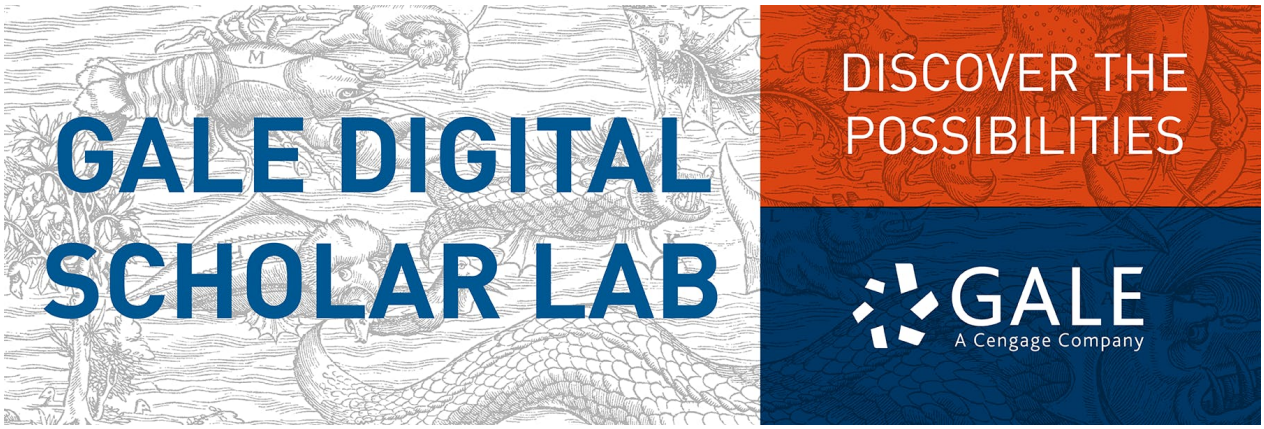


Text Analysis & Data Visualistations:

Gale Digital Scholar Lab

Topic Modelling exercise

Student Training and Support
Phone (07) 334 64312
Email askus@library.uq.edu.au
Web http://www.library.uq.edu.au/library-services/training

Service Points

St Lucia: Main desk of the Central, ARMUS and DHESL libraries
Hospitals Main desk of the PACE and Herston libraries
Gatton: Level 2, UQ Gatton Library

Library services provide the student I.T. Helpdesk service in the UQ Library. They can assist with general enquiries and IT support. This includes computing help and training for UQ students in: Study Management Applications like my.UQ and Learn.UQ (Blackboard), Microsoft Office and I.T. fundamentals like file management, printing and laptop setup.

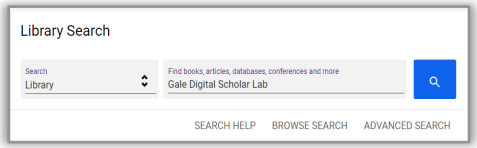
Table of Contents

Digital Scholar Lab – Text analysis & Data Visualisations

Gale Digital Scholar Lab is a subscription web platform, for analysing textual data and creating a range of data visualisations. Access the platform via UQ Library Search to use this subscription product for free. You will need to login using a Google or Microsoft account to start analysing and creating your own visualisations.

Exercise 1. ***Access & login***

1. Go to https://www.library.uq.edu.au/
   Note The Digital Scholar Lab works best in Firefox
2. Search for Gale Digital Scholar Lab

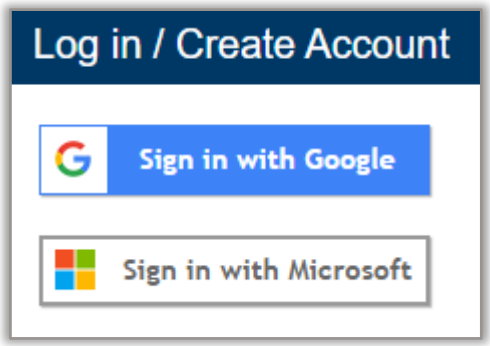


3. Click **Available Online** to open the Digital Scholar Lab platform
   Note If presented with the UQ log in, use your UQ credentials

4. Click **Log in / Create Account**

A dialogue box will open asking you to log in with either Google or Microsoft.
Note All UQ users will be able to log in with Microsoft.

