

DIY Computational Access: The Polytechnic Magazine 1879 - 1960

Jake Bickford
6 July 2022

THE
NATIONAL
ARCHIVES



‘DIY’ but not splendid isolation

| | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|

| | |
|----------|--|
| THE | |
| NATIONAL | |
| ARCHIVES | |

HOME



NEWS.

PUBLISHED

MONTHLY.

VOL. I.—No. 1.

JUNE, 1879.

PRICE 2D.

INTRODUCTION.

INTRODUCTION! Bless you, he wants no introduction. I have come across a good many cheeky young chaps in my time, but anything like this one I never saw. I have had no peace since he first turned up about three weeks ago, and demanded a printing habitation and a name. He was quite a little chap then, and was going to be content with four pages, and was willing to "make himself generally useful;" but when he once got his head out of the bag there was no holding him. He would have been bad enough by himself, but anything like the artful way he set Mr. Farmer and Ted Ravani on me I never heard tell of. Well, as I was saying, he declared he only wanted to be a little page-boy, but before he was half measured for his suit of livery he was for taking charge of the whole Institute. As for cheek, he doesn't care what he says to you; and one day when I told him he wasn't fit to be out of long clothes, he threatened to appear in Mr. P——m's unmentionables. Then he wanted to keep our books for us, and register

I have said, so I shan't give him one; but I will take the gag out of his mouth, and you can see what he has to say for himself. So, here goes——

Yes! Gentlemen, here I am, and I've been waiting some minutes while that precious dad of mine has been trying to take away my character. However, it's a good thing I've got a tongue of my own; so you needn't believe all that he says. You want to know my name, do you? Well, use your eyes. It's a very good name—quite as good as you have got, if not better. I'm the last new member, and I got in without ballot, which is more than you did. I had no one to propose or second me, though I could have had a Committee-man if I had asked him. Do you want to know what I think of you? No? Very well, but you shall all the same; and it won't be a bit of good for you to try and change my opinion when I have once spoke, because people of my type can't be altered. I think you are a pretty good lot as a whole, but how you've got on so long without me is more than I can think. One thing is pretty certain at any rate, and that is, you won't be able to get on without me in future. I am exceedingly modest, but I should be doing myself an injustice if I were to withhold the fact that I am the most important member in the Institute just now. In addition to my modesty I may men-

THE POLYTECHNIC MAGAZINE



Home News, 1879

The Polytechnic Magazine, 1960

THE

NATIONAL

ARCHIVES

Welcome to the GLAM Workbench

Here you'll find a collection of tools, tutorials, examples, and hacks to help you work with data from galleries, libraries, archives, and museums (the GLAM sector). The primary focus is Australia and New Zealand, but new collections are being added all the time. Let me know if there's some GLAM data you'd like me to explore – [suggestions](#) are always welcome!



Buy Tim a coffee

GLAM Workbench: <https://glam-workbench.net/>

The Archives Unleashed Project

The Archives Unleashed project aims to make petabytes of historical internet content accessible to scholars and others interested in researching the recent past. Our team develops web archive search and data analysis tools to enable scholars, librarians and archivists to access, share, and investigate recent history since the early days of the World Wide Web.

Read more about the next stage of the project [here](#).

Archives Unleashed:
<https://archivesunleashed.org/>

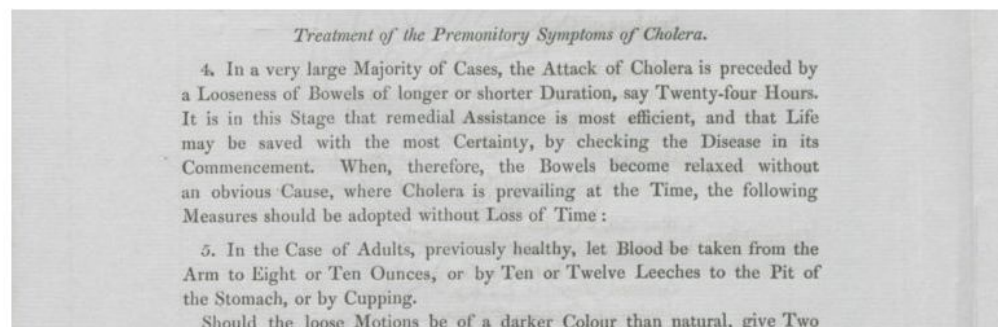
THE
NATIONAL
ARCHIVES

Select a category ▼ Select a month ▼ Select an author

Computing Cholera: Topic modelling the catalogue entries of the General Board of Health

Tuesday 2 June 2020 | [Chris Day](#) | [Records and research](#) | [Comment](#)

The 19th century saw a number of global cholera pandemics, beginning in India and spreading west. The British Isles were hit during the second global cholera pandemic (1829-1837). Cholera arrived in Gateshead via ship in October 1831. The epidemic spread over the following year, leaving around 52,000 dead.



