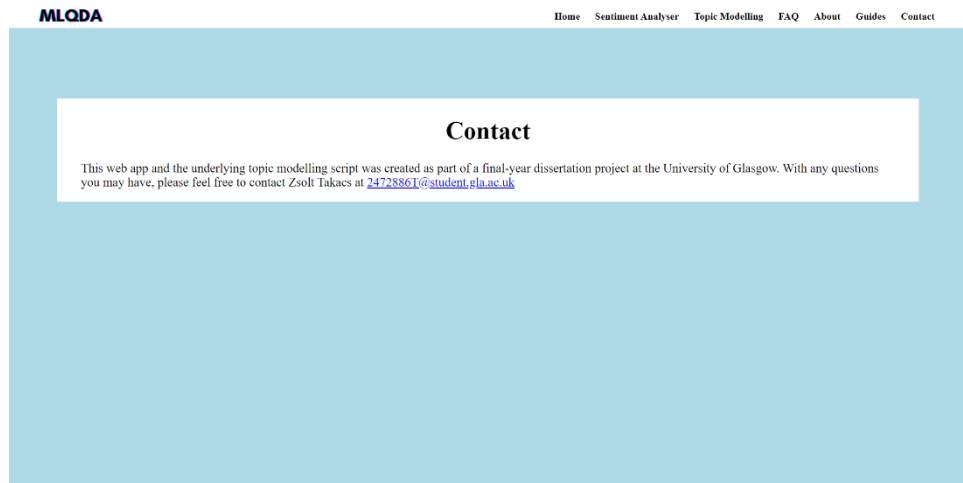
## Guides

The guides section walks through the user on how to run an analysis from start to finish. It also explains how to interpret the results.



## Contact

The contact page contains necessary information to get in touch with the developers of the site.



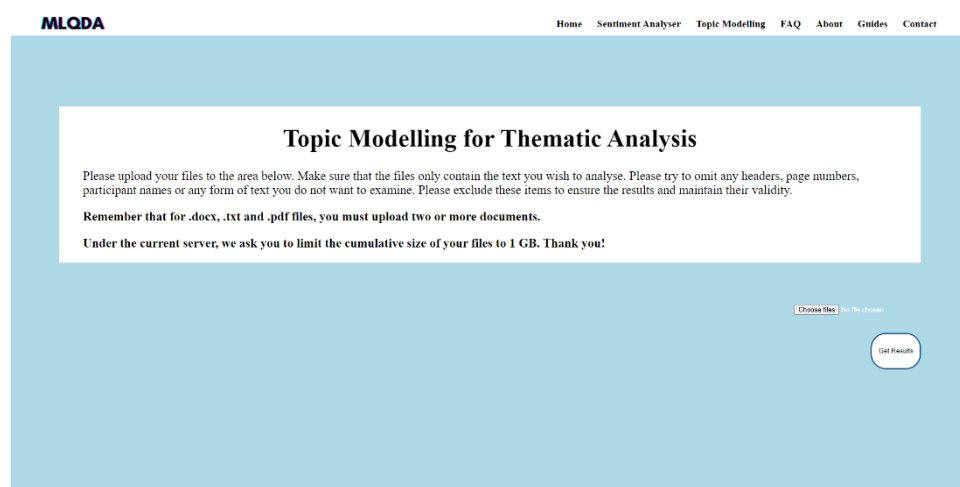
## Topic Modelling

As part of the Topic Modelling analysis, users can upload their files to the starting page and obtain their results on a landing page.

### Topic Modelling starting page

This page highlight some of the most important prerequisites for the underlying script. Users can upload multiple files. The types of files are automatically limited to the ones that are accepted by the system. The following workflow could be used to run an analysis.

1. Click on “Choose files”
2. Select desired files and click OK
3. Click on “Get Results”
4. Confirm pop-up message by clicking OK
5. Wait until system loads the landing page



### Topic Modelling landing page

The landing page contains high-level information about the topic modelling results as it displays the topics and their connected high probability contribution words. This page hosts two main functions:

1. Clicking “Download Results” will download a zip folder with all the in-depth result files.
2. Clicking “Delete Results” will remove every reference to the uploaded files completely and redirects the user to the home page.



MLODA
Home Sentiment Analyser Topic Modelling FAQ About Guides Contact

## Topic Modelling Results

The script identified 3 topics in your files.

