MLQDA

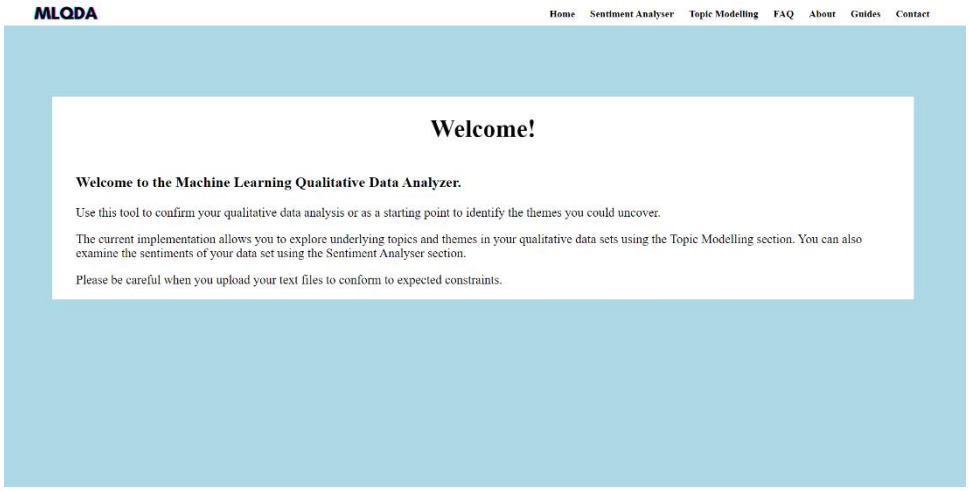
Machine Learning for Qualitative Data Analysis

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.

