

# Encyclopædia Britannica;

*James* OR, A *Gullerton*

## DICTIONARY

OF

ARTS and SCIENCES,

COMPILED UPON A NEW PLAN.

IN WHICH

The different SCIENCES and ARTS are digested into  
distinct Treatises or Systems;

AND

The various TECHNICAL TERMS, &c. are explained as they occur  
in the order of the Alphabet.

ILLUSTRATED WITH ONE HUNDRED AND SIXTY COPPER

By a SOCIETY of GENTLEMEN in SCOTLAND

IN THREE VOLUMES

VOL. I.

EDINBURGH

Printed for A. BELL and C. M.

And sold by COLIN MACFARQUHAR, at

M.DCC.

***frances: A Deep Learning NLP and  
Text Mining Web Tool to Unlock  
Historical Digital Collections***

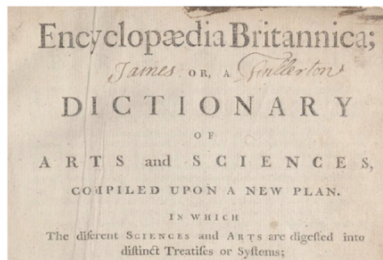
***A Case Study on the Encyclopaedia Britannica***

**Dr. Rosa Filgueira,  
Lecturer at the School of Computer Science,  
University of St Andrews,  
Email: [rf208@st-andrews.ac.uk](mailto:rf208@st-andrews.ac.uk)**

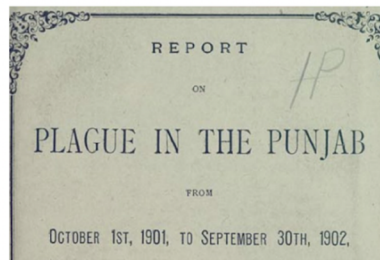
# frances: A Deep Learning NLP and Text Mining Web Tool to Unlock Historical Digital Collections

## Digitised collections

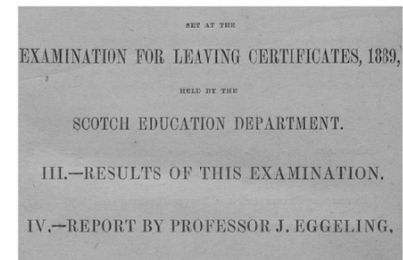
Download the ALTO, METS, image and plain text files for our digitised collections.



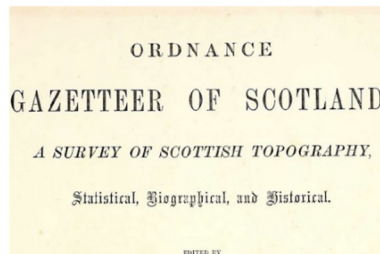
Encyclopaedia Britannica, 1768-1860



A Medical History of British India



Scottish School Exam Papers, 1888-1963





# *frances*: A Deep Learning NLP and Text Mining Web Tool to Unlock Historical Digital Collections

*frances* provides **abstractions** to a variety of ML/NLP techniques

→ *Extract complex knowledge without being an expert data scientist*

- Train and use text embedding models
- Employ topic mining, sentiment analysis, text summarization
- Build knowledge graph(s) visualizing the results

Using the **Encyclopaedia Britannica** – *frances* will allow us automatically to

- Group similar articles
- Detect how articles have changed across editions
- Extract the relationships between articles
- Classify articles into different categories
- Summarise articles
- Analyse the sentiment expressed in an article

*Overview of how the  
Encyclopaedia Britannica  
has changed over time*



New suite of ML functionalities that can be used to analyse  
any other Data Foundry collection

## The Journey

Phase 1: Information Extraction

Phase 2: EB Ontology and Knowledge Graph

Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

Phase 4: Defoe and Knowledge Graph

Phase 5: Flask Web-Application



## The Journey

### **Phase 1: Information Extraction**

Phase 2: EB Ontology and Knowledge Graph

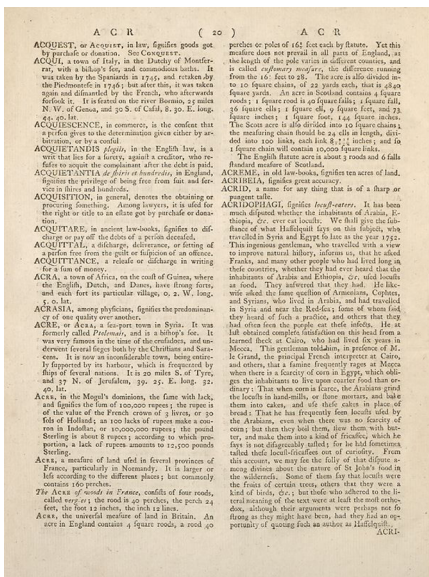
Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

Phase 4: Defoe and Knowledge Graph

Phase 5: Flask Web-Application

- Two types of Terms:
  - **Articles:** 1 or 2 paragraphs describing a term ~ entry in a dictionary
  - **Topics:** Several pages describing a term
- Extract Metadata from all EB editions and volumes.

How have we done it? By using *defoe* to analyze METS and ALTO XML files



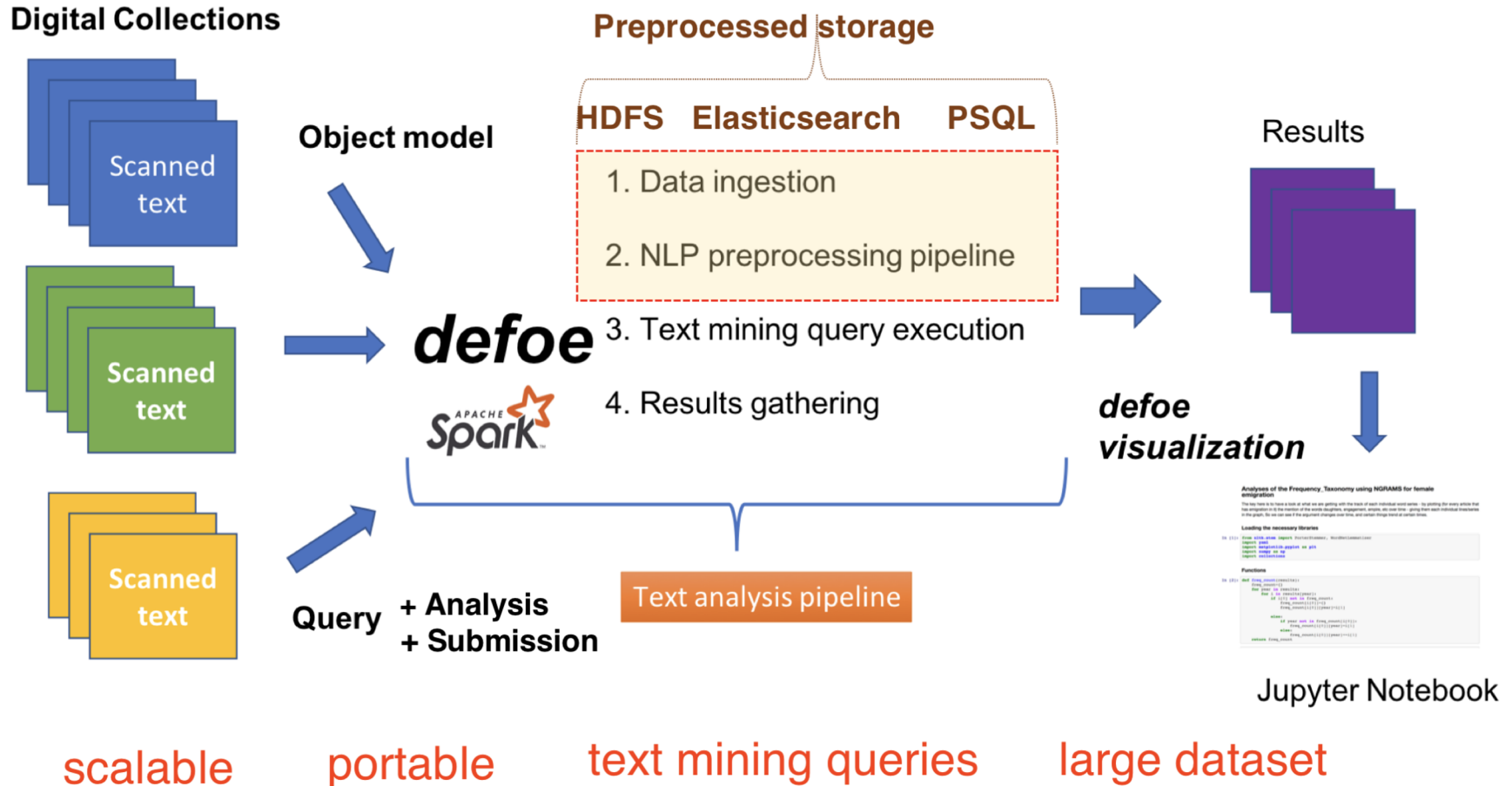
```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <alto
3   xmlns="http://www.loc.gov/standards/altto/v3/altto.xsd"
4   <description>
5     <sourceImageInformation>
6       <fileName>/data/pdfs/c_188936619/1_144133901/18882693.23.pdf</fileName>
7     </sourceImageInformation>
8     <ocrProcessing>
9       <ocrProcessingStep>
10        <processingDate>Fri Jul 12 10:36:14 2019
11      </ocrProcessingStep>
12    </ocrProcessing>
13    <processingSoftware>
14      <softwareName>pdfalto</softwareName>
15      <softwareVersion>0.1</softwareVersion>
16    </processingSoftware>
17  </description>
18  <styles>
19    <textStyle ID="font0" FONTFAMILY="times" FONTSIZE="10.000" FONTTYPE="serif" FONTWIDTH="proportional" FONTCOLOR="#000000" FONTSTYLE="normal">
20    </textStyle>
21    <textStyle ID="font1" FONTFAMILY="times" FONTSIZE="6.500" FONTTYPE="serif" FONTWIDTH="proportional" FONTCOLOR="#000000" FONTSTYLE="subscript">
22    </textStyle>
23  </styles>
24  <layout>
25    <Page ID="Page36" PHYSICAL_IMG_Nr="36" WIDTH="2466" HEIGHT="3327">
26      <PrintSpace>
27        <TextLine WIDTH="232" HEIGHT="75" ID="p36_11" HPOS="512" VPOS="188">
28          <String ID="p36_11" CONTENT="ACRID" HPOS="512" VPOS="188" WIDTH="42" HEIGHT="75" STYLE="font0">
29            <SP WIDTH="57" VPOS="188" HPOS="554">
30              <String ID="p36_12" CONTENT="C" HPOS="611" VPOS="188" WIDTH="37" HEIGHT="75" STYLE="font0">
31                <SP WIDTH="50" VPOS="188" HPOS="648">
32                  <String ID="p36_13" CONTENT="11" HPOS="706" VPOS="188" WIDTH="38" HEIGHT="75" STYLE="font0">
33                    <TextLine>
34                      <TextLine WIDTH="132" HEIGHT="75" ID="p36_12" HPOS="1119" VPOS="188">
35                        <String ID="p36_12" CONTENT="C" HPOS="1119" VPOS="188" WIDTH="19" HEIGHT="75" STYLE="font0">
36                          <SP WIDTH="62" VPOS="188" HPOS="1138">
37                            <String ID="p36_13" CONTENT="11" HPOS="1206" VPOS="188" WIDTH="51" HEIGHT="75" STYLE="font0">
38                              <TextLine>
39                                <TextLine WIDTH="1829" HEIGHT="75" ID="p36_13" HPOS="1711" VPOS="182">
40                                  <String ID="p36_13" CONTENT="ACQUIT" HPOS="1711" VPOS="182" WIDTH="243" HEIGHT="75" STYLE="font0">
41                                    <SP WIDTH="24" VPOS="182" HPOS="1954">
42                                      <String ID="p36_14" CONTENT="C" HPOS="1954" VPOS="182" WIDTH="19" HEIGHT="75" STYLE="font0">
43                                        <SP WIDTH="15" VPOS="182" HPOS="1973">
44                                          <String ID="p36_15" CONTENT="ACQUIT" HPOS="1973" VPOS="182" WIDTH="193" HEIGHT="75" STYLE="font0">
45                                            <SP WIDTH="24" VPOS="182" HPOS="2166">
46                                              <String ID="p36_16" CONTENT="ACQUIT" HPOS="2166" VPOS="182" WIDTH="124" HEIGHT="75" STYLE="font0">
47                                                <SP WIDTH="24" VPOS="182" HPOS="2290">
48                                                  <String ID="p36_17" CONTENT="ACQUIT" HPOS="2290" VPOS="182" WIDTH="134" HEIGHT="75" STYLE="font0">
49                                                    <SP WIDTH="24" VPOS="182" HPOS="2424">
50                                                      <String ID="p36_18" CONTENT="ACQUIT" HPOS="2424" VPOS="182" WIDTH="182" HEIGHT="75" STYLE="font0">
51                                                        <SP WIDTH="24" VPOS="182" HPOS="2606">
52                                                          <String ID="p36_19" CONTENT="ACQUIT" HPOS="2606" VPOS="182" WIDTH="114" HEIGHT="75" STYLE="font0">
53                                                        <TextLine>
54                  <TextLine WIDTH="792" HEIGHT="75" ID="p36_14" HPOS="2211" VPOS="223">
55