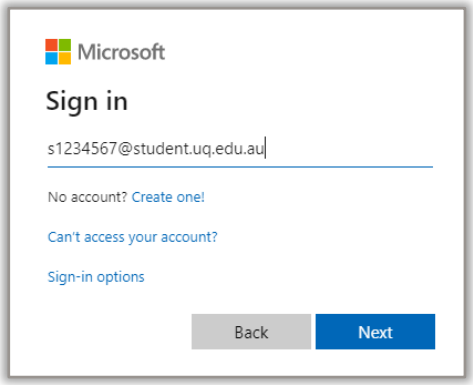
5. Click **Sign in with Microsoft**



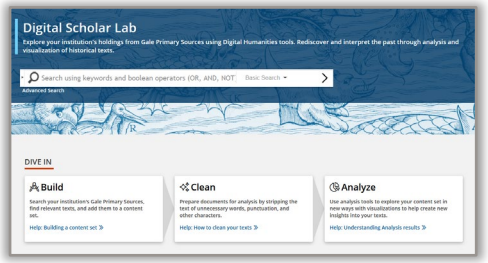
6. At the Mircosoft sign in, enter your UQ email address and click **Next**

   e.g. uqusername@student.uq.edu.au

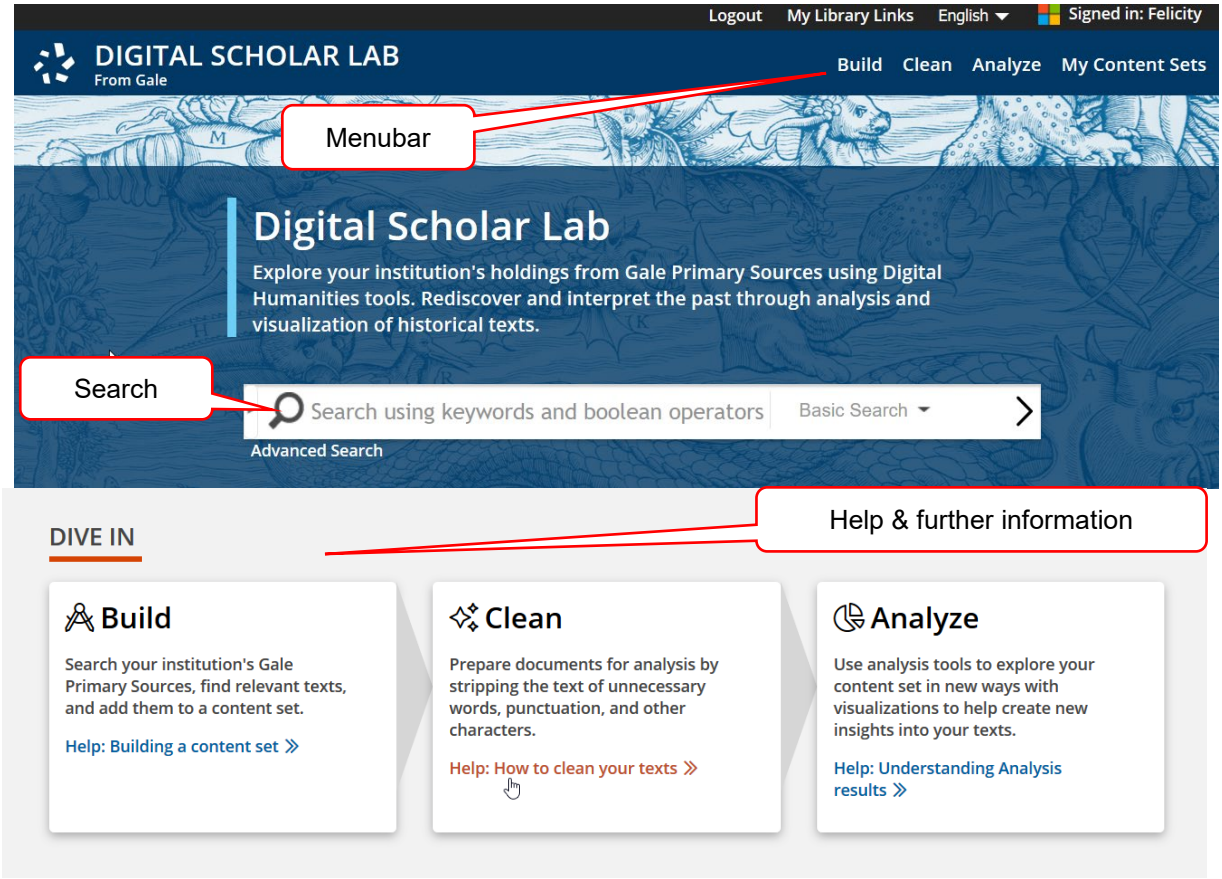
Note You will be redirected to a UQ log in screen. Log in with your UQ credentials