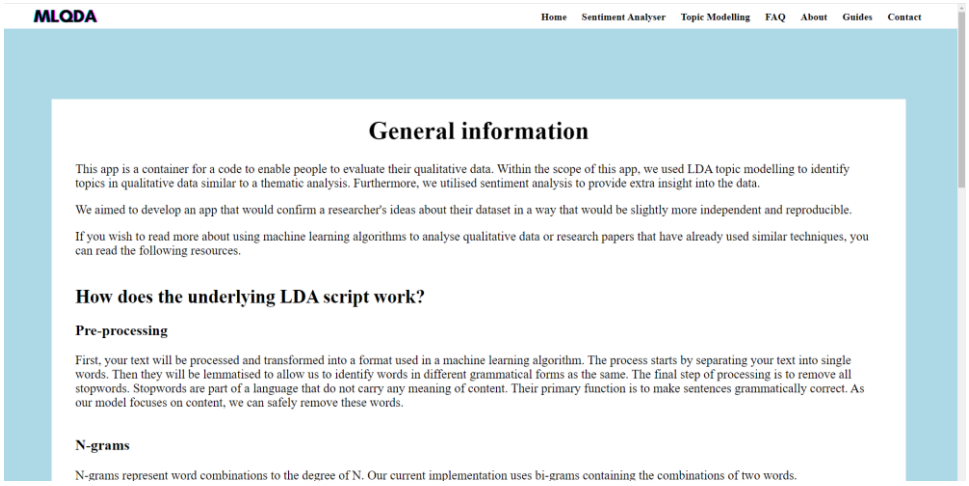


FAQ

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

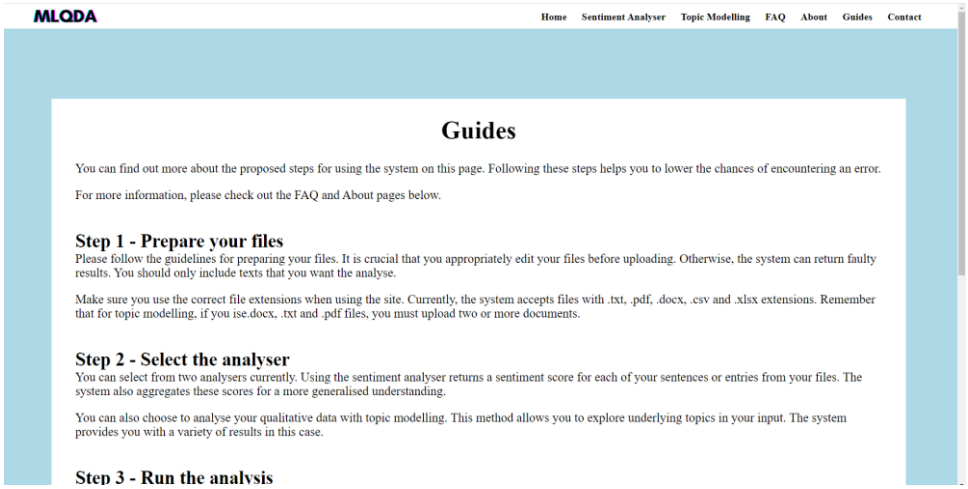
About

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



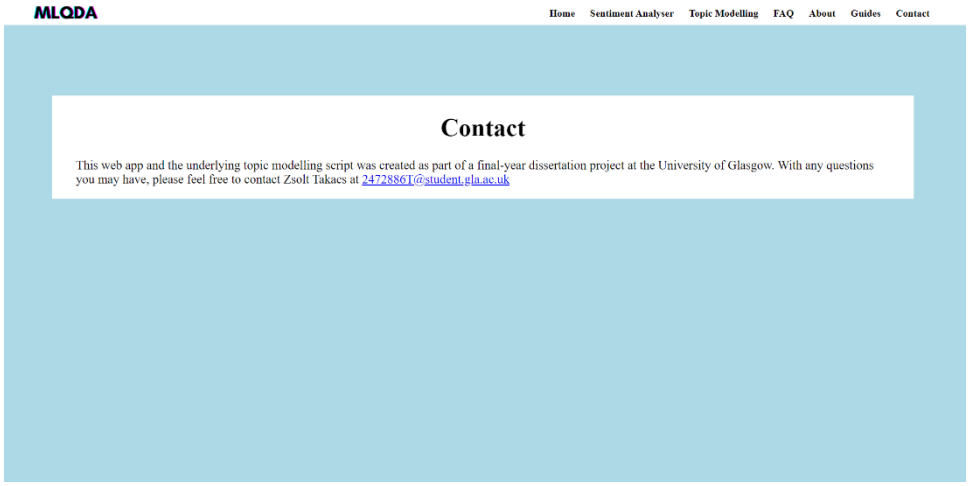
Guides

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 int ouch with the developers of the site.



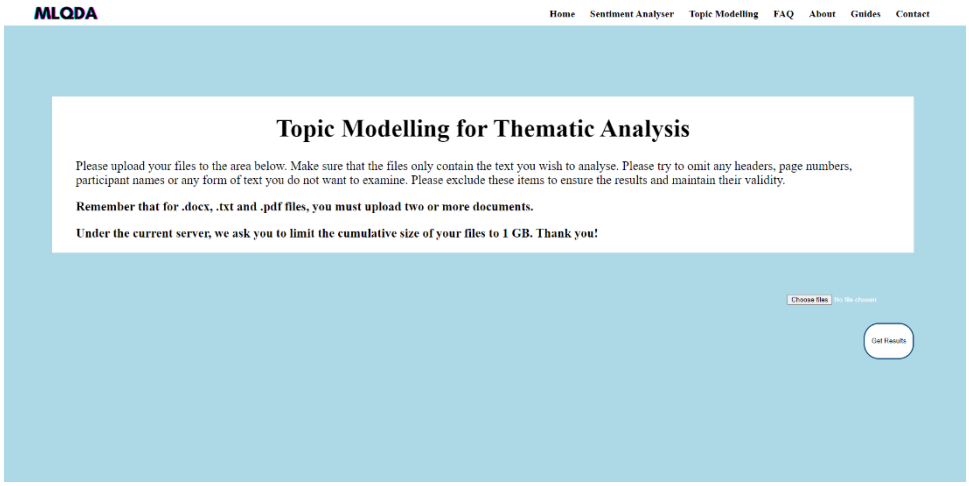
Topic Modelling

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.