Ideally, these words describe distinct topics in your documents. You will need to deduct these topics based on the top terms. Next to the words, you can also see their contribution as a percentage. Within the results, you can see highlighted sentences for each topic.

In a highlighted pdf, parts of your texts are highlighted based on the results of the machine learning script. Next to the snippets are the terms the script identified as a word of high importance for that topic.

You can check out the FAQ or the About section if you want to know more about the process or how to interpret your result.

**Please remember to click the Delete Files button after you download your results. If you forget to do so, the result file can potentially stay on the server for up to one hour before automatic deletion. However, you will not be able to access these results again after leaving the page.**

Download Results
Delete Results

### Topic Word Contribution

#### Topic 1

|                 |              |                |                  |                |
|-----------------|--------------|----------------|------------------|----------------|
| use: 0.02       | people: 0.02 | migrant: 0.016 | secretary: 0.012 | centre: 0.012  |
| language: 0.012 | house: 0.012 | asylum: 0.01   | minister: 0.01   | invasion: 0.01 |

#### Topic 2

|              |               |              |               |               |
|--------------|---------------|--------------|---------------|---------------|
| album: 0.038 | top: 0.033    | chart: 0.022 | single: 0.019 | number: 0.017 |
| sale: 0.014  | record: 0.014 | vinyl: 0.011 | first: 0.011  | sell: 0.009   |









#### Topic 3

|               |                 |                |                |            |
|---------------|-----------------|----------------|----------------|------------|
| school: 0.068 | rise: 0.033     | teacher: 0.026 | cost: 0.023    | cut: 0.023 |
| pay: 0.023    | increase: 0.017 | budget: 0.017  | funding: 0.014 | add: 0.014 |

## Topic Modelling results

The following files can be found in the compressed folder containing the results for a topic modelling analysis. Each file starts with the unique ID for the analysis instance.

1. A PNG file containing visual representation of the probability contributions for each word for each topic.
2. A JSON file mapping the probability contributions to each word for each topic.
3. A CSV file, containing a wide formatted table that presents a sentence-by-sentence breakdown of the entire corpus and shows which topic words are present in them.
4. As many PDF highlights as identified topics which show a long-formatted version of the previous table. Each row is a sentence, and a list of topic words present in the given sentence. If there are any, then it is highlighted.
5. Two text documents containing the representation of the corpus and id2word instances. These could be used to run and create an alternative model using the Gensim Library.

| Name   | Type                       |
|--|----------------------------|
|  105_result_visualisation | PNG File                   |
|  105_results              | JSON File                  |
|  105_results              | Microsoft Excel Comma Sepa |
|  105_topic1_highlights    | Chrome HTML Document       |
|  105_topic2_highlights    | Chrome HTML Document       |
|  105_topic3_highlights    | Chrome HTML Document       |
|  105corpus                | Text Document              |
|  105id2word               | Text Document              |

The most important results can be found in the highlighted PDF files. Each identified topic has a corresponding file, that highlights section of the data that is related to the topic. First, a list of the words and their probability contributions are presented. Then, each file is converted into a table, where the right-hand side contains a sentence from the data set. The left-hand side contains any of the topic words present in the given sentence. If there are any present, the sentence gets highlighted.

## Topic modelling results

### 1 Most important words for topic 3

- people--0.032
- use--0.032
- migrant--0.025
- centre--0.018
- language--0.018
- home--0.018
- secretary--0.018
- immigration--0.015
- invasion--0.015
- minister--0.015

Sentences that contain any of these words are highlighted below.