7. You will be redirected and logged in to Gale Digital Scholar Lab



The Digital Scholar Lab interface



Creating a Data Visualisation

The process in three steps

1.	**Build** your content/data set	Search across Gale Archives or upload your own textual data
2.	**Clean** your data	Use the default cleaning set-up, or create your own
3.	**Analyse** your data	Use the built-in tools to analyse your data and create visualisations

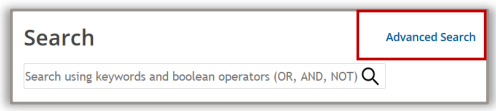
Building Your Content Set

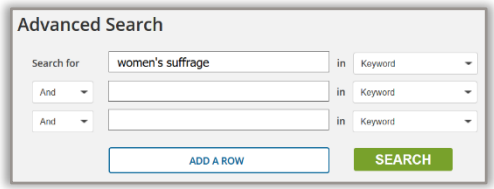
The first step in text and data mining is to create your Content Set, sometimes known as a corpus or data set. A Content Set is simply a collection of documents you wish to analyse. Create Content Sets by finding documents in UQ's Gale Primary Sources holdings or by uploading your own documents.

Exercise 2. ***Creating a new content set***

1. Click the **Build** link in the menubar

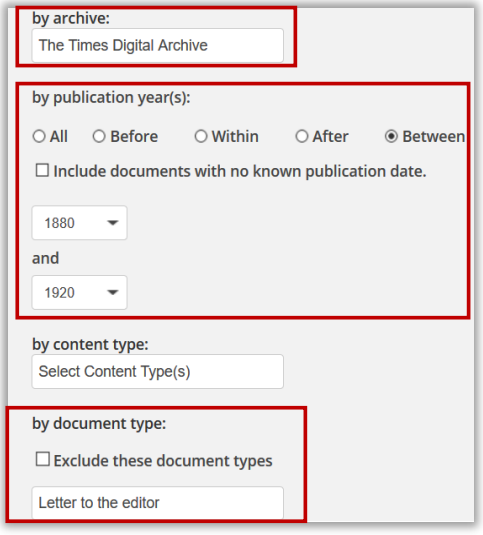
2. Click **Advanced Search**



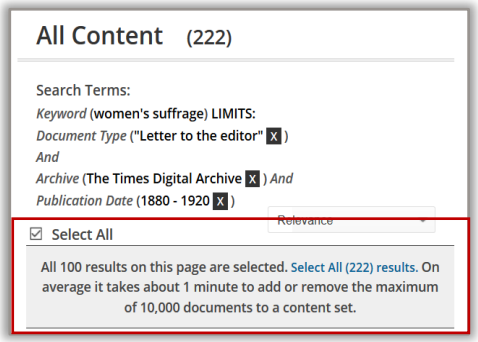


3. Search for **women's suffrage** in Keyword and apply the following options:
   a) **by archive: The Times Digital Archive**
   b) **by publication date: between 1880 – 1920**
   c) **by document type: Letter to the editor**

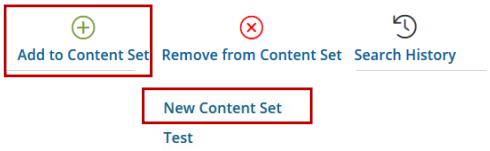
4. Click **Search**



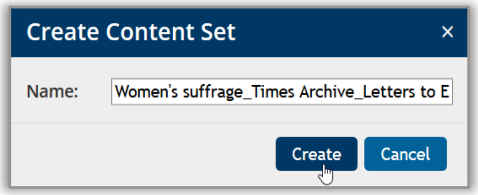
5. From the results screen, tick the box to **Select All**, and then click the link to **Select All (222) results**