Chris Day, *Computing Cholera*,
<https://blog.nationalarchives.gov.uk/computing-cholera-topic-modelling-the-catalogue-entries-of-the-general-board-of-health/>

Conduct Setup Episodes ▾ Extras ▾ License Improve this page 

Computational Analysis of Catalogue Data

roduces people working in catalogue-related roles in galleries, libraries, archives, and museums to analysing catalogue c
understand what the AntConc software does and how to use approaches from computational linguistics for the purposes

you will need to install [AntConc](#) and download the file [IAMS_Photos_1850-1950_selection](#)

| Setup | Download files required for the lesson |
|--------------------------------|--|
| 1. Introduction to AntConc | What is AntConc? What is corpus linguistics? How can they be combined t |
| 2. Importing data into AntConc | How do I get data into AntConc? |
| 3. Layout of AntConc | How is data organised in AntConc? How do I access different tools in AntConc? |
| 4. Settings in AntConc | What are the default settings in AntConc? What changes are recommended for analysing catalogue data? How can I keep a track of the settings I am using in AntConc? |
| 5. Word lists | What are word lists in AntConc and when would you use it? How do word lists work in AntConc? How might I interpret word lists generated from catalogue data? |

James Baker, *Computational Analysis of Catalogue Data*,
<https://cataloguelegacies.github.io/antconc.github.io/>

| | |
|----------|--|
| THE | |
| NATIONAL | |
| ARCHIVES | |

ment are four members of the Poly amongst Hogg, accompanied by Major Hoar W. the officers and some 400 amongst the men. They carry Pelham and Mr. and tchell, inspected the with them our most heartfelt and affectionate wishes and P made a short speech at the end of the prayers for their welfare and success. We inspection wishing them all God-speed and a safe return . Need envious and pro ns, knowing that they will not less to say, her visit was much appreciated by the

Il amply justify and uphold the honour of the of the Poly Company. Poly and of t whose behalf they are offering

Layout changes over the years

[illegible][illegible]


**THE POLYTECHNIC
MAGAZINE**
THE POLYTECHNIC. FOUNDED BY QUINTIN HOGG
309 REGENT STREET, LONDON, W.1

VOL. XXVII—No. 3 MARCH, 1957 PRICE ONE SHILLING

[illegible][illegible]



Andrew Akhlaghi 

This lesson covers how to convert images of text into text files and translate those text files. The lesson will also cover how to organize and edit images to make the conversion and translation of whole folders of text files easier and more accurate. The lesson concludes with a discussion of the shortcomings of automated translation and how to overcome them.

Peer-reviewed

 CC-BY 4.0

Support PH

EDITED BY
Anna-Maria Sichani

REVIEWED BY
Clemens Neudecker

Gustavo Candela

PUBLISHED | 2021-01-06

MODIFIED | 2021-11-11

DIFFICULTY | Medium

 <https://doi.org/10.46430/phen0091>

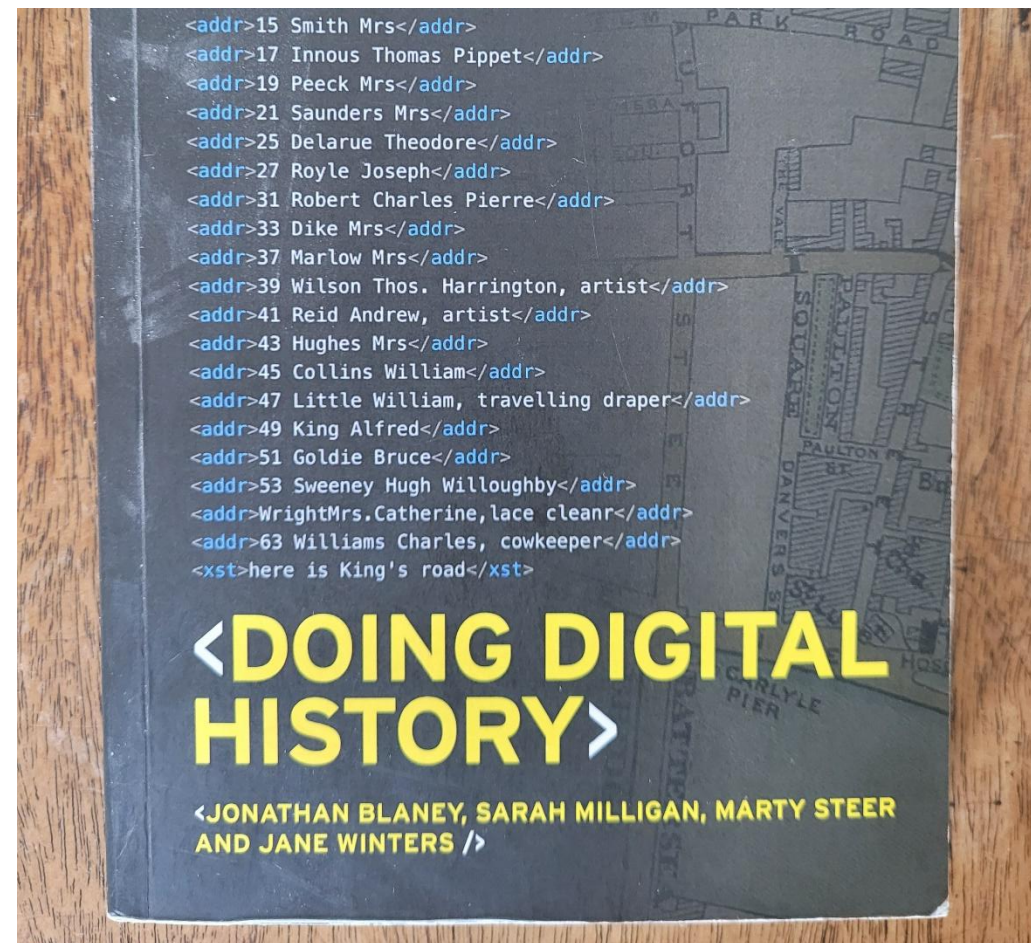
Donate today!