```

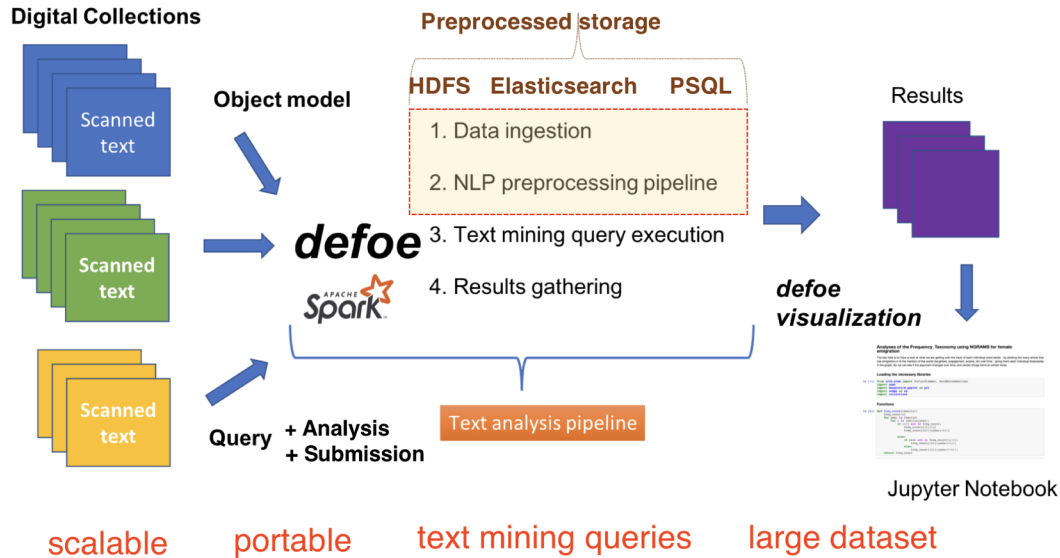


# Phase 1: Information Extraction



<https://github.com/francesNLP/defoe>

# Phase 1: Information Extraction



## Publications:

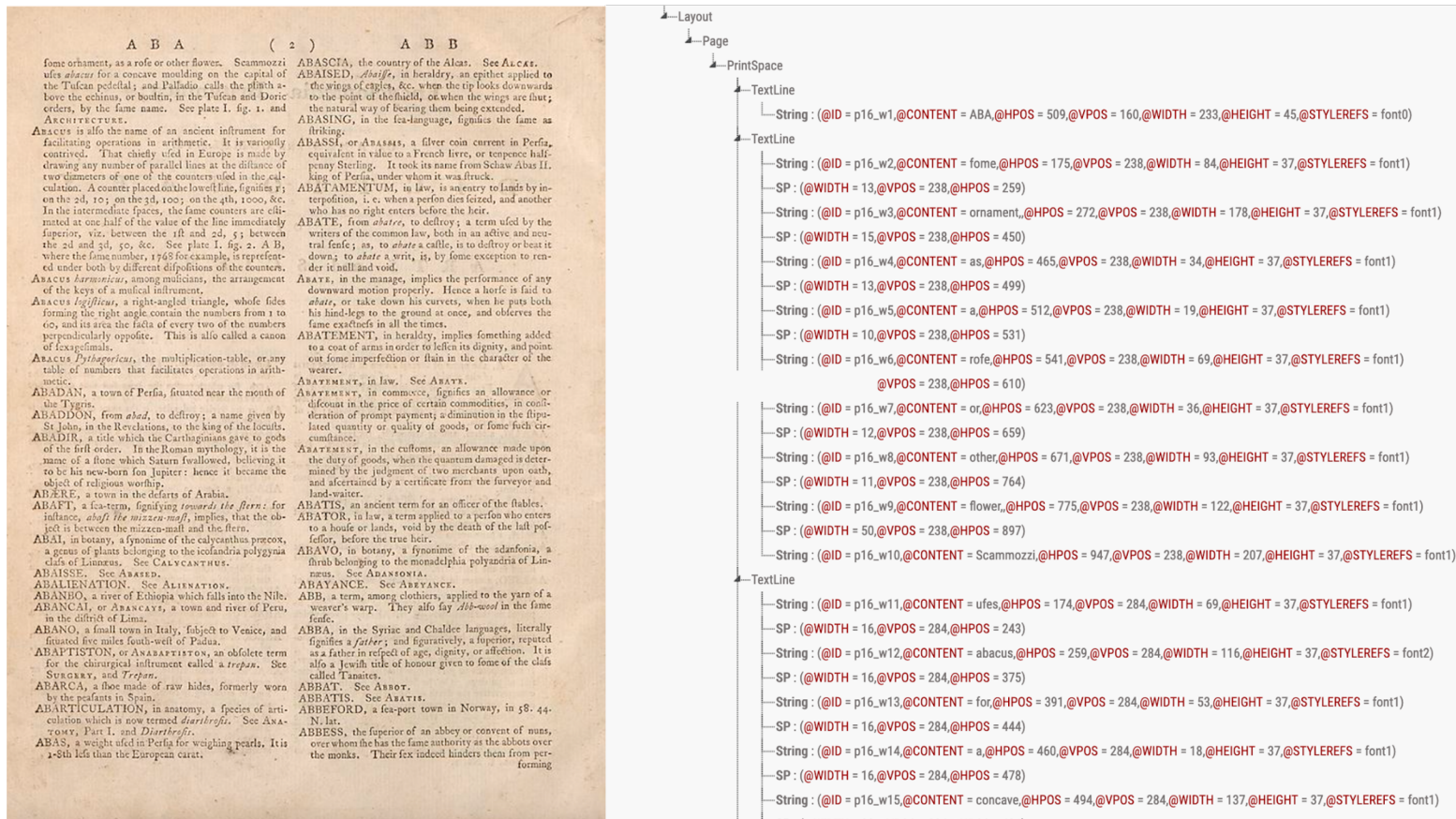
1. [defoe: A Spark-based Toolbox for Analysing Digital Historical Textual Data](#), 2019
2. [Geoparsing the historical gazetteers of scotland: Accurately computing location in mass digitised texts](#), 2020
3. [Extending defoe for the efficient analysis of historical texts at scale](#), 2021



# Phase 1: Information Extraction

## New *defoe* queries and heuristics

- *defoe* query to extract EB Terms:
  - By page and classify them between **articles** and **topics** → Using ALTO XML

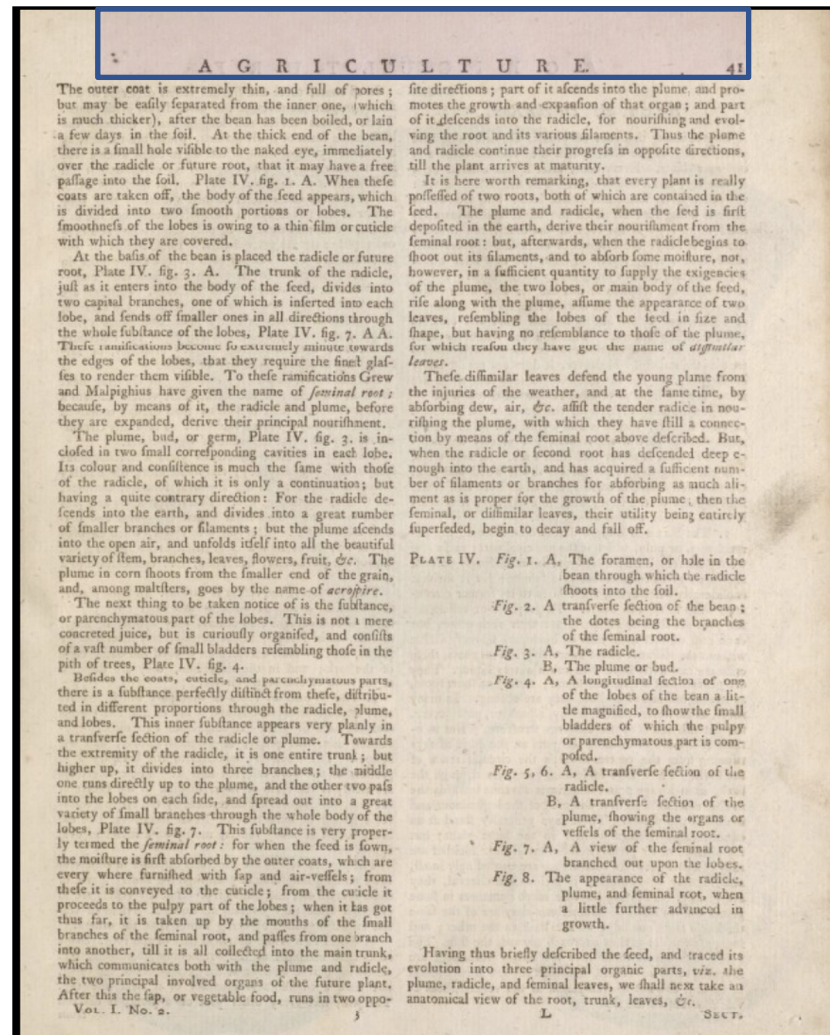
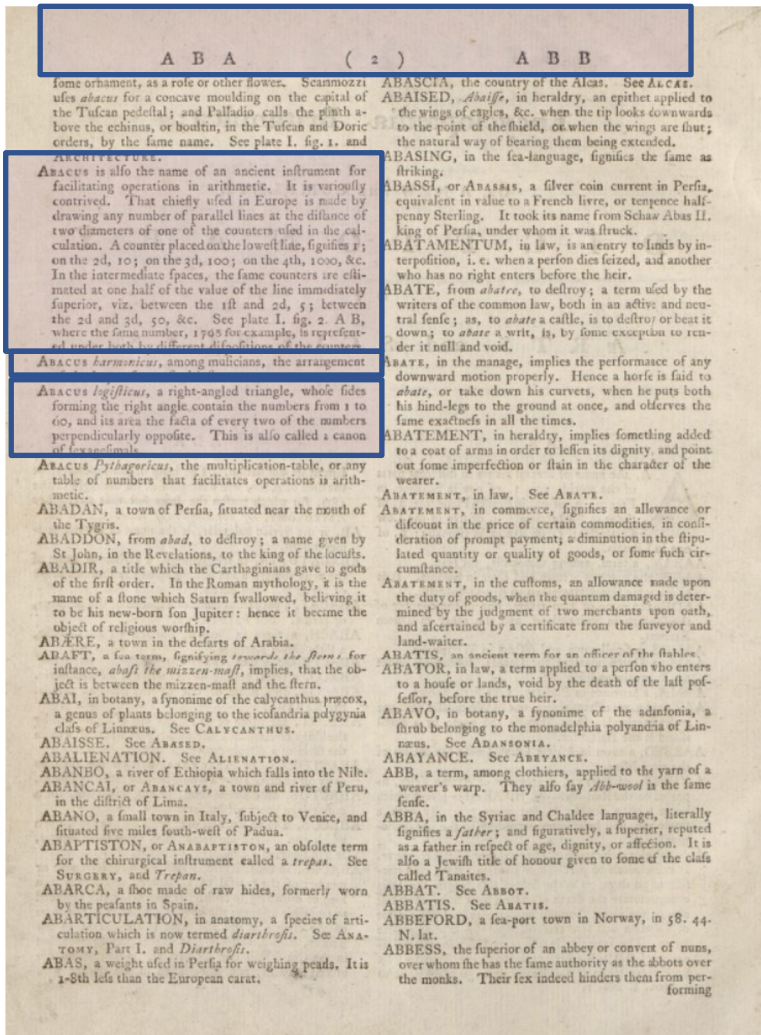


The image displays a digital library interface. On the left, a page from an old book is shown, featuring two columns of text. The text is in a historical font and includes various entries, some of which are marked with 'A B A' and '( 2 )'. The right side of the image shows the ALTO XML metadata for this page. The metadata is structured as a series of 'TextLine' elements, each containing attributes for ID, content, HPOS, VPOS, WIDTH, HEIGHT, and STYLEREFS. The 'TextLine' elements are organized into a 'Page' structure, which is further divided into 'Layout' and 'PrintSpace' sections. The 'TextLine' elements are numbered sequentially, starting from 1 and ending at 15. The 'TextLine' elements are organized into a 'Page' structure, which is further divided into 'Layout' and 'PrintSpace' sections. The 'TextLine' elements are numbered sequentially, starting from 1 and ending at 15.

# Phase 1: Information Extraction

## New *defoe* queries and heuristics

1. Detecting pages headers from ALTO XML
2. Using headers to classify terms into: Articles & Topics
3. Using ALTO Text for detecting the start of each article:
  - > Starting a line with TERM UPPERCASE + “,"