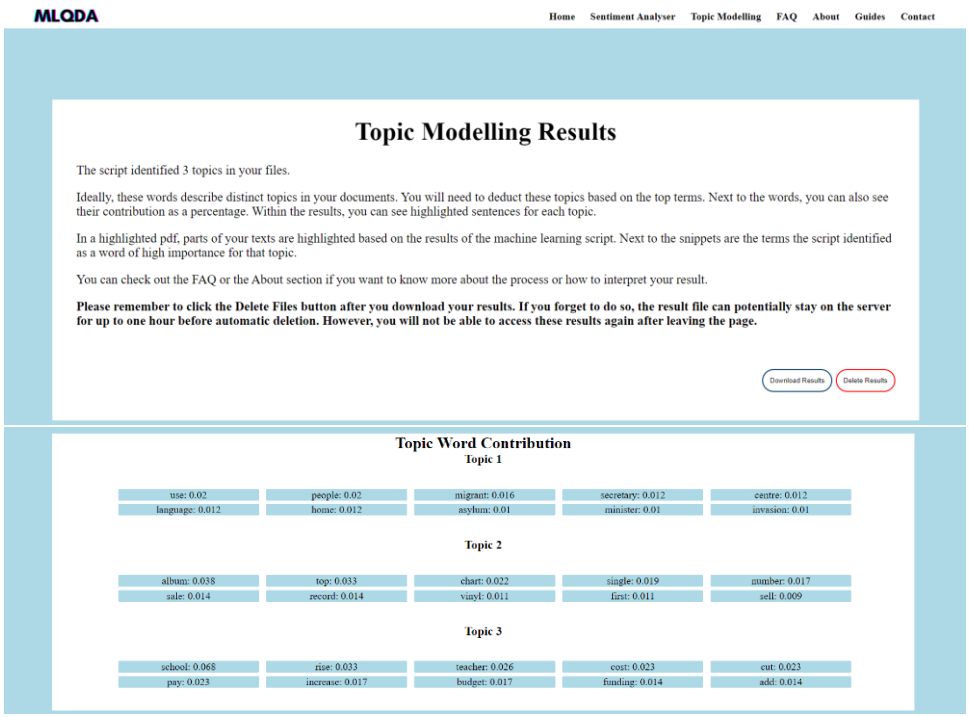
1. Click on “Choose files”
2. Select desired files and click OK
3. Click on “Get Results”
4. Confirm pop-up message by clicking OK
5. Wait until system loads the landing page



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:

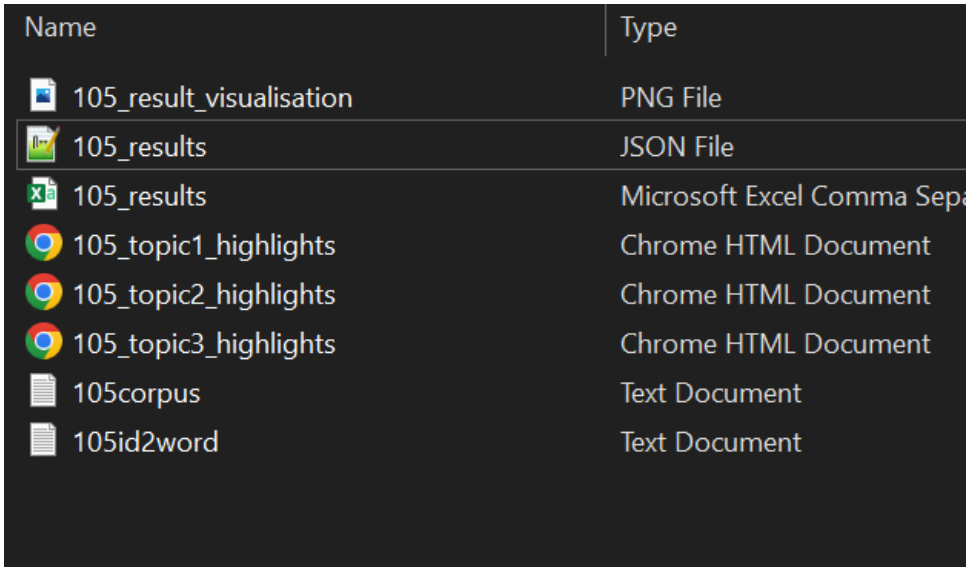
1. Clicking “Download Results” will download a zip folder with all the in-depth result files.
2. Clicking “Delete Results” will remove every reference to the uploaded files completely and redirects the user to the home page.