Great Open Access tutorials cost money to produce. Join the growing number of people supporting *Programming Historian* so we can continue to share knowledge free of charge.

Contents

- Introduction
 - Lesson Goals
- Setup
 - Command Line and Bash

Andrew Akhlaghi, *OCR and Machine Translation*, The Programming Historian,
<https://programminghistorian.org/en/lessons/OCR-and-Machine-Translation>



Doing Digital History, Jonathan Blaney, Sarah Milligan, Marty Steer and Jane Winters, *Doing Digital History*,
<https://manchesteruniversitypress.co.uk/9781526132680/>

THE

NATIONAL

ARCHIVES

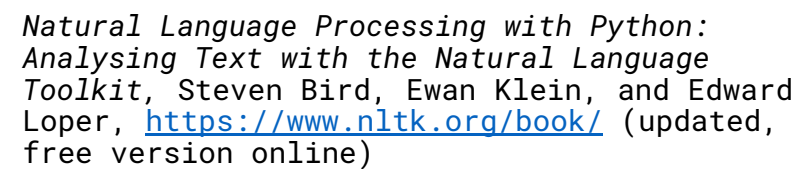
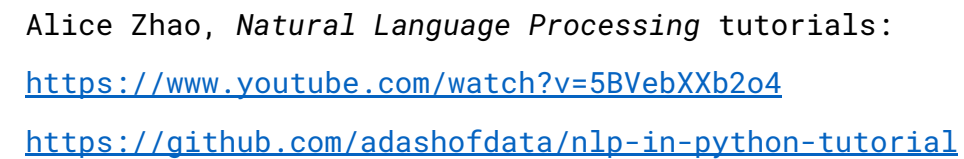
| Type | Date | Volume | Issue | Pages | OCR batch | Text | OCR filename |
|-------------------------------|---------|--------|-------|-------|-----------|--|----------------------------|
| Christian Workers' Supplement | 1891-07 | NaN | NaN | 4 | 1 | THE POLYTECHNIC MAGAZINE,\n\nChristian Workers... | CWSupp007.pdf.tiff.txt |
| Christian Workers' Supplement | 1891-09 | NaN | NaN | 4 | 1 | THE POLYTECHNIC MAGAZINE, 809, Regent Street,\n... | CWSupp008.pdf.tiff.txt |
| Christian Workers' Supplement | 1891-10 | NaN | NaN | 4 | 1 | THE POLYTECHNIC MAGAZINE,\n\nChristian Workers... | CWSupp009.pdf.tiff.txt |
| Christian Workers' Supplement | 1891-11 | NaN | NaN | 4 | 1 | THE POLYTECHNIC MAGAZINE, 809, Regent Street, ... | CWSupp010.pdf.tiff.txt |
| Christian Workers' Supplement | 1891-12 | NaN | NaN | 4 | 1 | THE POLYTECHNIC MAGAZINE, 809, Regent Street, ... | CWSupp011.pdf.tiff.txt |
| ... | ... | ... | ... | ... | ... | ... | ... |
| Quintinian | 1892-08 | NaN | 5 | 6 | 12 | Supplement to Podytechnte Magiesine,\n\nAugust... | Quintinian005.pdf.tiff.txt |
| Quintinian | 1892-09 | NaN | 6 | 2 | 12 | Supplement to Polytechnic Magazine, September ... | Quintinian006.pdf.tiff.txt |
| Quintinian | 1892-10 | NaN | 7 | 8 | 12 | [Supplement to Polytechnic Magazine, October 6... | Quintinian007.pdf.tiff.txt |
| Quintinian | 1892-11 | NaN | 8 | 8 | 12 | {Supplement to Polytechute Magazine, November ... | Quintinian008.pdf.tiff.txt |
| Quintinian | 1892-12 | NaN | 9 | 8 | 12 | [Supplement to Polytechnic Magazine, December ... | Quintinian009.pdf.tiff.txt |