6. Click **Add to Content Set** from the menubar, and then select **New Content Set.**



7. Name the new content set and click **Create**

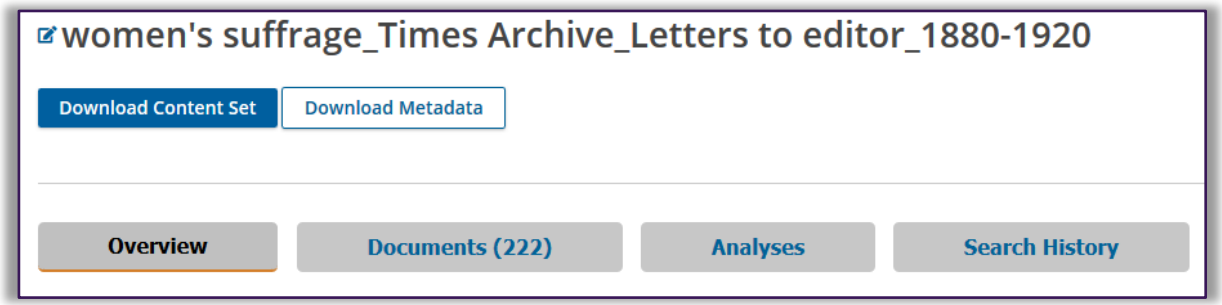


My Content Sets

My Content Sets is where you can download content sets, view any previous analyses you have performed over this content set, view the actual documents in your content set, and view the search history.

1. Click **My Content Sets** on the menubar.

2. Click the content set you just created from the list. An overview of the content set will be displayed.



Cleaning Your Data

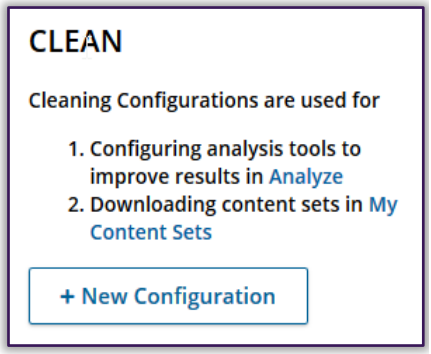
Most text is created and stored so that humans can understand it, and it is not always easy for a computer to process that text. Before you begin a text analysis project, you often need to clean the text to ensure it is in a format that a computer can use (machine readable).

Exercise 3. ***Creating a new cleaning configuration***

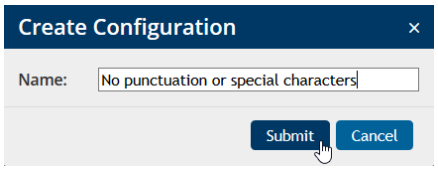
1. Click **Clean** from the menubar to go to the cleaning configuration screen.

   Note It will automatically load the default cleaning configuration.

2. To create a new cleaning configuration, click **+New Configuration.**



3. Give your new cleaning configuration a name and click **Submit**.

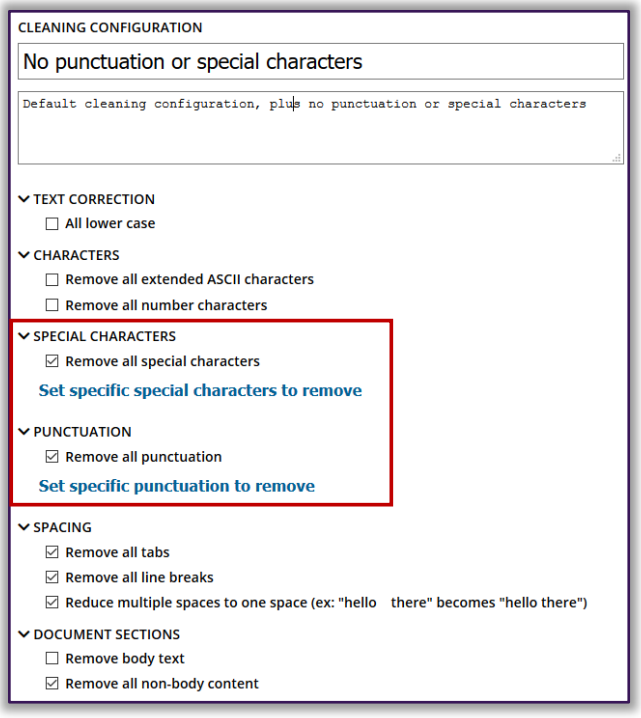


4. Give the new cleaning configuration a description and select options required.

   Note: For this example I have left the default options ticked and added selections to:

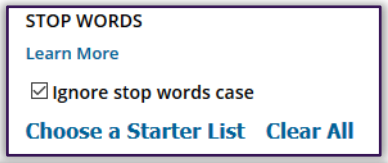
   a) Remove all special characters

   b) Remove all punctuation



5. On the right side of the screen click you will see a section for Stop Words. Stop words are words which are filtered out before or after processing of natural language text, e.g. common words like "the", "and", "at" etc.

   Click **Choose a Starter List**

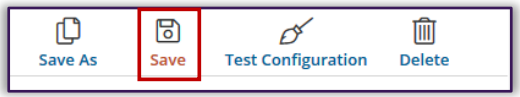


6. Select **English** from the list and then click **Select starter lists**

   A list of common words from the English language will be added to the Stop Words section. These words will be filtered out before analysis.

   Note You can add words to the stop word list by manually editing the list.

7. Click **Save** to save your new cleaning configuration.



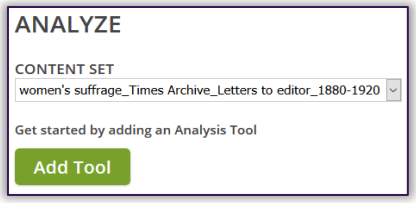
Analysing Your Data

With your Content Set curated and cleaned, you are ready to analyse it. Analysis allows you to take hundreds or thousands of documents and use digital tools to analyse them in ways that would have been too time consuming without the help of computational algorithms.

Exercise 4. ***Adding analysis tools and running an initial analysis***

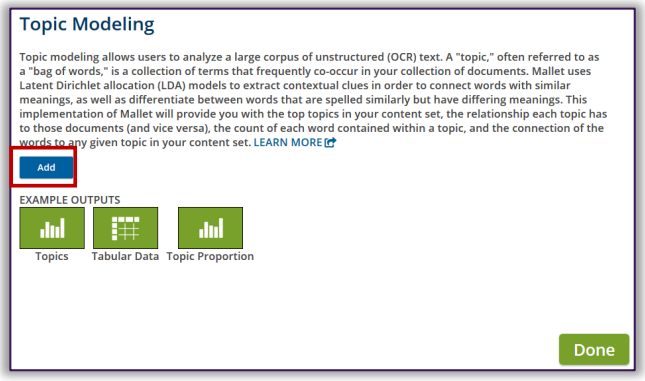
1. Click **Analyze** from the menubar.

2. Ensure the correct content set is selected and then click **Add Tool.**

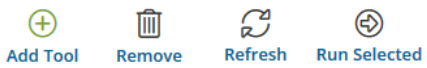


Note For this example we will be using the Topic Modelling tool.