Articles

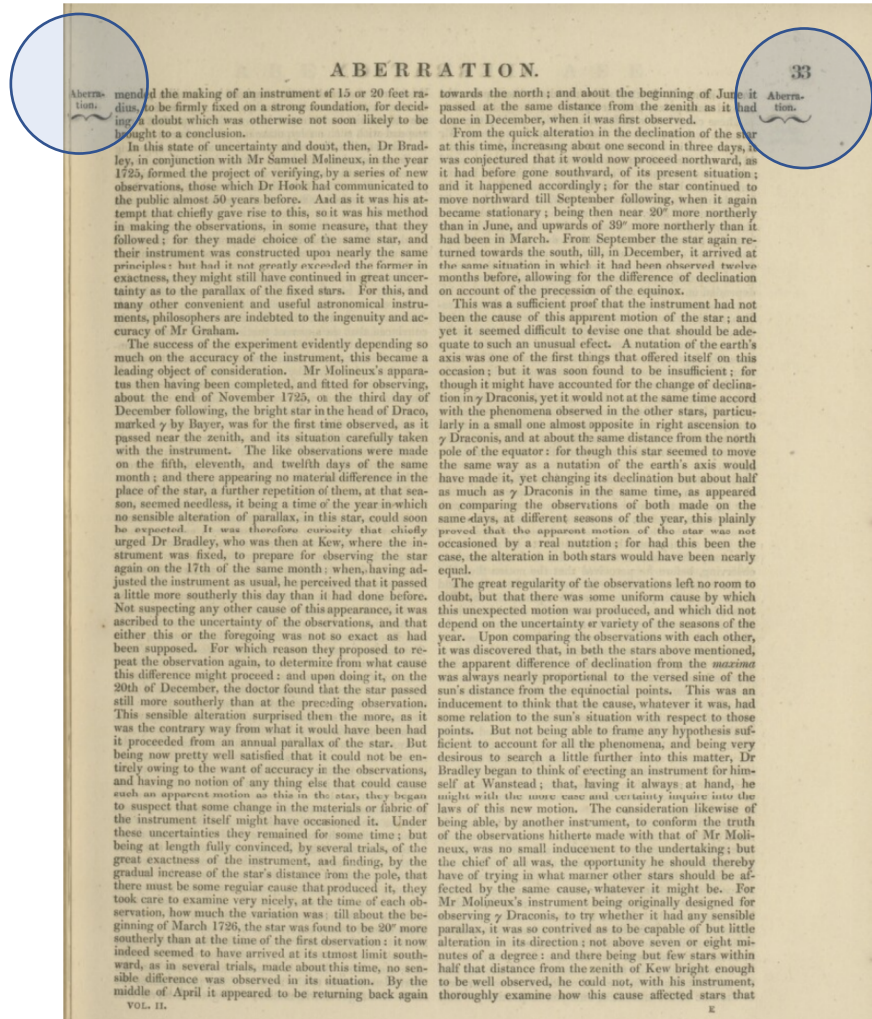
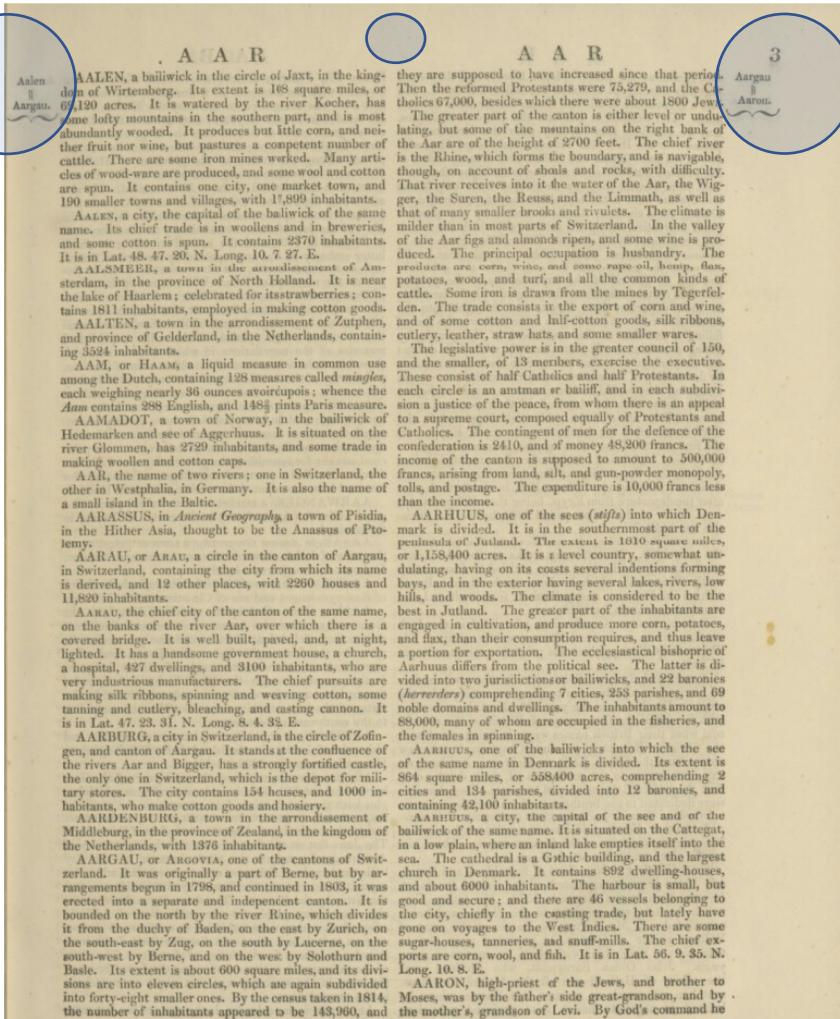
Topic

Edition 1 - 1771



## New *defoe* queries and heuristics

1. Detecting pages headers from ALTO XML
2. Using headers to classify terms into: Articles & Topics
3. Using ALTO Text for detecting the start of each article:  
--> Starting a line with TERM UPPERCASE + “,”



## Articles

# Topic

# Edition 7

## New *defoe* queries and heuristics

```
term SCIENCE
definition in philology, denotes any dpdfrine, deduced f...
relatedTerms []
header SCISCO
startsAt 658
endsAt 658
numberOfTerms 24
numberOfWords 15
numberOfPages 872
positionPage 7
typeTerm Article
editionTitle First edition, 1771, Volume 3, M-Z
editionNum 1
supplementTitle
supplementsTo []
year 1771
place Edinburgh
volumeTitle Encyclopaedia Britannica; or, A dictionary of ...
volumeNum 3
letters M-Z
part 0
altoXML 144133903/alto/144812443.34.xml
Name: 7454, dtype: object
```

**Science term** information extracted from the First Edition - 1771



# Phase 1: Information Extraction

## New *defoe* queries and heuristics

- *defoe* query to extract EB **metadata** :
  - Metadata per Edition and Volume → Using METS XML

	MMSID	editionTitle	editor	editor_date	genre	language	termsOfAddress	numberOfPages	physicalDescription	place	...
0	992277653804341	First edition, 1771, Volume 1, A-B	Smellie, William	1740-1795	encyclopedia	eng	None	832	3 v., 160 plates : ill. ; 26 cm. (4to)	Edinburgh	... h
1	992277653804341	First edition, 1771, Volume 2, C-L	Smellie, William	1740-1795	encyclopedia	eng	None	1018	3 v., 160 plates : ill. ; 26 cm. (4to)	Edinburgh	... h
2	992277653804341	First edition, 1771, Volume 3, M-Z	Smellie, William	1740-1795	encyclopedia	eng	None	872	3 v., 160 plates : ill. ; 26 cm. (4to)	Edinburgh	... h
3	9929192893804340	First edition, 1773, Volume 1, A-B	Smellie, William	1740-1795	encyclopedia	eng	None	844	3 v. (viii, 697, [1] p., LVIII leaves of plate...	London	... h
4	9929192893804340	First edition, 1773, Volume 2, C-L	Smellie, William	1740-1795	encyclopedia	eng	None	1032	3 v. (viii, 697, [1] p., LVIII leaves of plate...	London	... h
5	9929192893804340	First edition, 1773, Volume 3, M-Z	Smellie, William	1740-1795	encyclopedia	eng	None	864	3 v. (viii, 697, [1] p., LVIII leaves of plate...	London	... h

Subset of **metadata extracted** for the volumes of the **First Edition**. This edition is a 3-volume reference work, **issued twice, in 1771 and 1773**.

**New Post-processing scripts** to improve the previous results:

- Re-classification of **Terms** (articles and topics)
- Join terms split across pages



## The Journey

Phase 1: Information Extraction

**Phase 2: EB Ontology and Knowledge Graph**

Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

Phase 4: Defoe and Knowledge Graph

Phase 5: Flask Web-Application

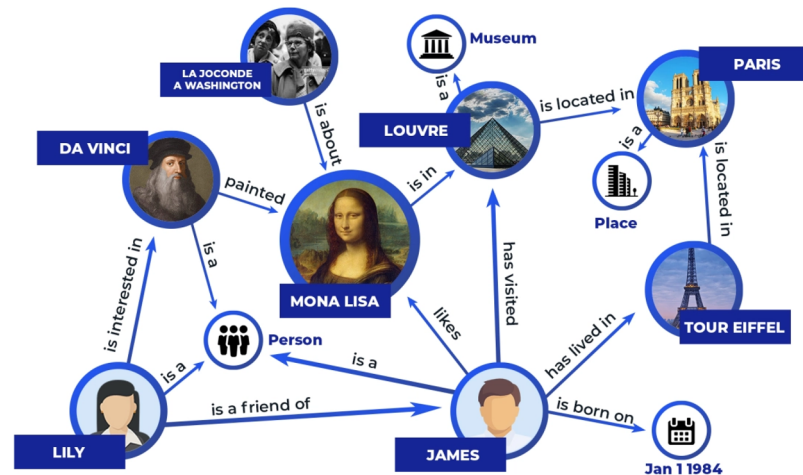
# Phase 2: EB Ontology and Knowledge Graph

## Background

KG: Incorporate human knowledge into intelligent systems, exploiting a semantic graph perspective

- A **knowledge graph** is a specialized graph or network of the things we want to describe and how they are related
- It is a **semantic** model since we want to capture and generate **meaning** with the model

"The application of graph processing and graph DBMSs will grow at 100 percent annually through 2022 to continuously accelerate data preparation and enable more complex and adaptive data science."  
– Gartner's *Top 10 Data and Analytics Technology Trends for 2019*



Aim: to capture a shareable and reusable knowledge representation of the EB.

EB Ontology → A formal description of knowledge as a set of concepts

In order to create and publish the EB-Ontology we used:

- diagrams.net : To create an UML with the EB information (classes, properties, relationships, etc.)
- Chowlk : To convert the UML into an OWL ontology
- Widoco: To publish and create an enriched and customized documentation of the ontology
- w3id.org: To configure my permanent Identifier for EB ontology

EB-Ontology available at:

- <https://github.com/francesNLP/EB-ontology>
- <https://w3id.org/eb/>

## EB Ontology

**This version:**

<https://w3id.org/eb/0.0.2>

**Latest version:**

<https://w3id.org/eb>

**Revision:**

0.0.2

**Authors:**

Rosa Filgueira

**Contributors:**

Daniel Garijo

**Download serialization:**

Format [JSON LD](#) Format [RDF/XML](#) Format [N Triples](#) Format [TTL](#)

**License:**

License [license name goes here](#)

**Visualization:**

Visualize with [WebVowl](#)

**Cite as:**

Rosa Filgueira. EB Ontology. Revision: 0.0.2. Retrieved from: <https://w3id.org/eb/0.0.2>

---

## Abstract

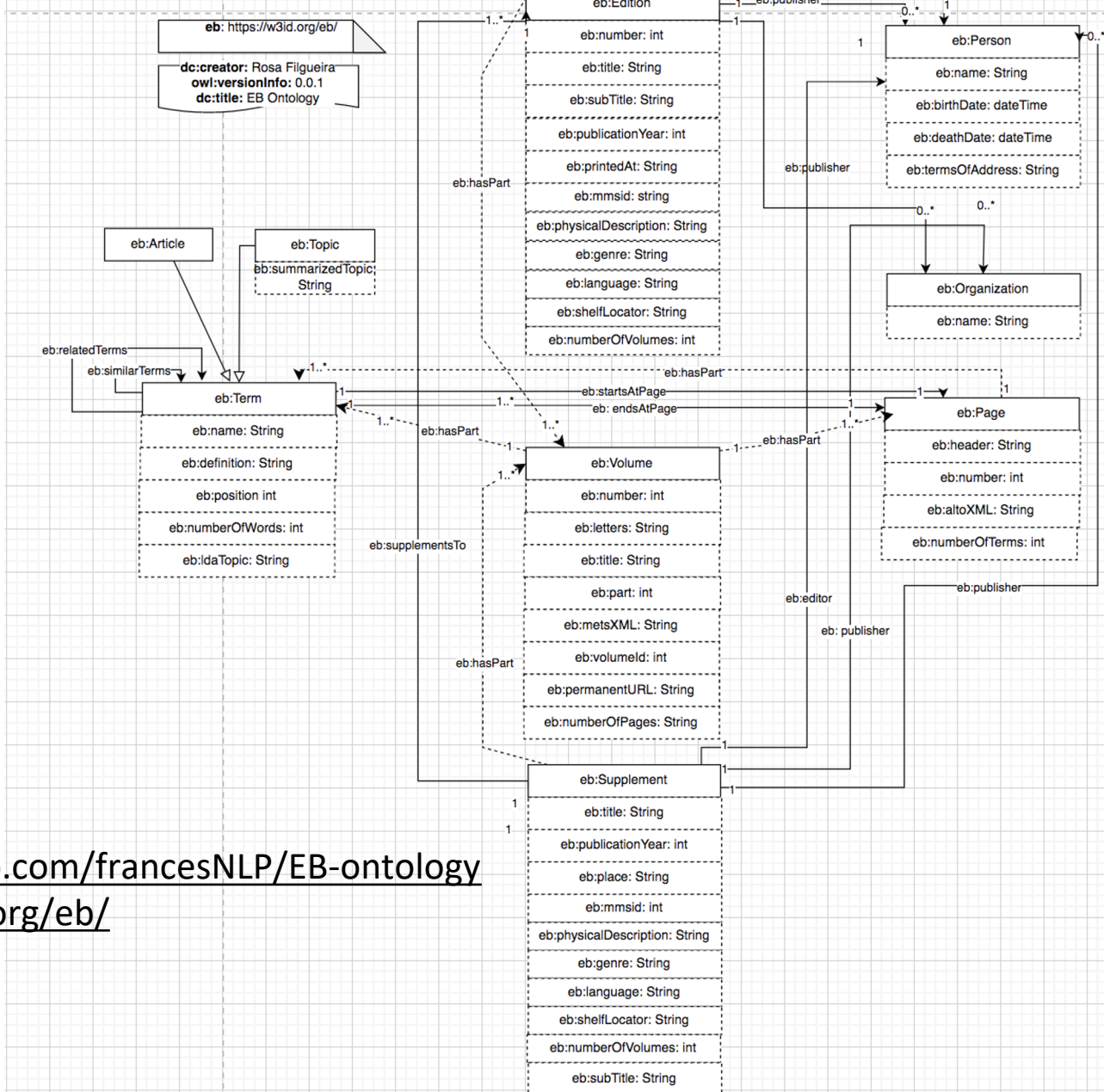
An ontology designed to represent volumes, terms and editions of the Encyclopaedia Britannica. The ontology extends Schema.org.

## Table of contents

- 1. [EB Ontology: Description](#)
- 2. [Cross reference for EB Ontology classes, properties and dataproperties](#)
  - 2.1. [Classes](#)
  - 2.2. [Object Properties](#)
  - 2.3. [Data Properties](#)
- 3. [Acknowledgments](#)

## 1. Introduction

This is a place holder text for the introduction. The introduction should briefly describe the ontology, its motivation, state of the art and goals.