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.



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 reuslts

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

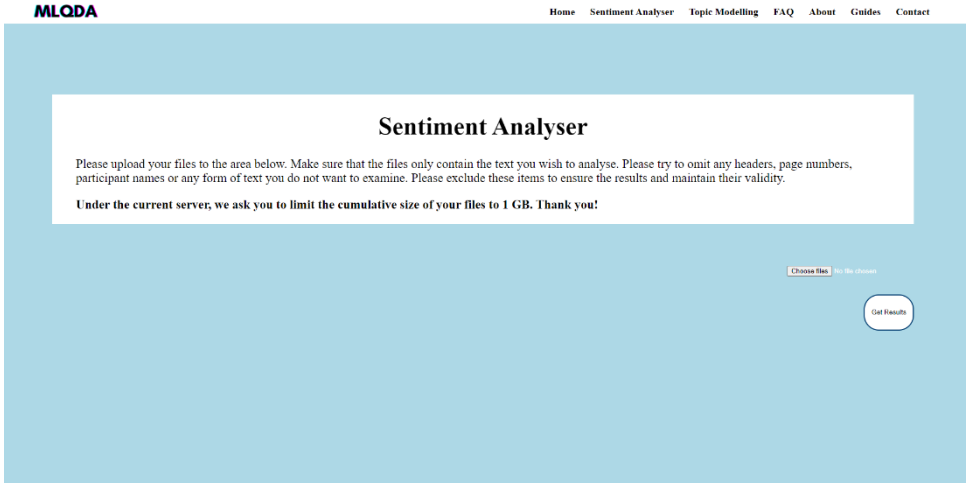
Topic Words	Sentence
migrant, language, home, secretary, invasion, minister	Immigration minister Robert Jenrick has said politicians must be careful with language after the home secretary claimed southern England was facing an 'invasion' of illegal migrants
use	Mr Jenrick told BBC Radio 4's Today: 'It's 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.

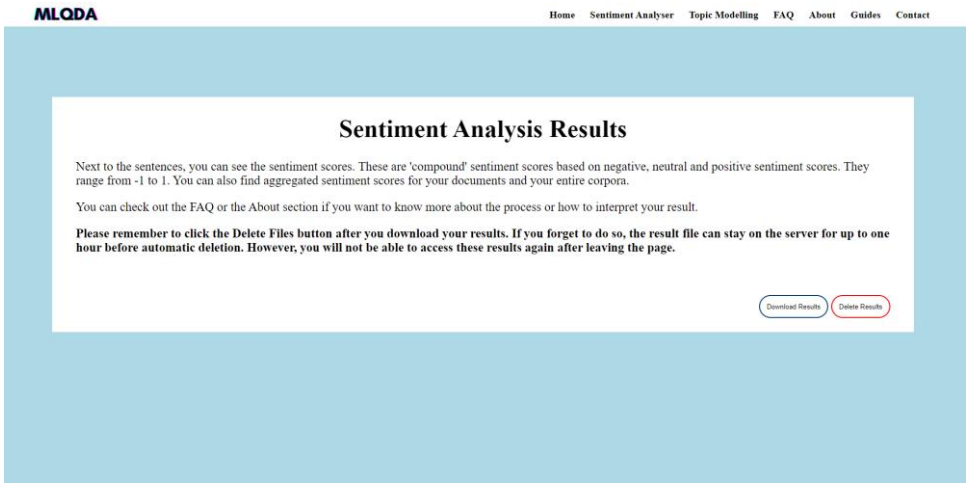
6. Click on “Choose files”
7. Select desired files and click OK
8. Click on “Get Results”
9. Confirm pop-up message by clicking OK
10. Wait until system loads the landing page



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.

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-