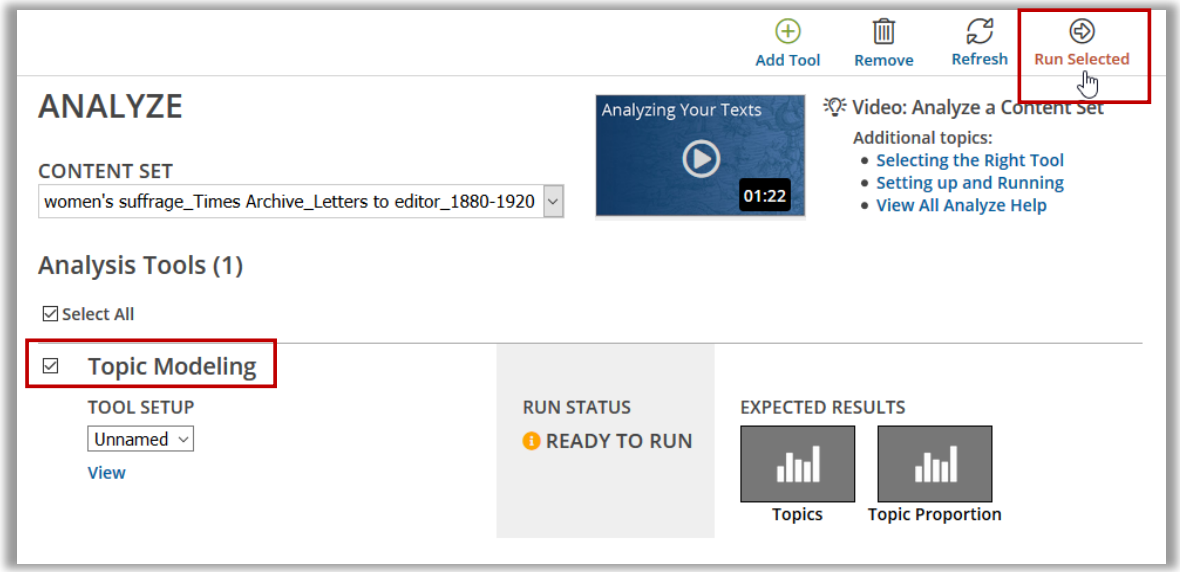
3. Find Topic Modelling in the list and click **Add**.
4. Click **Done** to return to the Analyze screen.



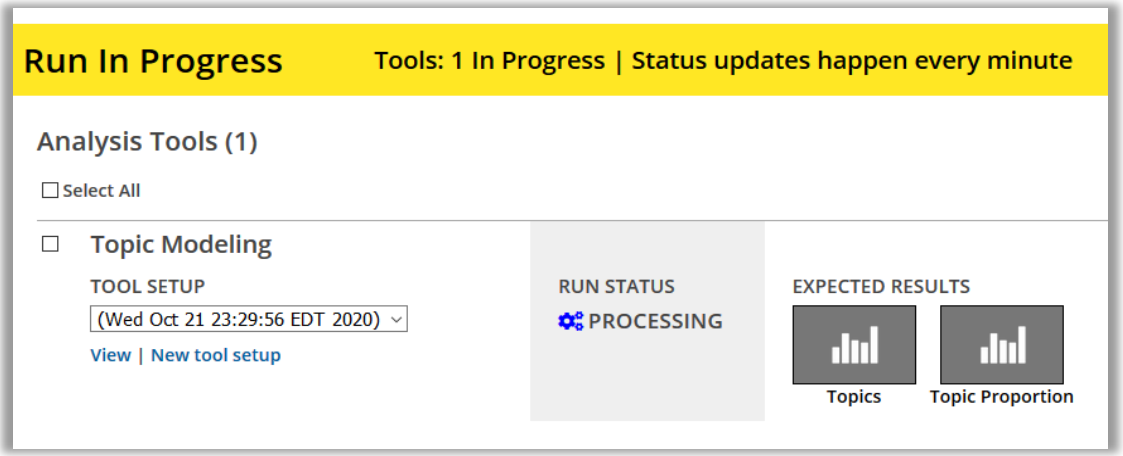
Note To add additional tools later, click **+ Add Tool** (top right of the Analyze screen, below the menubar.



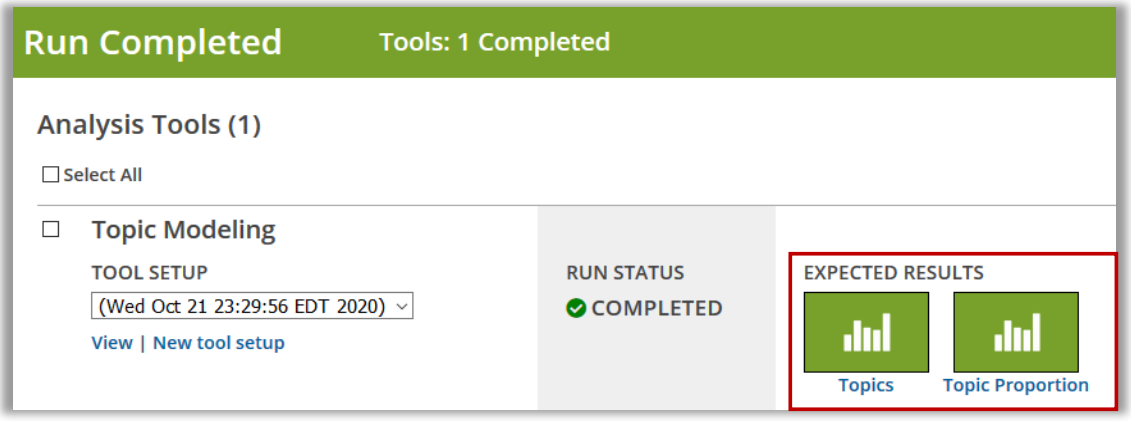
5. Select the **Topic Modelling** tool and click **Run Selected.**