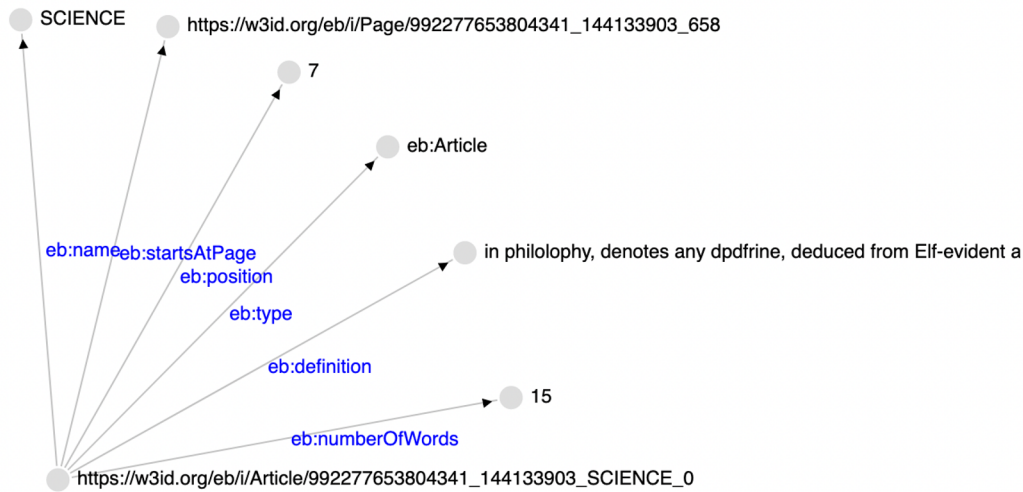
<https://w3id.org/eb/>



**EB Knowledge Graph:** Populating the post-processed information (extracted **Terms & Metadata**) into an RDF triple store + **EB Ontology:**

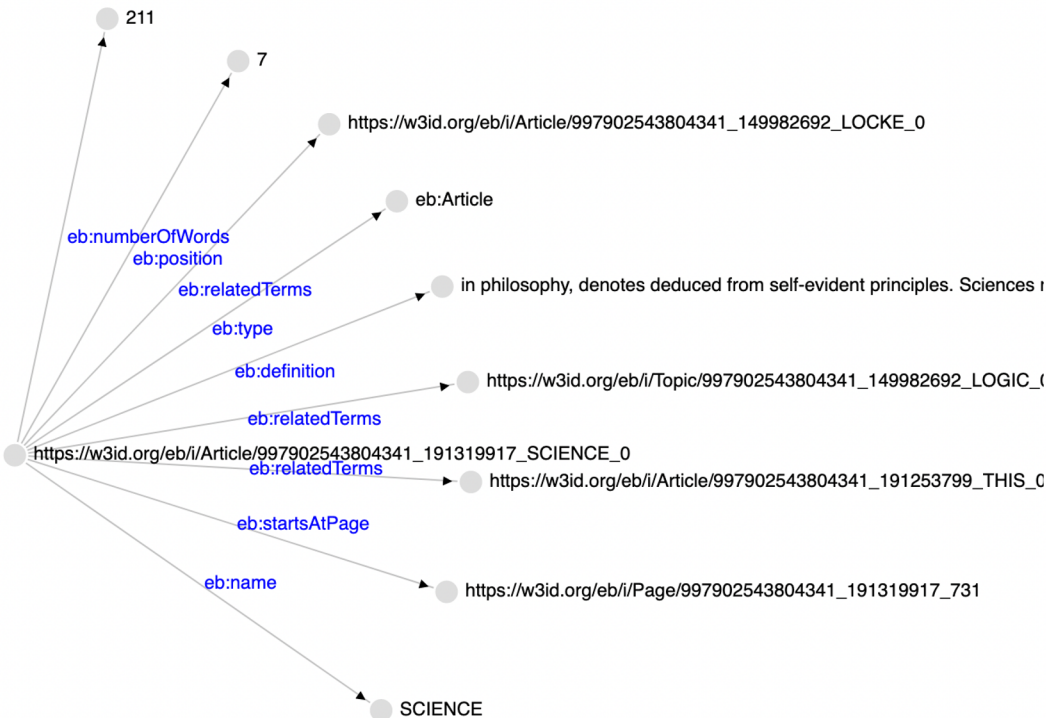
- Apache Jena FUSEKI SPARQL server

**NOTE:** Each Term, Edition, Page, Volume, etc ... is a **Resource** in our **Knowledge Graph** and has an **URI to identify it**.





- **Science** term in the First Edition issued in 1771

Edition 1 Year 1771  
 Edition 1 Year 1773  
 Edition 2 Year 1778  
 Edition 3 Year 1797  
 Edition 4 Year 1810  
 Edition 5 Year 1815  
 Edition 6 Year 1823  
 Edition 7 Year 1842  
 Edition 8 Year 1853



- **Science** term in the Third Edition issued in 1797

# Phase 2: EB Ontology and Knowledge Graph




UploadCommunities

There is a **newer version** of this record available.

June 21, 2022DatasetOpen Access

## EB-KG: Knowledge Graph of the first 8 editions Encyclopaedia Britannica (1768-1860)

 Rosa Filgueira

This Knowledge Graph represents the information of the first eight editions of Encyclopaedia Britannica (years: 1768 to 1860) in RDF (ttl format).

The raw dataset is provided by the NLS in this [link](#), and it comprises of eight editions and a total of 195 volumes with a total size of 44GB. It uses two XMLs schemas: METS for descriptive, structural, technical and administrative metadata (Title, Author, Publisher, etc); and ALTO for encoding the OCR text of a page.

In this work, we have extracted the information from METS and ALTO XMLS using [defoe](#) tool and developed [novel information extraction heuristics](#). With the extracted information, we created the EB-KG Knowledge Graph, which uses the [EB Ontology](#), to represent such information. Furthermore, during the information extraction phase, we have employed several techniques to mitigate two common OCR errors: long-S and the line-break hyphenation.

The EB-KG contains 1,638,239 RDF triples. It has information from 8 editions. Each edition can have several Volumes, references to Books, Supplements; it also has an Editor and a Publisher, which can be a Person or an Organization. A Volume has several Pages, which can contain several Terms. And a Term can be either a Topic (a term described across several pages, often combining text, pictures, and tables.) or an Article (a description of the term in one- or two-paragraph long text (similar to an entry in a dictionary)). The data model of the EB-KG can be found [here](#).

The original ALTO files do not indicate the start and end of each EB term, the first part of our work involved the automated extraction of all terms (along with their metadata) across editions, so they can be analysed independently without the surrounding text.

This work was performed during my 2021-2022 National Library of Scotland Digital Scholarship Fellowship.

## The Journey

Phase 1: Information Extraction

Phase 2: EB Ontology and Knowledge Graph

**Phase 3: Augmented Knowledge Graph with Deep Transfer Learning**

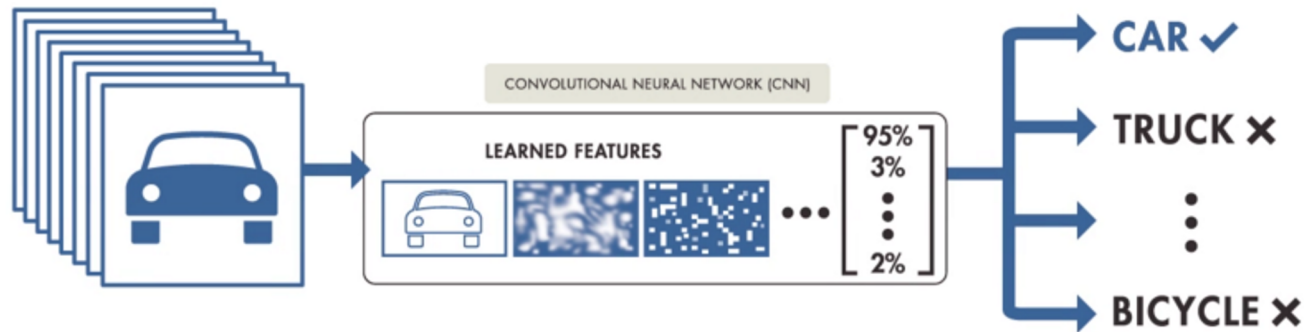
Phase 4: Defoe and Knowledge Graph

Phase 5: Flask Web-Application

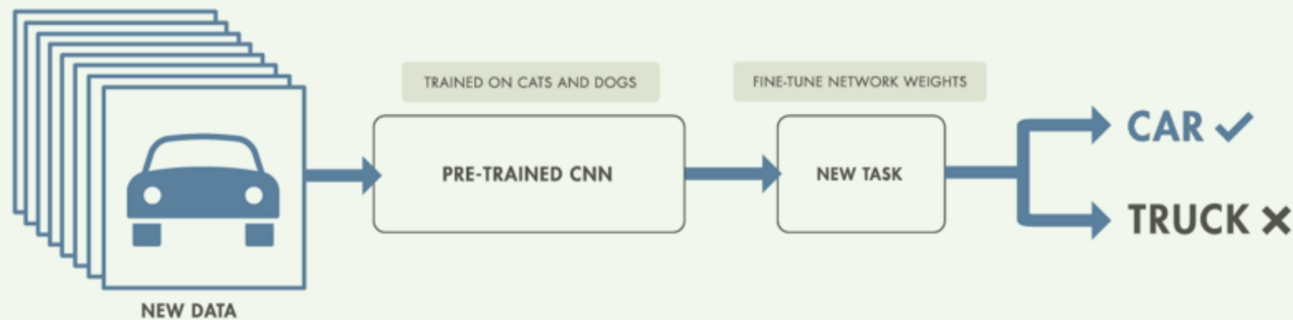
# Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

## Background

### TRAINING FROM SCRATCH



### TRANSFER LEARNING



We are going to use this approach for applying NLP/ML analysis



# Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

## Background



# Transformers

build **passing** license Apache-2.0 website **online** release v2.0.0

## A High-Level Look

Let's begin by looking at the model as a single black box. In a machine translation application, it would take a sentence in one language, and output its translation in another.



# Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

## Augmenting the EB-KG

**EB Knowledge Graph 2.0:** previous info + storing the result of applying different deep learning transformers analyses:

- **sentiment analyses:** Classifying text between positive and negative
  - **transformer:** siebert/sentiment-roberta-large-english
- **topic modelling:** Clustering terms into topics
  - **transformer:** all-mpnet-base-v2
- **term similarity:** Comparing text & semantic similarity
  - **transformer:** all-mpnet-base-v2
- **spelling checking:** Finding misspelling/ocr errors and fixing them
  - **transformer:** neuspell + *Elmosc lstmChecker*
- **summarization:** Summarizing the text of a topic term (XLNET)
  - **transformer:** XLNeT

**Note:** *Using pre-training models to calculate those analyses*

## Augmenting the EB-KG

### Example: Spelling Checking ☐ *Lewis* Term

#### Original Definition

the mort northerly of any of the w eften islands of Scotland, lying in 8\00b0 odd minutes W. long, and between 58\00b0 and 59 0 odd minutes N. lat.



#### Cleaned Definition

the most northerly of any of the w eastern islands of Scotland , lying in 8 and odd minutes W. long , and between 58 and and 59 0 odd minutes N. land .



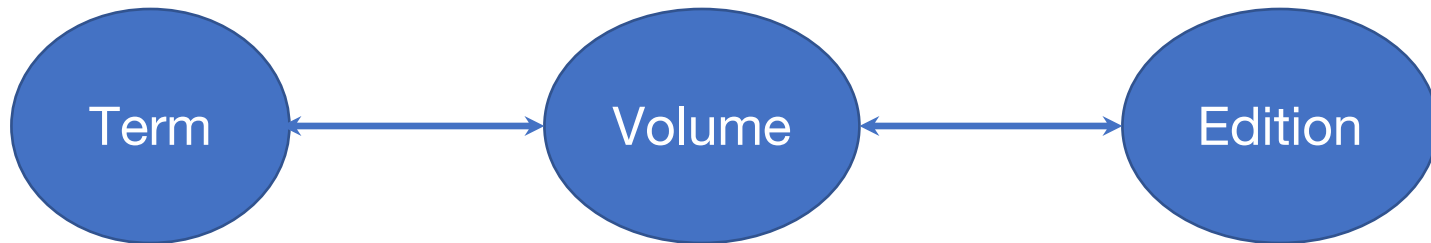
#### Compute Difference

the most northerly of any of the w eastern islands of Scotland , lying in 8\00b0 and odd minutes W. long , and between 58\00b0 and and 59 0 odd minutes N. latnd .

## Querying the Augmented EB-KG

Two types of “queries”:

- Type 1: **Extracting information from the EB Knowledge → SPARQL:**  
Navigating through the KG to get the desired information.



Example: Given a **Term** ( e.g., *Edinburgh*), I can get the **Edition** Information (e.g., *Edition Title*)

- Type 2: **Processing information from the EB Knowledge → *defoe*:**  
Processing further the definitions from the selected terms.
  - We needed to do some work on ***defoe*** first --> **Phase 4**



## The Journey

Phase 1: Information Extraction

Phase 2: EB Ontology and Knowledge Graph

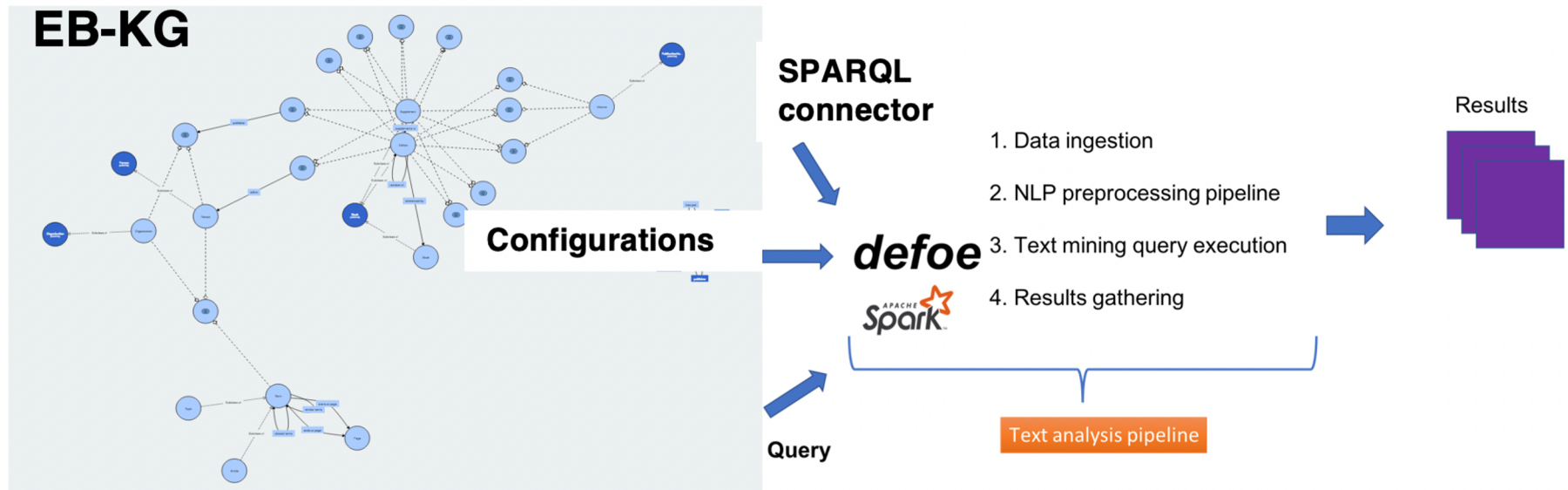
Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

**Phase 4: Defoe and Knowledge Graph**

Phase 5: Flask Web-Application

## defoe\_KG

- New **KG defoe connector** (based in SPARQL) to run *defoe* queries using the EB Knowledge Graph as a source of data → **defoe\_KG**



### New *defoe* queries

New *defoe* queries fully configurable: different filtering options, target, lexicon, etc.

