Contents

[Foreword 1](#_Toc101199372)

[Introduction to the Data 1](#_Toc101199373)

[Sample Use – Searching for a Word or Phrase 1](#_Toc101199374)

[Appendix 2](#_Toc101199375)

[Get Credentials 2](#_Toc101199376)

[Connect to the Database 2](#_Toc101199377)

# Foreword

If you wish to do your own queries on my data, then you will need to use SQL (structured query language), which is the standard language used to interrogate databases.

If your queries are lightweight – meaning ones that will not return more than 5,000 records – then there is no need for you to log in to my database. That is because I have provided a web page where you can simply type the SQL and press a button to get the results: <http://shakespearestext.com/can/sql.htm> .

If you need to do more heavyweight queries, without limit on the number of results, then you will need to log in directly to my database. In this case, follow the instructions in the Appendix.

# Introduction to the Data

You have access to three tables of data, called **words**, **plays**, and **divisions**. They are what I used to produce all my published CAN results. You can begin to browse them like this:

**select \* from words;**

**select \* from plays;**

**select \* from divisions;**

# Sample Use – Searching for a Word or Phrase

This can of course be done using EEBO-TCP or LION but if you wish to use this database then please follow these instructions:

Try running this command and look at the results it gives you:

select \* from words\_v where w0 = 'i' and w1 = 'love' and w2 = 'you';

The results are showing you all the occurrences of the 3-gram *i love you*. The first few columns in the output are obvious. The last 10 columns are w0 to w4, and a0 to a4. Of these, w0 to w4 are five consecutive words in the play, whereas a0 to a4 are the lemmatized forms of those words.

You can specify a longer search phrase, such as in this command:

select \* from words\_v where w0 = 'i' and w1 = 'love' and w2 = 'you' and w3 = 'i' and w4 = 'do';

All the w and a values are in lowercase, so **please always use lowercase** in your searches. You will still get results in whatever case is in the texts. And of course all words are in modern spelling.

You will get more hits if you search for lemmatized forms of words, e.g.:

select \* from words\_v where a0 = 'i' and a1 = 'love' and a2 = 'you';

With lemmatized searches you will need to guess what the lemma might be for your search words. But that's usually easy to find out. E.g. if you wanted to know what the lemma for the word *delighted* is then just do this query:

select \* from words\_v where w0 = 'delighted';

If you now look in the a0 column you will see the lemma for *delighted*.

# Appendix

This appendix explains how you can have a direct connection to my database.

## Get Credentials

Email me ([pervez.rizvi@gmail.com](mailto:pervez.rizvi@gmail.com)) to ask for your own username and password. As with the rest of CAN, this is free of charge, there is no charge for using the database, and any results you get from your searches are provided under the same Creative Commons licence as the rest of CAN.

## Connect to the Database

All you need to use the database is an internet connection and a browser. The screenshots below are from the Opera browser, but any modern browser should do.

Click on the URL below or paste it into your browser:

<https://ogdvjblzqo9ybxg-can.adb.uk-london-1.oraclecloudapps.com/ords/sql-developer>?

You should get a login screen like this:

A screenshot of a computer

Description automatically generated

Log in using the username and password I have given you.[[1]](#footnote-2) You should see a screen like this:

Graphical user interface, text, application, email

Description automatically generated

Click on the option pointed to by the arrow and you should see this:

Graphical user interface, application

Description automatically generated

You do not need the sidebar, so hide it by clicking on the little grey arrow as shown above.

Graphical user interface, application

Description automatically generated

The top half of this window is where you will type the commands to do your searches. The bottom half will then show you the results and allow you to download them. You can adjust the relative sizes of the two halves by dragging the grey arrow up or down.

1. The database shuts itself down after a few days of inactivity. If this happens you will not be able to log in. If you are sure that you have entered the correct username and password but cannot log in, then contact me to ask me to restart the database. [↑](#footnote-ref-2)