The initial Poymags dataframe

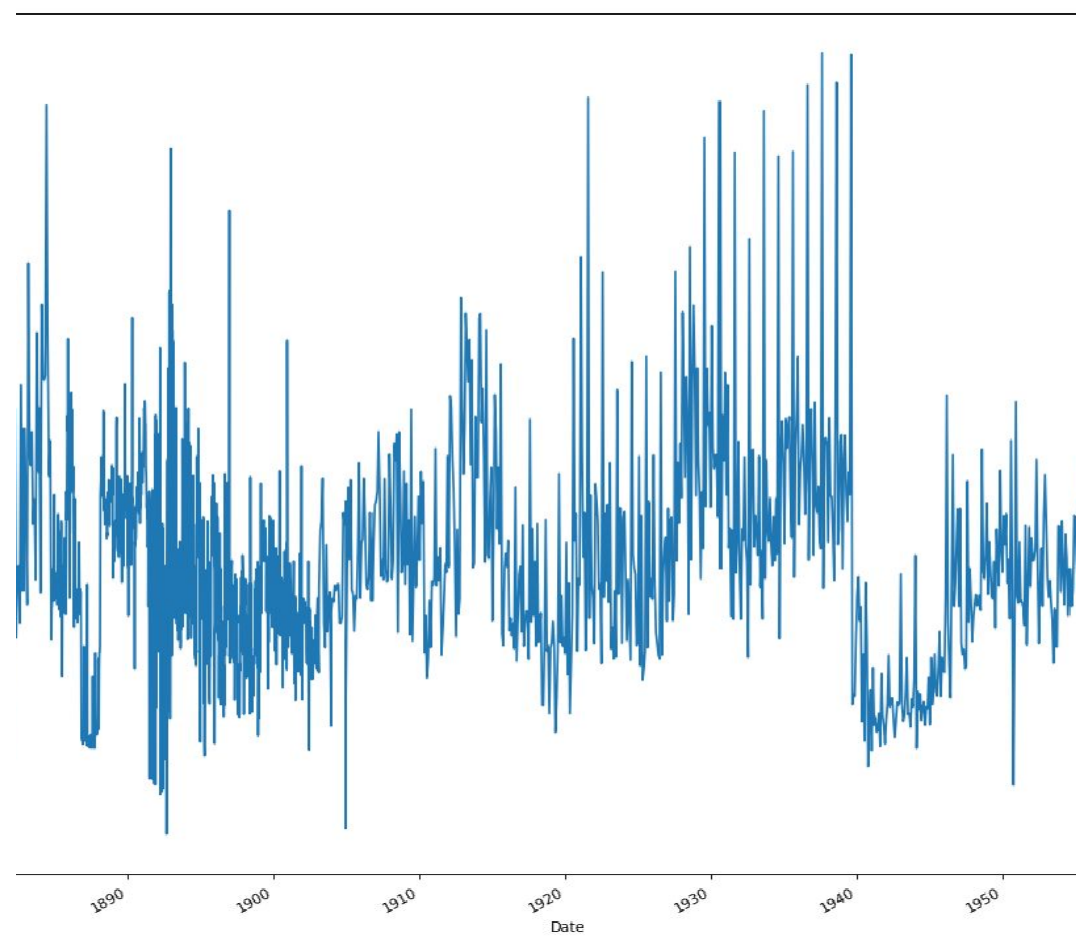
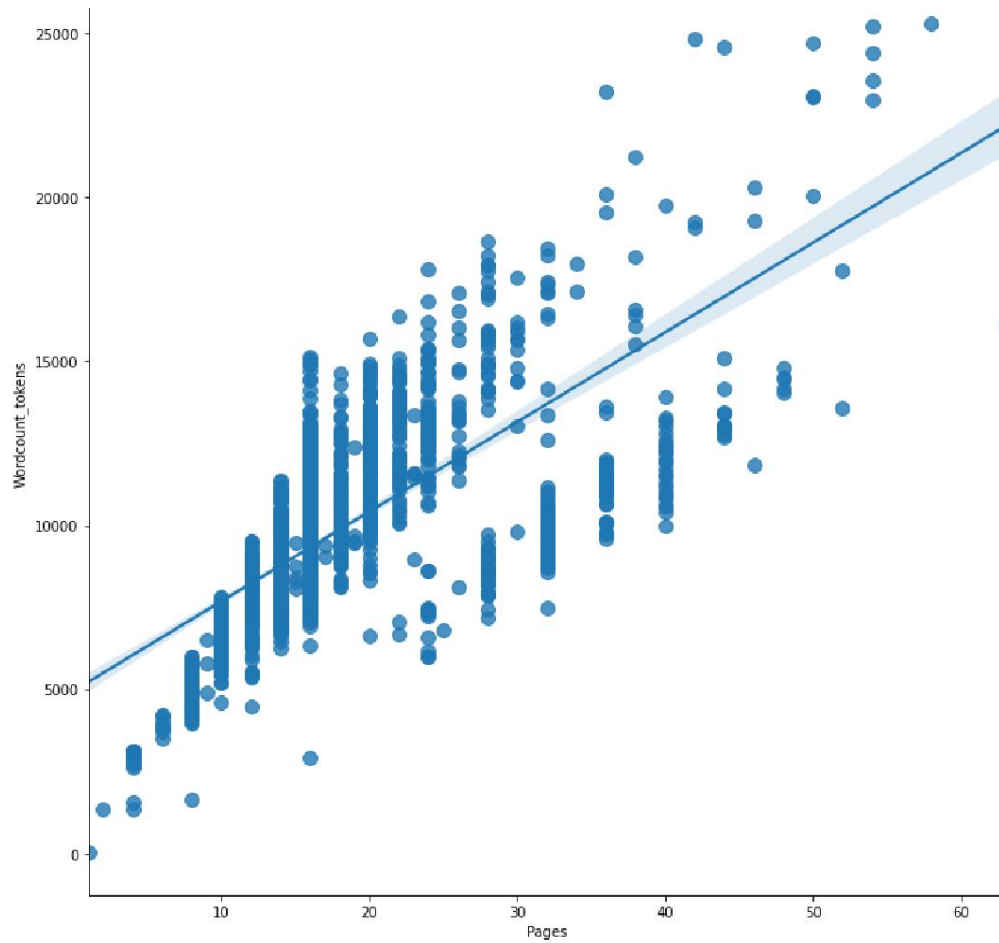
| | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|----------|--|
| | | | | | | | | | | | | | THE | |
| | | | | | | | | | | | | | NATIONAL | |
| | | | | | | | | | | | | | ARCHIVES | |