- *frequency keysearch*: **Count number of terms** or **times** in which appear keywords or keysentences and group by years. Several filtering options.
- *term fulltext keysearch*: **Extract terms definitions** in which appear keywords or keysentences and group by years. Several filtering options.
- *term snippet keysearch*: **Extract snippets of definitions** in which appear keywords or keysentences and group by years. Several filtering options, including the snippet size.
- *publication normalization*: **Extract the number of documents, pages, words** per year.
- *uris keysearch*: **Extract uris of terms** in which appear keywords or keysentences and group by years. Several filtering options.
- *geoparser terms*: **Geoparsing the term definition** in which appear keywords or keysentences and group by years. Several filtering options.

## The Journey

Phase 1: Information Extraction

Phase 2: EB Ontology and Knowledge Graph

Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

Phase 4: Defoe and Knowledge Graph

**Phase 5: Flask Web-Application**



### Semantic Web

- Users DO NOT have to create SPARQL or *defoe* queries → the web-application does it for them → Abstractions to SPARQL & DEFOE
- The web-app runs both types of queries and visualizes the results

## Exploring the Encyclopaedia Britannica (1768-1860)

### Term Search

Enter the **term** that you would like to search for. In case that the **Term Type is a Topic**, only the **summary** of the definition is displayed. If not term is introduced, it will search for the first term in the Encyclopaedia.



## Exploring the Encyclopaedia Britannica (1768-1860)

### Term Search

Enter the **term** that you would like to search for. In case that the **Term Type is a Topic**, only the **summary** of the definition is displayed. If not term is introduced, it will search for the first term in the Encyclopaedia.



## Term Search

Results for **FLOWER**.

Note that if you click over an **URI** in this table, it will take you to the **Visualization of Resources page** . However, if you instead click over a **related term** , it will conduct a **term search**, showing all the searching results for that term. And if you click over a **topic model** it will take you to the **Topic Modelling page** , listing all the terms belonging to that particular topic model.

displaying 1 - 2 records in total 9

URI	Year	Edition	Volume	Start Page	End Page	Term Type	Definition/Summary	Related Terms	Topic Modelling	Sentiment_Score	Advanced Options
<a href="https://w3id.org/eb/i/Article/992277653804341_144133902_FLOWER_0">https://w3id.org/eb/i/Article/992277653804341_144133902_FLOWER_0</a>	1771	1	2	566	566	Article	among botanifts and gardeners, the most beautiful part of trees and plants, containing the organs or... <a href="#">More</a>		3190_pea_pease_pea_with_footstalk	POSITIVE_1.00	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850367_FLOWER_0">https://w3id.org/eb/i/Article/9929192893804340_144850367_FLOWER_0</a>	1773	1	2	574	574	Article	among botanifts and gardeners, the most beautiful part of trees and plants, containing the organs or... <a href="#">More</a>	BOTANY	3190_pea_pease_pea_with_footstalk	POSITIVE_1.00	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a>

- Results obtained when we search for Flower.
- Term found nine times across the eight editions
- Note that the First Edition was issued twice, in 1771 and 1773).

## Exploring the Encyclopaedia Britannica (1768-1860)

### Term Similarity

Enter an **URI of a term** or **some text** that you would like to search similar terms for. If not term is introduced, it will search for similar terms of the first term in the Encyclopaedia.





## Term Similarity

URI	Edition	Year	Volume	Term	Definition	Topic Modelling	Similitud Rank
<a href="https://w3id.org/eb//Article/992277653804341_144133902_FLOWER_0">https://w3id.org/eb//Article/992277653804341_144133902_FLOWER_0</a>	1	1771	2	FLOWER	among botanifts and gardeners, the most beautiful part of trees and plants, containing the organs or... <a href="#">More</a>	<a href="#">3190_pea_pease_pea_with_footstalk</a>	0.9776949
<a href="https://w3id.org/eb//Article/997902543804341_149981189_FLOWER_0">https://w3id.org/eb//Article/997902543804341_149981189_FLOWER_0</a>	3	1797	7	FLOWER	Flos, among botanifts and gardeners, the most beautiful part of trees and plants, containing the org... <a href="#">More</a>	<a href="#">-1_he_his_was_in</a>	0.81424284
<a href="https://w3id.org/eb//Article/9910796253804340_192015836_FLOWER_0">https://w3id.org/eb//Article/9910796253804340_192015836_FLOWER_0</a>	6	1823	8	FLOWER	Flos, among botanists and gardeners, the most beautiful part of trees and plants, containing the org... <a href="#">More</a>	<a href="#">-1_he_his_was_in</a>	0.8098287
<a href="https://w3id.org/eb//Article/9922270543804340_191678898_FLOWER_0">https://w3id.org/eb//Article/9922270543804340_191678898_FLOWER_0</a>	5	1815	8	FLOWER	Flos, among botanifts and gardeners, the molt beautiful part of trees and plants, containing the org... <a href="#">More</a>	<a href="#">-1_he_his_was_in</a>	0.8073469
<a href="https://w3id.org/eb//Article/9929192893804340_144850367_LILIADEOUS_0">https://w3id.org/eb//Article/9929192893804340_144850367_LILIADEOUS_0</a>	1	1773	2	LILIADEOUS	an appellation given to such flowers as resemble that of the lily.	<a href="#">3798_mint_horehound_horehound_and_in_rings_the_lily</a>	0.704466
<a href="https://w3id.org/eb//Article/992277653804341_144133902_LILIADEOUS_0">https://w3id.org/eb//Article/992277653804341_144133902_LILIADEOUS_0</a>	1	1771	2	LILIADEOUS	an appellation given to such flowers as resemble that of the lily.	<a href="#">3798_mint_horehound_horehound_and_in_rings_the_lily</a>	0.704466
<a href="https://w3id.org/eb//Article/9910796253804340_192692756_CALYX_0">https://w3id.org/eb//Article/9910796253804340_192692756_CALYX_0</a>	6	1823	5	CALYX	among botanists, a general term, expressing the cup of a flower*, or that part of a plant which surr... <a href="#">More</a>	<a href="#">226_botanifts_among_botanifts_of_flower_petals</a>	0.68875444
<a href="https://w3id.org/eb//Article/997902523804341_144850377_LILIACEOUS_0">https://w3id.org/eb//Article/997902523804341_144850377_LILIACEOUS_0</a>	2	1778	6	LILIACEOUS	in botany, an appellation given to such flowers as referable those of the lily.	<a href="#">3798_mint_horehound_horehound_and_in_rings_the_lily</a>	0.68756175

# Phase 5: Flask Web-Application

## Term Similarity

Enter an **URI of a term** or **some text** that you would like to search similar terms for. If not term is introduced, it will search for similar terms of the first term in the Encyclopaedia.

person who does scientific experiments



URI	Edition	Year	Volume	Term	Definition	Topic Modelling	Similitud Rank
<a href="https://w3id.org/eb/i/Article/997902523804341_144850379_PHYSICIAN_0">https://w3id.org/eb/i/Article/997902523804341_144850379_PHYSICIAN_0</a>	2	1778	8	PHYSICIAN	a person who profesles medicine, or the art of healing diseases. See Medicine. PHYSICS, or NA Tu RAL... <a href="#">More</a>	-1_he_his_was_in	0.6265975
<a href="https://w3id.org/eb/i/Article/9910796253804340_192015835_ARMORIST_0">https://w3id.org/eb/i/Article/9910796253804340_192015835_ARMORIST_0</a>	6	1823	2	ARMORIST	a person skilled in the knowledge of	3715_knowledge_acquired_any_teacher_use_without_teacher_mr	0.58429337
<a href="https://w3id.org/eb/i/Article/992277653804341_144133901_BOTANIST_0">https://w3id.org/eb/i/Article/992277653804341_144133901_BOTANIST_0</a>	1	1771	1	BOTANIST	a person skilled in botany. See Bo-	-1_he_his_was_in	0.56840813
<a href="https://w3id.org/eb/i/Article/9910796273804340_193469092_PROFESSOR_0">https://w3id.org/eb/i/Article/9910796273804340_193469092_PROFESSOR_0</a>	7	1842	18	PROFESSOR	in the universities, a person who teaches or reads public lectures in some art or science. See Uni	441_attends_on_teaches_or_reads_public_or_reads	0.5636317

## Exploring the Encyclopaedia Britannica (1768-1860)

### Topic Modelling Search

Enter a **topic modelling name** or just the **number** of a topic to see all the terms within the same topic. All topics modelling names start with a number. If not topic is introduced, it will use the first topic modelling, which name starts with '0'.

We have **4306** topics modelling items.





# Exploring the Encyclopaedia Britannica (1768-1860)

## Topic Modelling Search

Enter a **topic modelling name** or just the **number** of a topic to see all the terms within the same topic. All topics modelling names start with a number. If not topic is introduced, it will use the first topic modelling, which name starts with '0'.

We have **4306** topics modelling items.

3190\_pea\_pease\_pea\_with\_footstalk

Q

12 terms found for the topic 3190\_pea\_pease\_pea\_with\_footstalk

Note that if you click over an **URI** in this table, it will take you to the **Visualization of Resources** page . However, if you instead click over a **term** , it will take you to the **Term Search** page, showing all the searching results for that term.

displaying 1 - 10 records in total 12

					the organs or... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/9929777383804340_193819046_PEA_0">https://w3id.org/eb/i/Article/9929777383804340_193819046_PEA_0</a>	8	1853	17	<a href="#">PEA</a>	the English name applied to the seed of several leguminous plants, but chiefly to those of the culti... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850367_CHICKLING_0">https://w3id.org/eb/i/Article/9929192893804340_144850367_CHICKLING_0</a>	1	1773	2	<a href="#">CHICKLING</a>	pea, in botany, a name given to the lathyrus. See Lathyrus s. ' C HI CUITO, or Cuyo, a province o... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/997902543804341_192200061_PISUM_0">https://w3id.org/eb/i/Article/997902543804341_192200061_PISUM_0</a>	3	1797	14	<a href="#">PISUM</a>	PEASE ; a genus of the belonging to the diadelphia class of cies are, 1. The fativum, or greater low... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/997902523804341_144850375_EVERLASTING_0">https://w3id.org/eb/i/Article/997902523804341_144850375_EVERLASTING_0</a>	2	1778	4	<a href="#">EVERLASTING</a>	pea, a genus of plants, otherwise called lathyrus. See Lathyrus. EVESHAM, a borough-town of Worcefte... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/997902523804341_144850379_PISUM_0">https://w3id.org/eb/i/Article/997902523804341_144850379_PISUM_0</a>	2	1778	8	<a href="#">PISUM</a>	pease ; a genus of the deeadndria order, belonging to the diadelphia class of plants. The species are... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850366_ANTHEMIS_0">https://w3id.org/eb/i/Article/9929192893804340_144850366_ANTHEMIS_0</a>	1	1773	1	<a href="#">ANTHEMIS</a>	or Camomile, in botany,, a genus of ^ the fyngenefia polygamia superflua class. The receptacle of th... <a href="#">More</a>

## Topic Modelling

12 terms found for the topic 3190\_pea\_pease\_pea\_with\_footstalk

Note that if you click over an **URI** in this table, it will take you to the **Visualization of Resources page** . However, if you instead click over a **term** , it will take **Search page**, showing all the searching results for that term.

displaying 1 - 10 records in total 12

« 1 2 »

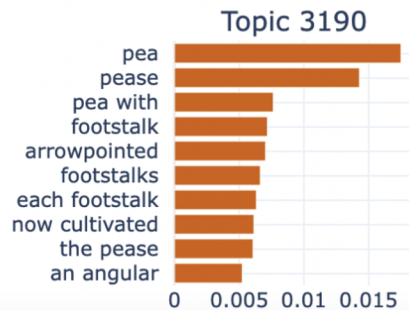
URI	Edition	Year	Volume	Term	Definition
<a href="https://w3id.org/eb/i/Article/9922270543804340_192692638_PISUM_0">https://w3id.org/eb/i/Article/9922270543804340_192692638_PISUM_0</a>	5	1815	16	PISUM	Pease; a genus of plants belonging diadelphia class. See Botany Index. The l. The fativum, or greate... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/992277653804341_144133902_FLOWER_0">https://w3id.org/eb/i/Article/992277653804341_144133902_FLOWER_0</a>	1	1771	2	FLOWER	among botanifts and gardeners, the most beautiful part of trees and plants, containing the organs or... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/9929777383804340_193819046_PEA_0">https://w3id.org/eb/i/Article/9929777383804340_193819046_PEA_0</a>	8	1853	17	PEA	the English name applied to the seed of several leguminous plants, but chiefly to those of the culti... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850367_CHICKLING_0">https://w3id.org/eb/i/Article/9929192893804340_144850367_CHICKLING_0</a>	1	1773	2	CHICKLING	pea, in botany, a name given to the lathyrus. See Lathyrus s. ' C HI CUITO, or Cuyo, a province o... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/997902543804341_192200061_PISUM_0">https://w3id.org/eb/i/Article/997902543804341_192200061_PISUM_0</a>	3	1797	14	PISUM	PEASE ; a genus of the belonging to the diadelphia class of cies are, 1. The fativum, or greater low... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/997902523804341_144850375_EVERLASTING_0">https://w3id.org/eb/i/Article/997902523804341_144850375_EVERLASTING_0</a>	2	1778	4	EVERLASTING	pea, a genus of plants, otherwise called lathyrus. See Lathyrus. EVESHAM, a borough-town of Worcefte... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/997902523804341_144850379_PISUM_0">https://w3id.org/eb/i/Article/997902523804341_144850379_PISUM_0</a>	2	1778	8	PISUM	pease ; a genus of the deeandria order, belonging to the diadelphia class of plants. The species are... <a href="#">More</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850366_ANTHEMIS_0">https://w3id.org/eb/i/Article/9929192893804340_144850366_ANTHEMIS_0</a>	1	1773	1	ANTHEMIS	or Camomile, in botany,, a genus of ^ the fyngenefia polygamia fuperflua class. The receptacle of th... <a href="#">More</a>



### Topic Modelling Visualizations

We are visualizing the topic model **3190\_pea\_pease\_pea\_with\_footstalk**. For each topic model, we plot their most common 10 words , along with their [c-TF-IDF scores](#) .

