Benefit of this database is that it is a massive repository of different points of view

Illuminated by the different journals, talking about similar things, in the context of time changing

Time change = march of time

Time as temporality = Defined by Critical Events

* Napoleon Defeated at Waterloo at 1815
  + Soldiers getting paid by British Government
  + Former soldiers unemployed on the streets on London

1. Temporality

* Want a subset of that time (it is worth knowing within a year range)

1. Relationship of certain words

* Linking this into Publisher
* Type of Article, Title of Journal

Having a reference case

-------------------------------------------------------------------------------------------------------------------------------

How to make a parameterization optional

Questions:

*What are the different forms that a word appears in journals?*For each five year period, what was the count of each way a word appeared that year

Needed: Line Chart, Regex, Conditional (if “ “ in query then return two words)

*How many articles in the archive did each journal publish cumulatively over time?*

* How many articles did Journals with titles rlike {{parameter}} publish cumulatively overtime

Needed: JournalTitle, RLIKE

Charts: Line Chart (with logarithmic scale)

*Miscellaneous*  
- In a journal, what is the distribution of the types of articles?

Needed: TypeOfArticle, JournalTitle RLIKE (Parameterization)

* Who was the most prolific publisher over a time?

Needed: Publisher, Time

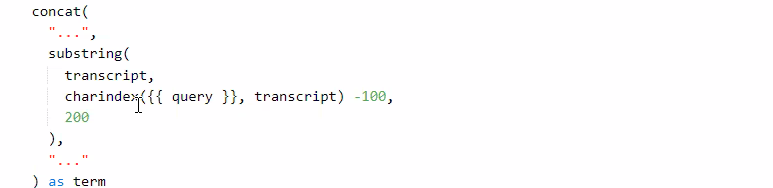
*What context does a word appear in?*

* What are the top 10 most common words that appear in sentences that contain the word law

Needed: Text Column, Regex knowledge, SQL function knowledge

Charts: Table, Word Cloud

We would have to remove or not consider stop words



FAILED:

Query -

SELECT split(array\_join(collect\_set(s), " "), " ") as w

FROM

(SELECT

array\_join(regexp\_extract\_all(

lower(text),

"[^\\?\\!\\.]\*(\\W)" || {{word1}} || "(\\W)[^\\?\\!\\.]\*.", 0)

, ' ') as s

FROM

century1\_avro

WHERE

(

to\_date(cast(NumericPubDate as string), 'yyyyMMdd') BETWEEN DATE '{{ date.start}}'

AND DATE '{{ date.end }}'

)

AND (TypeOfArticle IN ({{Type of Article}}) OR {{Type of Article}} == '-- All Categories --'))