Data cleansing and normalisation

- Finding and correcting more OCR errors
- Making text lower case
- Removing punctuation & numbers
- Removing ‘stop words’ (e.g. ‘and’, ‘the’, ‘a’)
- Tokenizing (splitting sentances into individual words)
- Lemmatization (reducing different forms of a word to a single term)



| | |
|----------|--|
| THE | |
| NATIONAL | |
| ARCHIVES | |



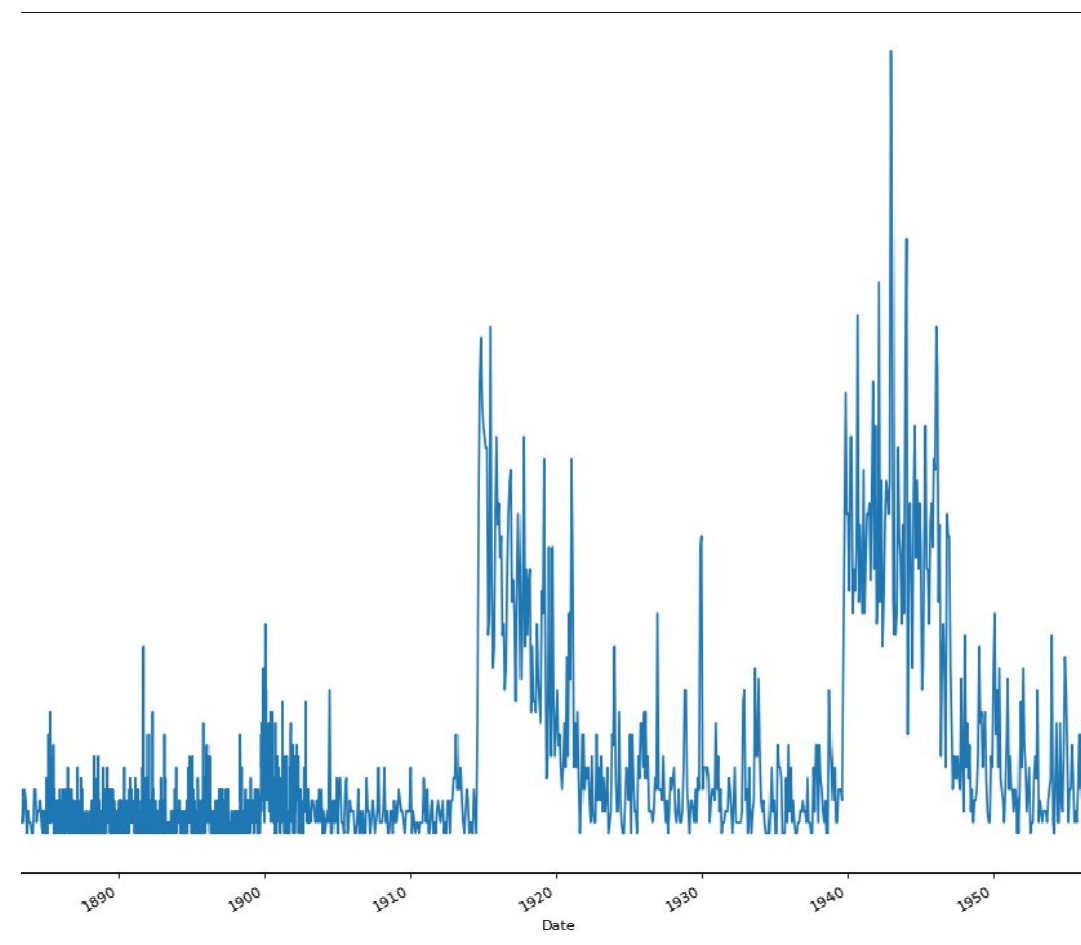
| | | | | | |
|----------------------|--|-----------------------|--|----------|--|
| Word counts vs pages | | Word counts over time | | THE | |
| | | | | NATIONAL | |
| | | | | ARCHIVES | |


```
#can do multiple
nltk_text.dispersion_plot(['Glen-Haig', 'Fencing', 'trophy'])
#this helps with the scale issue
#N.B. scale is in tens of millions (number of words from beginning)
```

```
#can do multiple
nltk_text.dispersion_plot(['Glen-Haig', 'Fencing', 'trophy'])
#this helps with the scale issue
#N.B. scale is in tens of millions (number of words from beginning)
```