### Topic Word Scores



## Spell Checker

Enter an **URI of a term** that you would like to check its spelling. If not term is introduced, it will check the spelling of the first term in the Encyclopaedia.

URI	Edition	Year	Volume	Term
<a href="https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1">https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1</a>	1	1771	1	ABACUS

### Original Definition

in architeflure, signifies the superior part or member of the capital of a column, and serves as a kind of crowning to both. It was originally intended to represent a square tile covering a basket. The form of the abacus is not the same in all orders: in the Tuscan, Doric, and Ionic, it\u2018is generally square; but in the Corinthian and Compofite, its four sides are arched ir Avards, and embellilhed in the middle withornament, as a rose or other flower, Scammozzi uses abacus for a concave moulding on the capital of the Tuscan pedefial; and Palladio calls the plinth above the echinus, or bouldin, in the Tufean and Doric orders, by the same name. See plate I. fig. i. and

### Cleaned Definition

in architecture , signifies the superior part or member of the capital of a column , and serves as a kind of crowning to both . It was originally intended to represent a square tile covering a basket . The form of the abacus is not the same in all orders : in the Tuscan , Doric , and Ionic , its generally square ; but in the Corinthian and Compofite , its four sides are arched in Avards , and embellished in the middle withornament , as a rose or other flower , Scammozzi uses abacus for a concave moulding on the capital of the Tuscan pedestal ; and Palladio calls the plinth above the echinus , or bouldin , in the Tufean and Doric orders , by the same name . See plate I. fig . i. and



### Cleaned Definition

in architecture , signifies the superior part or member of the capital of a column , and serves as a kind of crowning to both . It was originally intended to represent a square tile covering a basket . The form of the abacus is not the same in all orders : in the Tuscan , Doric , and Ionic , it\u2018is generally square ; but in the Corinthian and Compofite , its four sides are arched in Avards , and embellished in the middle withornament , as a rose or other flower , Scammozzi uses abacus for a concave moulding on the capital of the Tuscan pedestal ; and Palladio calls the plinth above the echinus , or bouldin , in the Tufean and Doric orders , by the same name . See plate I. fig . i. and

Select Edition

Edition 1 Year 1771

Select Volume

1 A-B

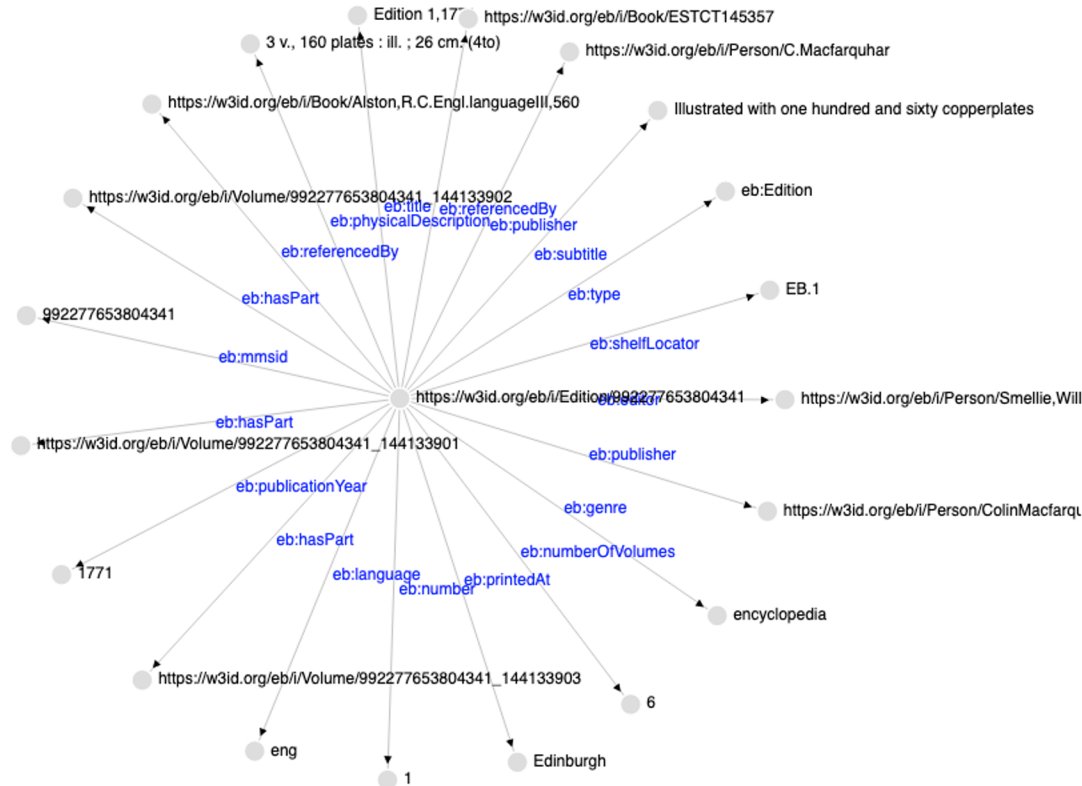


Edition	Details
Year	1771
Edition Number	1
Edition URI	< <a href="https://w3id.org/eb/I/Edition/992277653804341">https://w3id.org/eb/I/Edition/992277653804341</a> >
Edition Title	Edition 1,1771
Edition Subtitle	Illustrated with one hundred and sixty copperplates
Printed at	Edinburgh
Physical Description	3 v., 160 plates : ill. ; 26 cm. (4to)
MMSID	992277653804341
Shelf Locator	EB.1
Genre	encyclopedia
Language	English
Number of Volumes	3

Volume	Details
Volume Number	1
Volume URI	< <a href="https://w3id.org/eb/I/Volume/992277653804341_144133901">https://w3id.org/eb/I/Volume/992277653804341_144133901</a> >
Volume Title	Encyclopaedia Britannica; or, A dictionary of arts and sciences, compiled upon a new plan
Volume Letters	A-B
Volume Permanent URL	<a href="https://digital.nls.uk/144133901">https://digital.nls.uk/144133901</a>
Volume Number of Pages	832

Volume	Statistics
Number of Articles	2505
Number of Topics	15
Number of Distinct Articles	2440
Number of Distinct Topics	15

## Visualization of Resources



Visualization of the First Edition (issued in 1771) resource stored in the EB-KG

## Defoe Queries – Text Mining Platform

[Term Search](#) [Term Similarity](#) [Topic Modelling](#) [Spell Checker](#) [Defoe Queries](#) [EB Details](#) [Visualization of Resources](#)

### Exploring the Encyclopaedia Britannica (1768-1860)

#### Defoe Queries

Select the **defoe query** that you would to run, and introduce your configurations:

**Defoe query**

It counts the number of terms/words in which keywords/keysentences appear. It groups results by years.

**frequency\_keysearch\_by\_year:** it counts number of terms or times in which appear keywords or keysentences and groups the results by years

**publication\_normalized:** it extracts the number of documents, pages, and words per year

**uris\_keysearch:** it extracts uris of terms in which appear keywords or keysentences and groups the results by years

**terms\_snippet\_keysearch\_by\_year:** it extracts snippets of definitions in which appear keywords or keysentences and groups the results by years

**terms\_fulltext\_keysearch\_by\_year:** it extracts terms definitions in which appear keywords or keysentences and groups the results by years

**terms\_geoparser:** it geoparsers the term definition in which appear keywords or keysentences and groups the results by years



## Defoe Queries – Text Mining Platform

### Defoe Queries

Select the **defoe query** that you would to run, and introduce your configurations:

**Defoe query** terms\_snippet\_keysearch\_by\_year ▾

It extracts snippets of terms definitions in which appear your selected keywords/keysentences grouping results by years.

#### Preprocess Treatment

- ☐ None
- ☒ **Normalize**
- ☐ Normalize & Numbers
- ☐ Normalize & Lemmatize
- ☐ Normalize & Stemming

It converts all words to lower-case removing all characters that are not 'a',..., 'z'.

#### Upload your lexicon file

Browse... sc\_philosophers.txt



● ● ● sc\_philosophers.txt  
Francis Hutcheson  
David Hume  
Adam Smith  
Dugald Stewart  
Thomas Reid

The file should contain a line per keyword and/or keysentence that you want to use in your query.

#### Filtering Options

**Introduce Target words/sentences separated by "," (Optional)**

Those are the list of words and/or sentences that must appear in a term in order to select it. Leave it empty to select all.