The analysis will run automatically and display a status of **In Progress**.



Once completed the status will show as **Completed.**

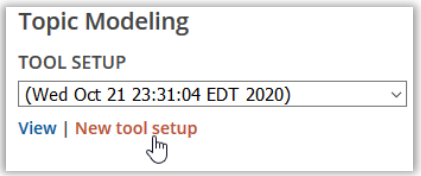
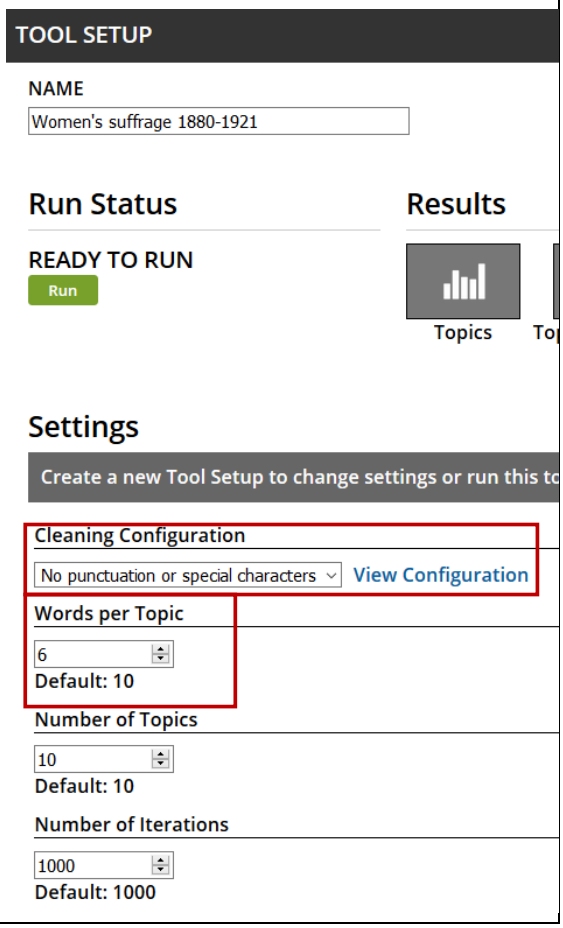


6. Click on the Expected Results to view the analysis and visualisation

Changing the Tool Setup

The initial analysis will use the default cleaning configuration. To apply the new cleaning configuration we created above, and to make other changes to the tool set-up, you will need to create a new tool set-up.

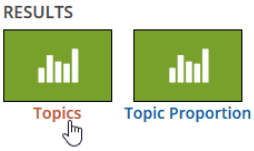
Exercise 5. ***Creating a new tool setup***

1. From the Analyze screen click **New tool setup**	
2. Give your new setup a name. 3. Under **Settings > Cleaning Configuration** select the new cleaning configuration you created above. 4. Adjust the **Words per Topic** to 6 5. Click **Run**	

Exercise 6. ***Editing topic names***

By default the topics in the visualisation are given numbers. To make the visualisation more meaningful you may like to edit the topic names.