Concordances and lexical dispersion plots for Mary Glen-Haig



| Frequency plot for 'War' | | | | | | THE |
|--------------------------|--|--|--|--|--|--------|
| | | | | | | NATION |
| | | | | | | ARCH |

Topic Modelling of Australian Parliamentary F

An exploration by [Adel Rahmani](#)

This Jupyter notebook can be found on [GitHub](#)

As part of his wonderful work with [Trove](#) data, [Tim Sherratt](#) has harvested and made available releases (the data can be downloaded from his [GLAM Workbench website](#)). The collection words relating to immigration/refugee issues.

On his [GLAM Workbench website](#) Tim Sherratt explains how the documents were harvested

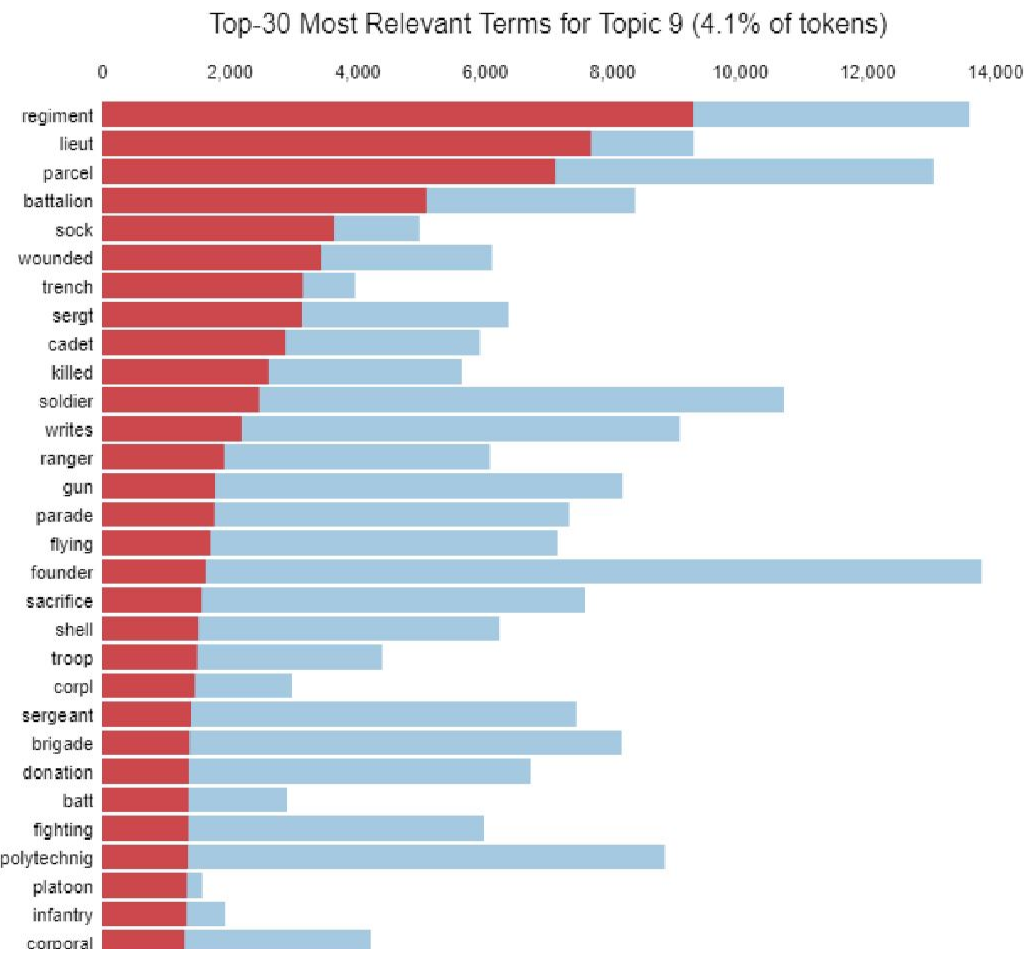
Politicians talking about 'immigrants' and 'refugees'

Using the notebook above I harvested parliamentary press releases that included 'illegal arrivals', or 'boat arrivals'. A total of 12,619 text files were harvested.

I was curious about the contents of the press releases, however, at more than 12,000 documents get the computer to do it for me, and use topic modelling to poke around the corpus.