**Select term which contains**

- ☒ **Any of the target words/sentences**
- ☐ All the target words/sentences

**Start Year (Optional)**  **End Year (Optional)**

**Snippet Window**

Submit Defoe query

Configuring a query

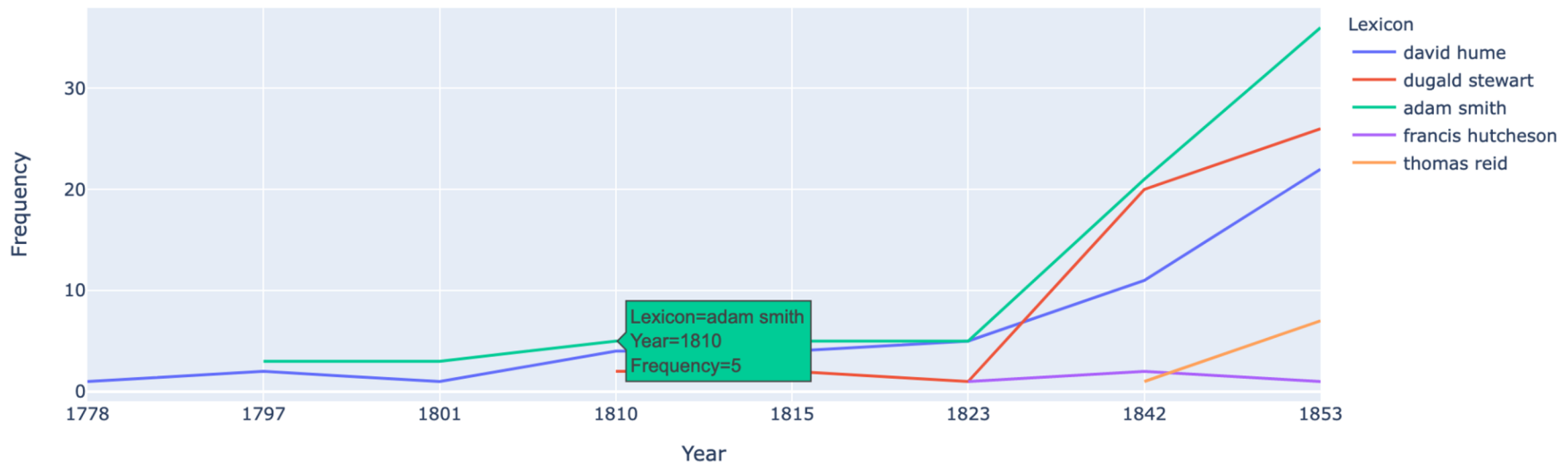
# Defoe Queries – Text Mining Platform

Year	Data									
1778	archive_filename	edition	header	keysearch-term	letters	page number	part	snippet	uri	volume
	144850375-mets.xml	2	DOU	david hume	D-F	179	None	his admirers are too fond of repeating to his advantage david hume of godfcroft an author of uncommon merit and an	<a href="https://w3id.org/eb//Article/9910796233804341_144850375_DOUGLAS_0">https://w3id.org/eb//Article/9910796233804341_144850375_DOUGLAS_0</a>	4
Year	Data									
1810	archive_filename	edition	header	keysearch-term	letters	page number	part	snippet	uri	volume
	191678900-mets.xml	4	WAR	david hume	SUI-THE	315	1	and published them in a pamphlet entitled remarks on mr david humes natural history of religion by a gentleman of cambridge	<a href="https://w3id.org/eb//Article/9910796233804340_191678901_A_4">https://w3id.org/eb//Article/9910796233804340_191678901_A_4</a>	20
	192692192-mets.xml	4	HOIHOI	adam smith	GOT-Heraldry	221	1	vicechancellor of the university latter end of the year dr adam smithsletter containing death of mr david hume i publicly animadverted	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1">https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1</a>	10
	192692192-mets.xml	4	HOIHOI	adam smith	GOT-Heraldry	221	1	smithsletter containing death of mr david hume i publicly animadverted adam smith i l d on the sophy of his friend	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1">https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1</a>	10
	192692192-mets.xml	4	HOIHOI	adam smith	GOT-Heraldry	221	1	duty made him universally beloved lifed letters on infidelity similar adam smith the books against his ridicule are an apology of	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1">https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1</a>	10
	192692192-mets.xml	4	HOIHOI	david hume	GOT-Heraldry	221	1	of the year dr adam smithsletter containing death of mr david hume i publicly animadverted adam smith i l d on	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1">https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1</a>	10
	192692192-mets.xml	4	HOIHOI	david hume	GOT-Heraldry	221	1	smith i l d on the sophy of his friend david hume people called christians in aofidities of the scotch philosophers	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1">https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1</a>	10
	192692192-mets.xml	4	HOIHOI	david hume	GOT-Heraldry	221	1	smith the books against his ridicule are an apology of david hume esq j humes ral religion an effay on filicide	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1">https://w3id.org/eb//Article/9910796233804340_192692193_HORNE_1</a>	10
	193108316-mets.xml	4	SMISMI	adam smith	SCR-SLE	17	1	and causes of the nations was the only son of adam smith of the customs at kirkaldy and of margaret daughter	<a href="https://w3id.org/eb//Article/9910796233804340_193108317_SMITH_2">https://w3id.org/eb//Article/9910796233804340_193108317_SMITH_2</a>	19
	191320558-mets.xml	4	POB	dugald stewart	Projectiles-RHI	221	2	of the preceding theory it may be added that professor dugald stewart in a paper read a considerable time ago before	<a href="https://w3id.org/eb//Article/9910796233804340_191320559_EO_0">https://w3id.org/eb//Article/9910796233804340_191320559_EO_0</a>	17
	191678900-mets.xml	4	Empty	david hume	SUI-THE	405	1	became acquainted also in the course of his education with david hume and dr adam ferguson o before he completed his	<a href="https://w3id.org/eb//Article/9910796233804340_191678901_WILKIE_0">https://w3id.org/eb//Article/9910796233804340_191678901_WILKIE_0</a>	20
	191678902-mets.xml	4	M	adam smith	Mathematics-Medicine	399	1	with feme strictures on the selfcontradidions historical errors of dr adam smith for the abolitor of the said company this time	<a href="https://w3id.org/eb//Article/9910796233804340_191678903_MICKLE_0">https://w3id.org/eb//Article/9910796233804340_191678903_MICKLE_0</a>	13
	192692192-mets.xml	4	HOMHOM	david hume	GOT-Heraldry	195	1	of p julofoficial neceffity deed been taught by plobbes celebrated david hume were proffed infidels or excited as coming from ked	<a href="https://w3id.org/eb//Article/9910796233804340_192692193_HOME_0">https://w3id.org/eb//Article/9910796233804340_192692193_HOME_0</a>	10
	191678902-mets.xml	4	METAPHYSICS	dugald stewart	Mathematics-Medicine	164	1	excellent observations on the common dodrine concerning abftradtion by professor dugald stewart of edinburgh elements of the philosophy of the human	<a href="https://w3id.org/eb//Topic/9910796233804340_191678903_METAPHYSICS_0">https://w3id.org/eb//Topic/9910796233804340_191678903_METAPHYSICS_0</a>	13
	191320558-mets.xml	4		adam smith	Projectiles-RHI	108	2	approbation of thinking men and has gradually superseded all others adam smith being professor in the first commercial city of scotland	<a href="https://w3id.org/eb//Topic/9910796233804340_191320559_POLITICAL_0">https://w3id.org/eb//Topic/9910796233804340_191320559_POLITICAL_0</a>	17
	191679019-mets.xml	4	PT	adam smith	HYD-JUN	191	1	that and some other works upon the same principles dr adam smith observes still continue to make an impression upon many	<a href="https://w3id.org/eb//Topic/9910796233804340_191679020_PART_2">https://w3id.org/eb//Topic/9910796233804340_191679020_PART_2</a>	11

Results of the **terms\_snippet\_keysearch\_by\_year** query

### Defoe Queries – Text Mining Platform

Frequency of Lexicon Terms per Years



frequency n-gram obtained after running the  
**frequency\_keyseach\_by\_year** query using the same  
configurations  
as in the previous query

## Defoe Queries – Text Mining Platform

[Term Search](#) [Term Similarity](#) [Topic Modelling](#) [Spell Checker](#) [Defoe Queries](#) [EB Details](#) [Visualization of Resources](#)

### Exploring the Encyclopaedia Britannica (1768-1860)

#### Results of publication\_normalized query

Here we can see the number of volumes, pages and words per year.

Year	Num. Volumes	Num. Pages	Num. Words
1771	3	2722	2587557
1773	3	2740	2581066
1778	10	9448	10218591
1797	11	9512	10495573
1801	2	1740	1943036
1803	2	1760	1956641
1810	40	17662	18876327
1815	20	17470	18764604
1823	20	17546	18742386
1824	15	13436	12258062
1842	22	19532	22407301
1853	22	19628	23475295

[Download Results](#)

[Return](#)

### Results of publication\_normalization query

[Term Search](#)[Term Similarity](#)[Topic Modelling](#)[Spell Checker](#)[Defoe Queries](#)[EB Details](#)[Visualization of Resources](#)

### Exploring the Encyclopaedia Britannica (1768-1860)

#### Results of publication\_normalized query

Here we can see the number of volumes, pages and words per year.

1st edition was published on years: [1771, 1773]  
2nd edition was published on years: [1778]  
3rd edition was published on years: [1797, 1801, 1803]  
4th edition was published on years: [1810, 1824]  
5th edition was published on years: [1815, 1824]  
6th edition was published on years: [1823, 1824]  
7th edition was published on years: [1842]  
8th edition was published on years: [1853]

Year	Num. Volumes	Num. Pages	Num. Words
1771	3	2722	2587557
1773	3	2740	2581066
1778	10	9448	10218591
1797	11	9512	10495573
1801	2	1740	1943036
1803	2	1760	1956641
1810	40	17662	18876327
1815	20	17470	18764604
1823	20	17546	18742386
1824	15	13436	12258062
1842	22	19532	22407301
1853	22	19628	23475295

[Download Results](#)[Return](#)

### Results of publication\_normalization query



## Defoe Queries – Text Mining Platform

### Results of frequency\_keysearch\_by\_year query

Your configurations:

- **preprocess:** normalize
- **target\_filter:** or
- **start\_year:** 1771
- **end\_year:** 1773
- **hit\_count:** word
- **data :** sport.txt

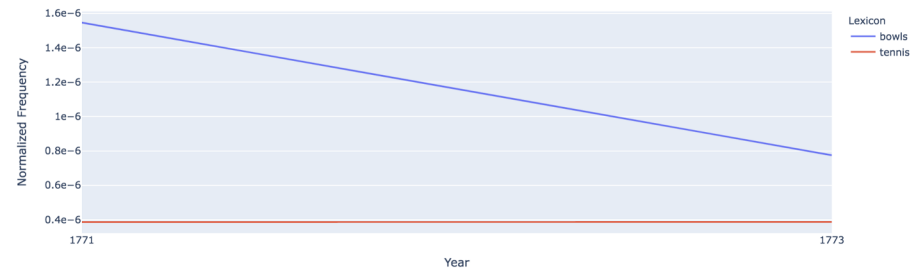
Here we can see the frequency of your lexicon words/sentences taking into acc

Year	Keyword : Frequency
1771	<ul style="list-style-type: none"><li>• bowls : 4</li><li>• tennis : 1</li></ul>
1773	<ul style="list-style-type: none"><li>• bowls : 2</li><li>• tennis : 1</li></ul>

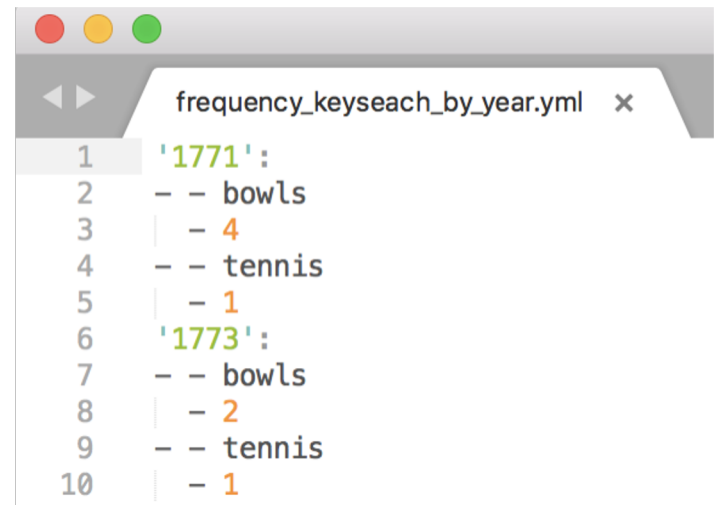
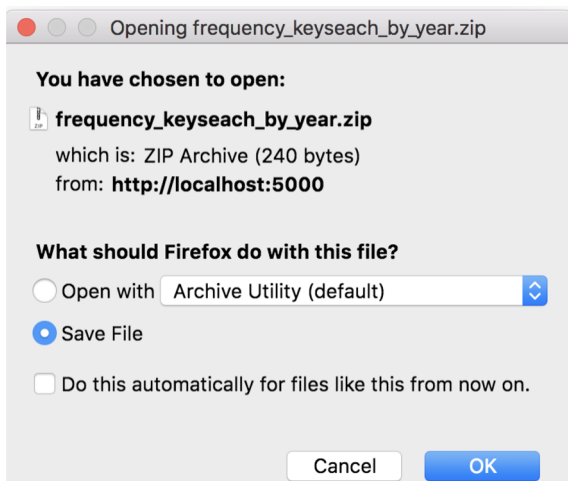
Frequency of Lexicon Terms per Years



Normalized Frequency of Lexicon Terms per Years



Download Results





## Text Mining Platform

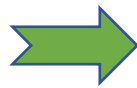
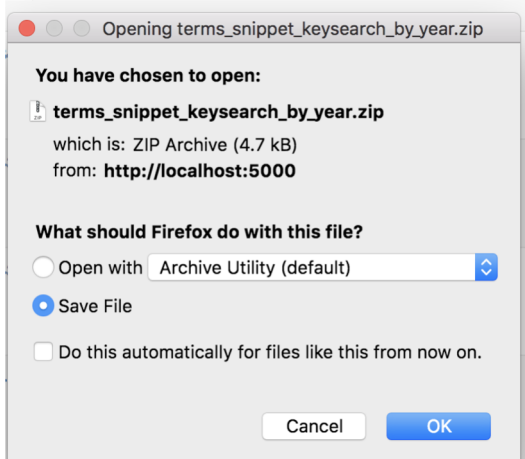
### Results of terms\_snippet\_keysearch\_by\_year query

Your configurations:

- preprocess: normalize
- target\_filter: or
- window: 15
- start\_year: 1771
- end\_year: 1773
- data : sport.txt

The full results are available for download in the below link . Here we list the terms founded in your search. For each those terms, you can see which words or sentences have been found from your lexicon, taking into account the target words/sentences (if any) and as well other configurations.

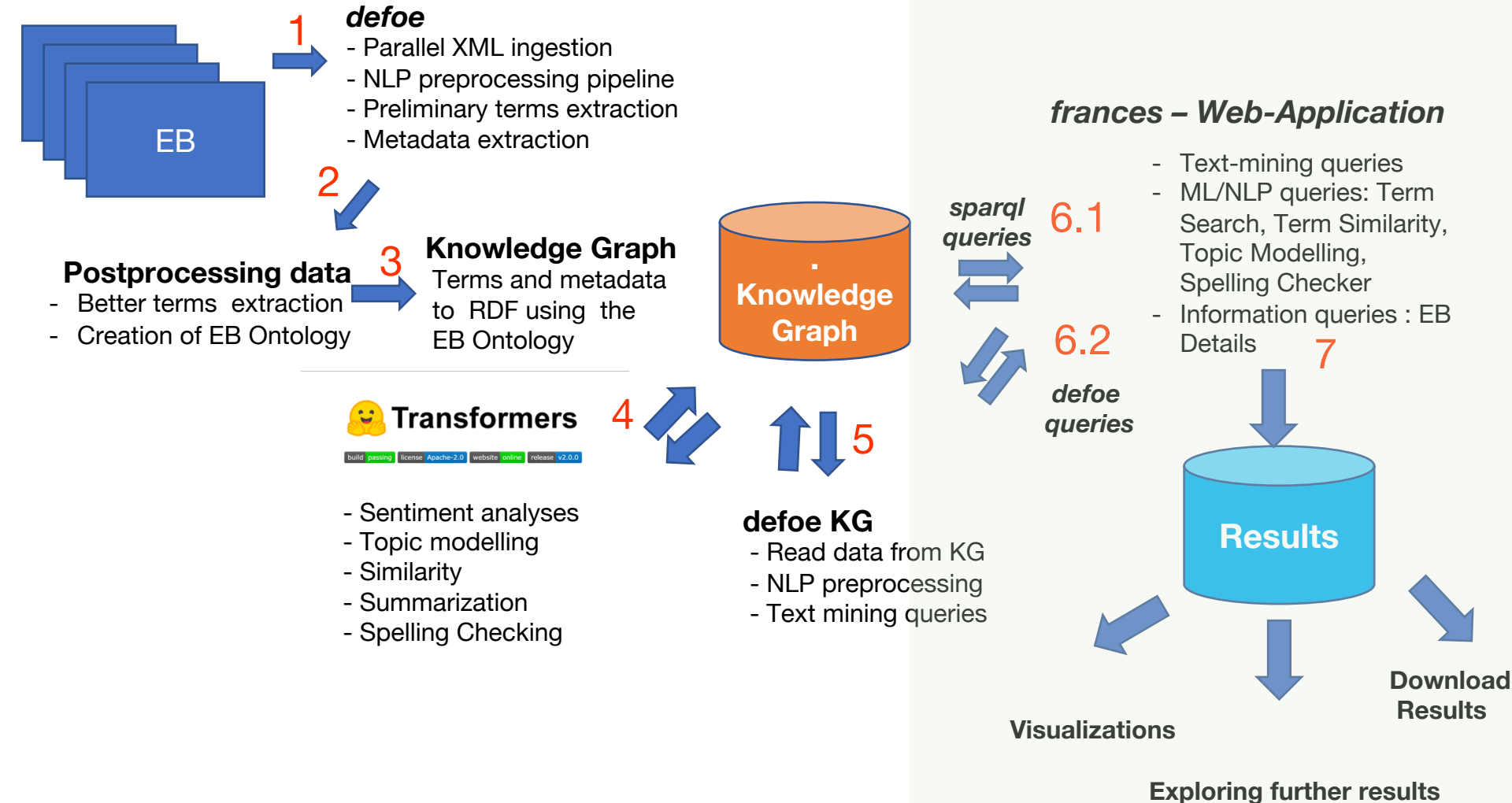
Year	Data	Advanced Options
<a href="https://w3id.org/eb/i/Article/992277653804341_144133902_CANDLESTICK_0">https://w3id.org/eb/i/Article/992277653804341_144133902_CANDLESTICK_0</a>	• bowls	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a> <a href="#">Term Evolution</a>
<a href="https://w3id.org/eb/i/Article/992277653804341_144133902_CANDYING_0">https://w3id.org/eb/i/Article/992277653804341_144133902_CANDYING_0</a>		
<a href="https://w3id.org/eb/i/Article/992277653804341_144133901_BOWL_0">https://w3id.org/eb/i/Article/992277653804341_144133901_BOWL_0</a>		
<a href="https://w3id.org/eb/i/Article/992277653804341_144133901_BOWLING_0">https://w3id.org/eb/i/Article/992277653804341_144133901_BOWLING_0</a>		
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWLING_0">https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWLING_0</a>		



```
terms_snippet_keysearch_by_year.yml x
1 |terms_details:
2 |-- '1773'
3 |-- archive_filename: 144850366-mets.xml
4 |  edition: '1'
5 |  header: BOWBOW
6 |  keysearch-term: bowls
7 |  letters: A-B
8 |  page number: '796'
9 |  part: None
10 |  snippet: bowling the art of playing at bowls the first thing to be observed in
11 |           bowling is the right chufing your bowl which
12 |  term: BOWLING
13 |  title: Edition 1,1773
14 |  uri: https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWLING_0
15 |  volume: '1'
16 |-- archive_filename: 144850366-mets.xml
17 |  edition: '1'
18 |  header: BOWBOW
19 |  keysearch-term: bowls
20 |  letters: A-B
21 |  page number: '796'
22 |  part: None
23 |  snippet: wood for the use of bowling or a vessel of capacity wherein to hold
24 |           li bowls and buckets of wood imported pay a duty f prgvi tb dozen whereof
25 |           sjd is
26 |  term: BOWL
27 |  title: Edition 1,1773
28 |  uri: https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWL_0
29 |  volume: '1'
30 |-- archive_filename: 144850368-mets.xml
31 |  edition: '1'
32 |  header: METAPHYSICS
33 |  keysearch-term: tennis
34 |  letters: M-Z
35 |  page number: '199'
36 |  part: None
37 |  snippet: there may be will there may be volition where there is no liberty thus
38 |           a tennis ball whether in motion by the stroke of a racket or lying till at
39 |           rest
40 |  term: METAPHYSICS
41 |  title: Edition 1,1773
42 |  uri: https://w3id.org/eb/i/Topic/9929192893804340_144850368_METAPHYSICS_0
43 |  volume: '3'
44 |-- '1771'
45 |-- archive_filename: 144133901-mets.xml
46 |  edition: '1'
47 |  header: GBOW
48 |  keysearch-term: bowls
49 |  letters: A-B
50 |  page number: '788'
51 |  part: None
52 |  snippet: wood for the use of bowling or a vessel of capacity wherein to hold
53 |           li bowls and buckets of wood imported pay a duty of qjgd the dozen whereof
54 |           sd is
55 |  term: BOWL
56 |  title: Edition 1,1771
```