6. Click **Topics**

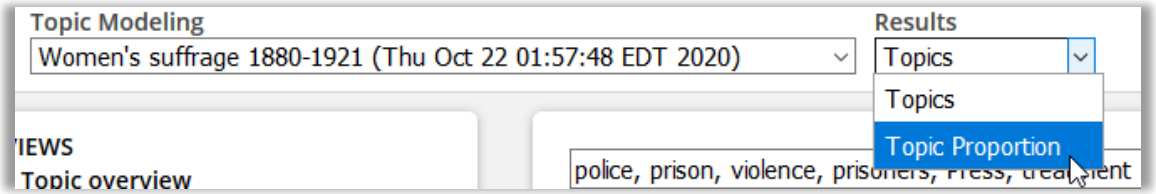


7. Place the cursor in the topic name box, e.g. where is states Topic 0, and delete the default text and add your own.

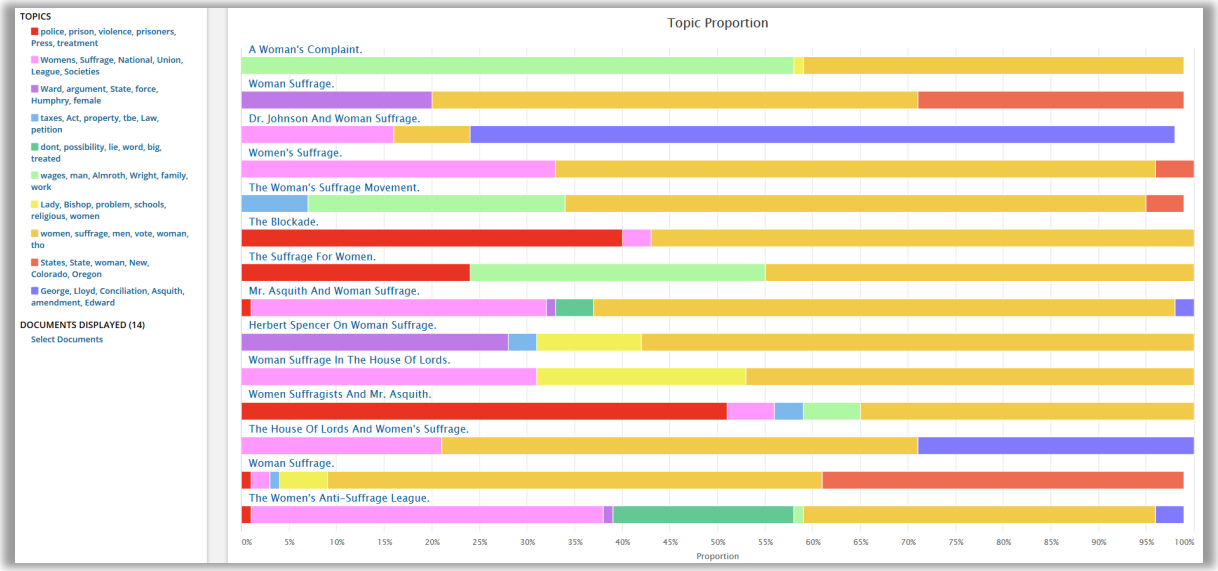
Note I have used the 6 topic terms used to create this topic cluster as the name of the topic.

8. Do the same for each of the 10 topics and click **Save.**

9. Select **Topic Proportion** from the Results dropdown menu.



10. The Topic Proportion visualisation will display.

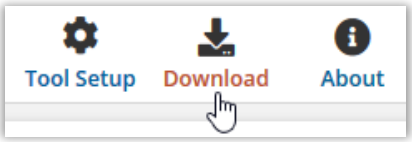


Data and visualisation outputs

Data and visualisations created automatically within the platform can be exported in multiple formats.

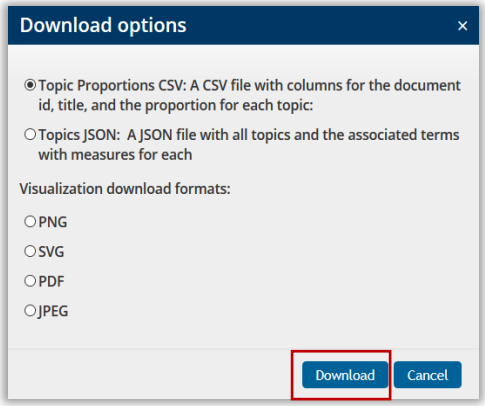
Exercise 7. ***Exporting data and visualisations***

1. In the Topics/Topic Proportion screen click **Download**.



The **Download options** dialogue box will open.

- Data can be downloaded as a CSV or JSON file for further analysis.
- Visualisations can be downloaded as image files in various formats.



2. Make your selection and click **Download**