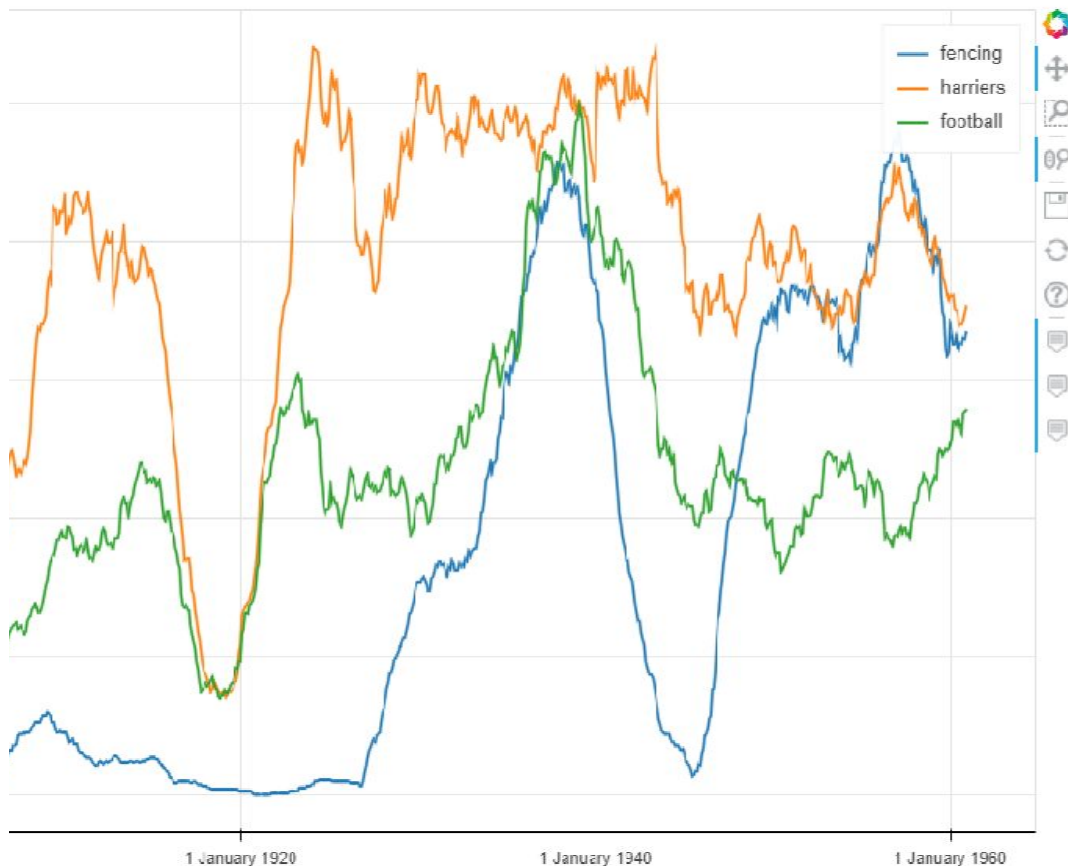
Let's start by importing several modules that we will need.

Adel Rahmani, *Topic Modelling of Australian Parliamentary Press Releases*, <https://github.com/adelr/trove-refugee>



Examining the 'war' topic

| | |
|----------|--|
| THE | |
| NATIONAL | |
| ARCHIVES | |



Frequency analysis with sport teams, from
https://colab.research.google.com/drive/1Wa44qr8xMK2kd-eIk-M0_-IqW41D2nB3?usp=sharing

Set parameters for including common and rare words

These parameters will allow you to control what words are included in your topic model.

MinDf determines the number of documents a word must occur in to be included, while **MaxDf** is the the maximum while a high **MaxDf** will allow more common words from across the corpus, potentially resulting in more general topics.

You can either enter a whole number (e.g. 10) or a decimal which will be understood as a percentage of the corpus (i.e. 50%) would mean that each word may not appear in more than half of the documents.

For a further discussion of how this works, see Adel Rahmani's excellent notebook on [Topic Modelling of Australia](#).

MinDf: 10

MaxDf: 0.5

[Show code](#)

[6] Select the date range you would like to model

The digitised run of the Polytechnic Magazine covers the years 1879 to 1960. You can select a date range to model.

StartDate: 1894

EndDate: 1960

[Show code](#)

[7] Create document-term matrix

When you have finished setting the parameters above, run this cell and a document-term matrix will be created. **BUT** if you want to change the parameters above later on then as well as running the cells above, make sure you run this cell again.




✓ 7s cc

Setting parameters for topic modelling in Google Colab, from
https://colab.research.google.com/drive/1HAS-Xo3EH_093Lp_nTV4pZT3eV8F5bBX?usp=sharing

THE

NATIONAL

ARCHIVES

| | | | |
|---|-----------------|-------------------|---------------|
|  | PublicNotebooks | Update README.txt | 10 months ago |
|  | LICENSE | Initial commit | 10 months ago |
|  | README.md | Update README.md | 10 months ago |

☰ README.md

Computational access to the Polytechnic Magazine (1879 - 1960)

This project aimed to unlock new ways of accessing a digitised collection of historic magazines held by the University of Westminster Archive. It was undertaken as part of my PG Cert. in Applied Data Science at Birkbeck.