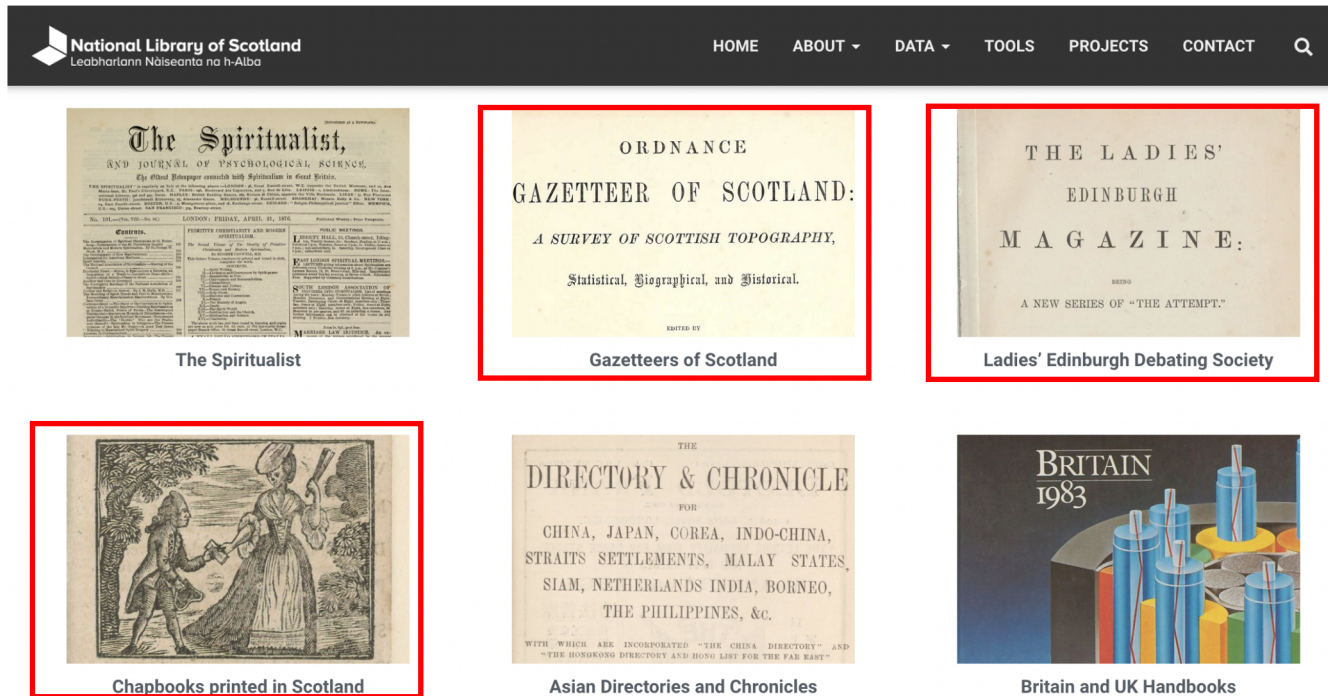
# Summary: Applying advanced IE/AI/NLP to historical digital collections

## NSL Data Foundry Collections



# Ongoing work: Working with other NLS digital Collections

- New **defoe** queries to extract NLS digital collections pages and metadata
- New **NLS ontology**
- New **Knowledge Graphs (KGs)** per digital collection
- New **notebooks** to show how to **query (SPARQL)** KGs
- Adapting **frances** for working with other **KGs**
- Improving the **frances** user interface & REACT web- application.
- Dockerise **frances** & defoe-RPC



The screenshot displays the National Library of Scotland website. The header includes the library's name in English and Scottish Gaelic, along with navigation links: HOME, ABOUT, DATA, TOOLS, PROJECTS, CONTACT, and a search icon. Below the header, six digital collections are showcased, each with a thumbnail image and a caption. The collections are: 'The Spiritualist' (a newspaper page), 'Gazetteers of Scotland' (a title page for a survey of Scottish topography), 'Ladies' Edinburgh Debating Society' (a title page for a magazine), 'Chapbooks printed in Scotland' (a woodcut illustration of two figures), 'Asian Directories and Chronicles' (a title page for a directory and chronicle for various Asian regions), and 'Britain and UK Handbooks' (a cover for a handbook featuring a bar chart).

National Library of Scotland  
Leabharlann Nàiseanta na h-Alba

HOME ABOUT DATA TOOLS PROJECTS CONTACT

**The Spiritualist**

**Gazetteers of Scotland**

**Ladies' Edinburgh Debating Society**

**Chapbooks printed in Scotland**

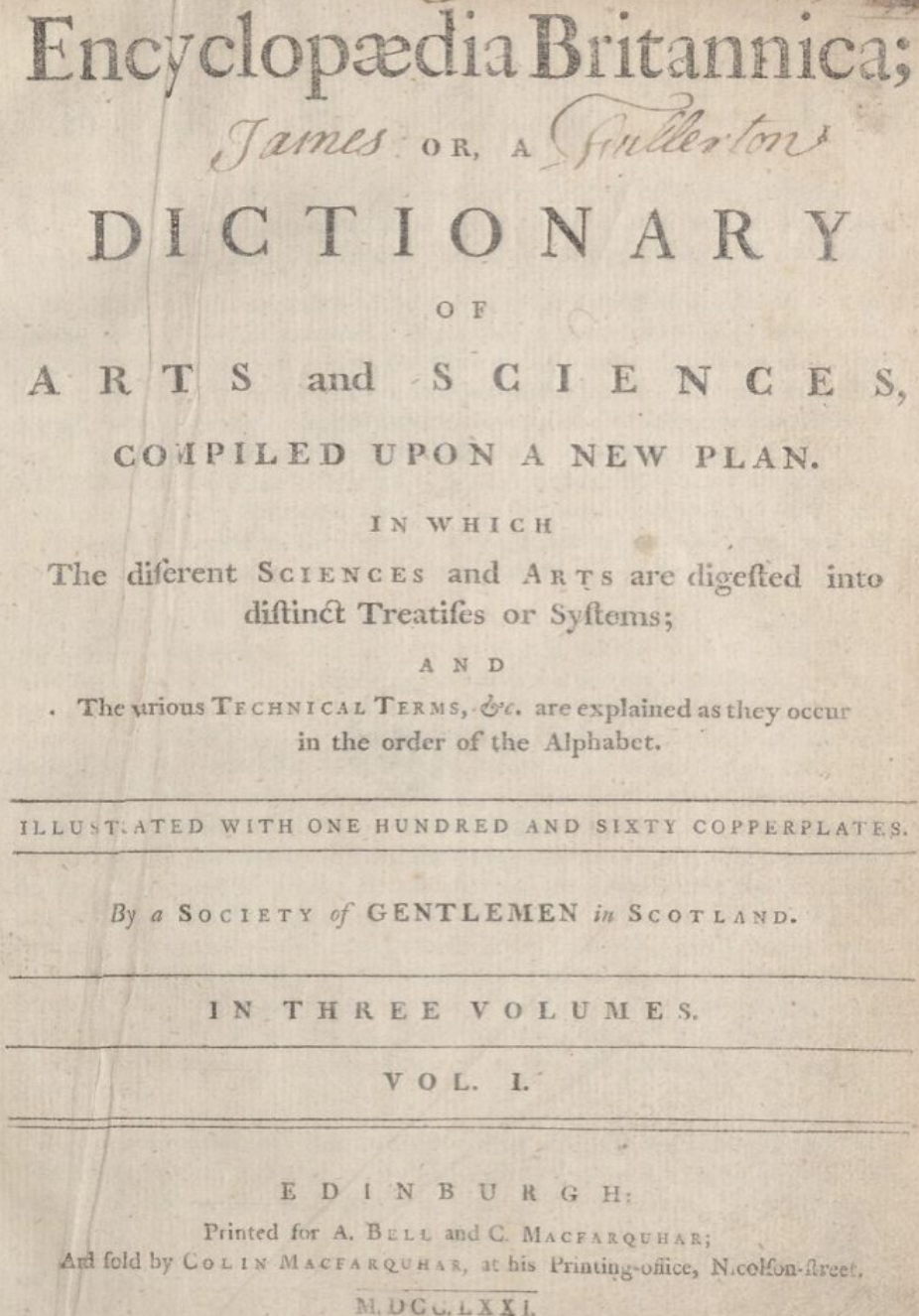
**Asian Directories and Chronicles**

**Britain and UK Handbooks**

# Questions?

Dr. Rosa Filgueira,  
Lecturer, School of Computer Science  
University of St Andrews,

Email: [rf208@st-andrews.ac.uk](mailto:rf208@st-andrews.ac.uk)



# Additional slides – New Frontend (REACT App)



frances

Term Search

Term Similarity

Topic Modelling

Defoe Queries

Collection Details

Create an account

Sign in

## Exploring the Encyclopaedia Britannica (1768-1860)

### Term Search

Enter the **term** that you would like to search for. In case that the **Term Type is a Topic**, only the **summary** of the definition is displayed. If not term is introduced, it will search for the first term in the Encyclopaedia.

1

Q

Results for **FLOWER**.

Note that if you click over an **URI** in this table, it will take you to the **Visualisation of Resources page** . However, if you instead click over a **related term** , it will conduct a **term search**, showing all the searching results for that term. And if you click over a **topic model** if will take you to the **Topic Modelling page** , listing all the terms belonging to that particular topic model.

displaying **1 - 4** records in total **9**

<<

1

2

3

>>

Year	Edition	Volume	Start Page	End Page	Term Type	Definition/Summary	Related Terms	Topic Modelling ID	Sentiment Score	Advanced Options
1771	1	2	566	566	Article	among botanifts and gardeners, the most beautiful part of trees and plants, containing the organs or.. <a href="#">More</a>		3190	POSITIVE_1.00	<div><div>Visualise</div><div>Check Spell</div><div>Similar terms</div></div>
						among botanifts and gardeners, the most beautiful				<div><div>Visualise</div></div>



# Additional slides – User Accounts (REACT App)

frances

## AI tool for Historical digital Collection

sign in with you france account

Sign in

[Forget password?](#)

[Create an account](#)

frances

## AI tool for Historical digital Collection

create a france account

Create an account



## Defoe Queries

Select Collection to query

Encyclopaedia Britannica

Configuration

Query type ?

publication\_nomalrized

Submit Query

### Previous Queries

[Publication normalized](#)

Frequency keysearch by year

[EB-animal-normalize-1710-1810](#)

[Chapbook-animal-normalize-1710-1810](#)

Uris keysearch

[EB-animal-normalize-1710-1810](#)

[Chapbook-animal-normalize-1710-1810](#)

[Chapbook-animal-normalize-1710-1810](#)

[Chapbook-animal-normalize-1710-1810](#)

[Chapbook-animal-normalize-1710-1810](#)

[More ...](#)

[All previous queries](#)



St Andrews / CS

Query type ?

frequency\_keysearch\_by\_year

#### Preprocess Treatment

It does not apply any type of treatment to the text.

None

#### Hit Count

It counts the number of times a keyword / key sentence appears in a term (article or topic)

Word ☒ Term

#### Upload Lexicon File

The file should contain a line per keyword and/or key sentence that you want to use in your query.

[Choose File](#)

#### Filtering Options

Target words/sentences

[Add](#)

Select term which contains **All** the target words/sentences

Or contains any of the target words/sentences.

☒

Year Range

To

Submit Query

# Additional slides – New Architecture