### 2 Highlights from text 1

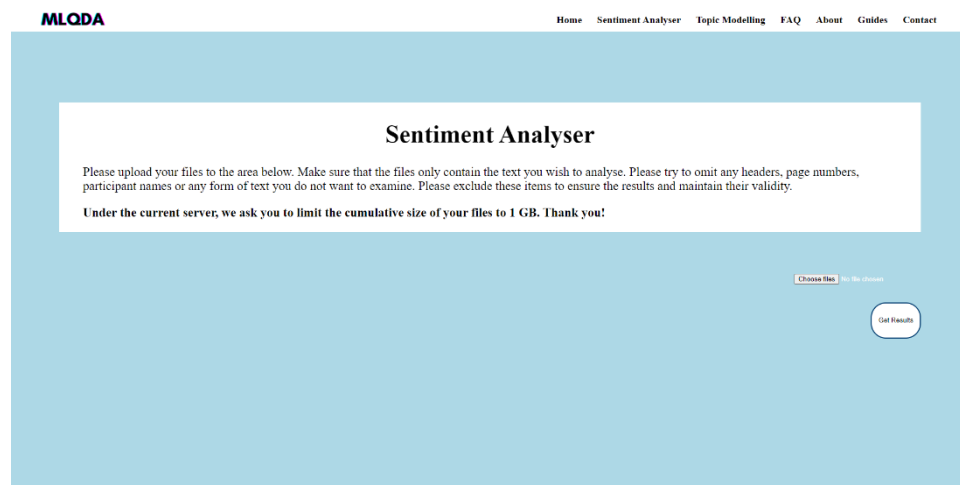
| Topic Words  | Sentence  |
|--|---|
| migrant, language, home, secretary, invasion, minister | Immigration minister Robert Jenrick has said politicians must be careful with language after the home secretary claimed southern England was facing an 'invasion' of illegal migrants           |
| use  | Mr Jenrick told BBC Radio 4's Today: 'It's not a phrase I've used.' But he said his boss Suella Braverman was right to be straight with the public about the 'sheer scale' of illegal migration |
|  | He said 'more radical options' were needed to deal with the 'crisis'  |
| centre   | Ms Braverman is under pressure to deal with overcrowding at the Manston asylum processing centre in Kent, which has reportedly led to outbreaks of  |

# Sentiment Analysis

## Sentiment Analysis start

This page highlight some of the most important prerequisites for the underlying script. Users can upload multiple files. The types of files are automatically limited to the ones that are accepted by the system. The following workflow could be used to run an analysis.

6. Click on “Choose files”
7. Select desired files and click OK
8. Click on “Get Results”
9. Confirm pop-up message by clicking OK
10. Wait until system loads the landing page



## Sentiment Analysis finish

The landing page contains high-level information about how to interpret the Sentiment Analysis results. Furthermore, this page hosts two main functions:

3. Clicking “Download Results” will download a zip folder with all the in-depth result files.
4. Clicking “Delete Results” will remove every reference to the uploaded files completely and redirects the user to the home page.



## Sentiment Analysis results

The compressed folder that can be downloaded from the sentiment analysis results landing page contains two files. Both files start with the unique ID for the analysis instance.

1. A CSV file containing broken down data set. Each row contains a sentence or an entry of the original files. Next to them the compound sentiment score is presented.
2. A PDF file containing the same table with additional information on how to interpret the results.

| Name   | Type                           |
|--|--------------------------------|
|  103_results            | Microsoft Excel Comma Separ... |
|  103_sentiment_analysis | Chrome HTML Document           |

## Sentiment Analysis results

### 1 Sentiment Analysis

The sentiment scores are present on the left-hand side of the table. The analyser uses compound sentiment scores calculated from positive, negative and neutral sentiment scores. The displayed scores range from -1 to 1, which you can interpret as percentages. For example, you can interpret a score of +0.25 as 25% positive sentiment. When compiling the document, the system calculates the sentiment score for every sentence.

### 2 Sentiments from text 1

| Sentiment | Sentence  |
|-----------|---|
| -0.4588   | Immigration minister Robert Jenrick has said politicians must be careful with language after the home secretary claimed southern England was facing an 'invasion' of illegal migrants   |
| -0.5499   | Mr Jenrick told BBC Radio 4's Today: 'It's not a phrase I've used.' But he said his boss Suella Braverman was right to be straight with the public about the 'sheer scale' of illegal migration   |
| 0.0       | He said 'more radical options' were needed to deal with the 'crisis'  |
| -0.743    | Ms Braverman is under pressure to deal with overcrowding at the Manston asylum processing centre in Kent, which has reportedly led to outbreaks of disease and violence   |
| -0.8738   | Braverman was warned about migrant overcrowding Why are there problems at Manston migrant centre? How many migrants cross the Channel in small boats? In a Commons statement, she rejected claims she had deliberately blocked the use of hotels to ease the pressure on Manston and promised to fix the 'hopelessly lax' asylum system |
| 0.0       | Some 4,000 people are being held at the former RAF base which is only designed to accommodate 1,600 people on a temporary basis   |
| 0.0       | Hundreds of people were moved there on Sunday after a man threw firebombs at a separate immigration centre in Dover   |
| -0.6486   | He was later found dead nearby  |
| -0.154    | A report from a visit by the chief inspector of prisons in July found the centre had consid-  |