The Polytechnic Magazine was the in-house magazine of the [Regent Street Polytechnic](#), one of the predecessor institutions of the [University of Westminster](#). Since 2011, a digitised run of more than 1700 issues of the magazine covering the years 1879 to 1960 has been made available by the [University Archive](#). This has proved an invaluable resource for academic researchers, family historians and university staff. You can search and read the digitised magazines themselves via the University Archive's [Polymags site](#). This project aimed to complement the existing resource by opening up computational methods of access to the collection.

Project outputs

Public notebooks

These are designed to be accessible to all users regardless of their level of experience with digital analysis of texts.

The project GitHub page: <https://github.com/jakebickford/PolyMags>

| | |
|----------|--|
| THE | |
| NATIONAL | |
| ARCHIVES | |

[illegible]

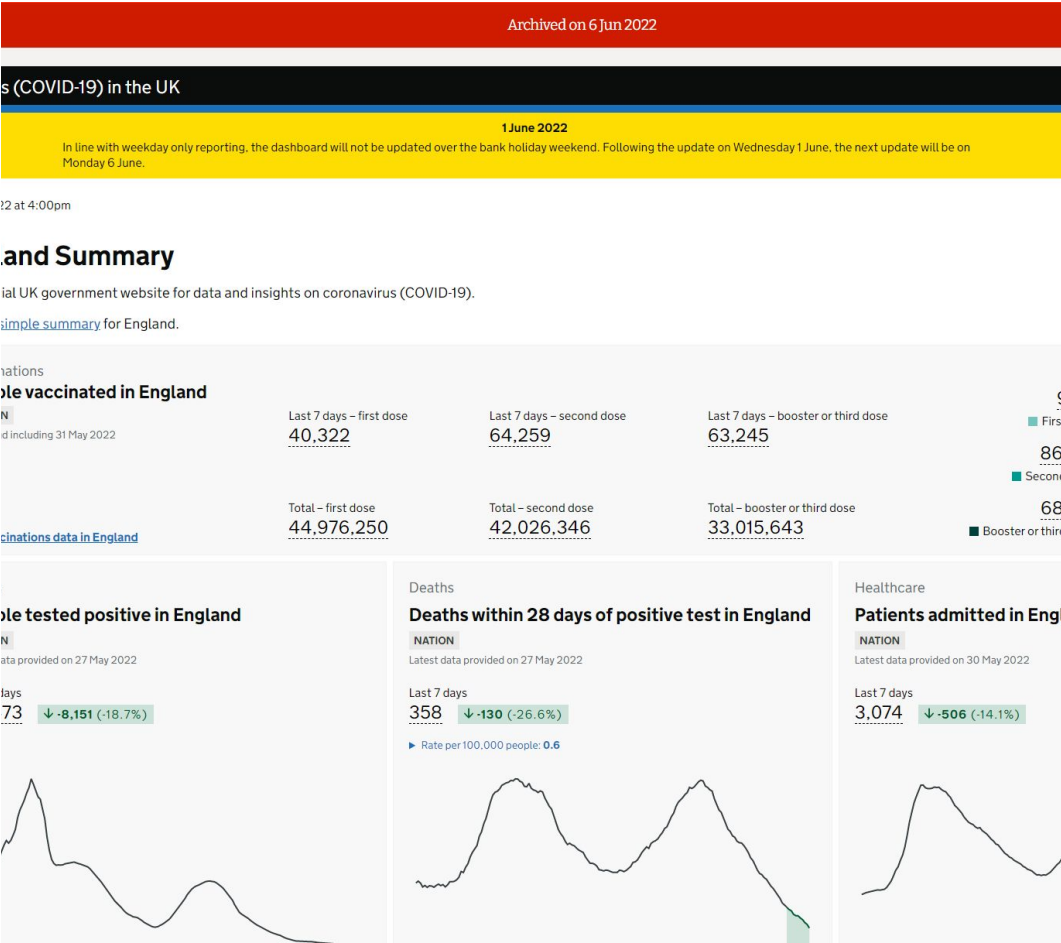
Things to think about

- Collaboration and user engagement
- Be realistic in scoping your project
- Think about how your work will be received
- How will your project be sustained?
- There are loads of great resources out there!

[illegible]



HM Treasury, 5 Nov 1996,
<https://webarchive.nationalarchives.gov.uk/ukgwa/19961105224031/http://www.hm-treasury.gov.uk:80/>



Covid Dashboard, 6 Jun 2022,
<https://webarchive.nationalarchives.gov.uk/ukgwa/20220606101231/https://coronavirus.data.gov.uk/>

THE
POLYTECHNIC
MAGAZINE

Vol. LXXXV—No. 12.

DECEMBER, 1945.

PRICE TWOPENCE

Don't forget the
BOVRIL



-it adds the appetising
flavour of Beef

Feel free to get in touch!

- jake.Bickford@nationalarchives.gov.uk
- [@b1ninch](#)

Polytechnic Magazine Resources

- <https://polymags.westminster.ac.uk/> (the original website where you can read the magazines)
- <https://github.com/jakebickford/PolyMags/> (github for the project)

| | |
|----------|--|
| THE | |
| NATIONAL | |
| ARCHIVES | |