

Encyclopædia Britannica;

James OR, A Fullerton

DICTIONARY

O F

A R T S and S C I E N C E S,

COMPILED UPON A NEW PLAN.

IN WHICH

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

A N D

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 SCOTLA

IN THREE VOLUMES

VOL. I.

EDINBURGH

Printed for A. BELL and C. M

And sold by COLIN MACFARQUHAR, at

M.DCC

*frances: An AI-toolbox to
discover automatically insights
from Data Foundry collections*

Dr. Rosa Filgueira,
Assistant Professor,
Heriot-Watt University,
Email: R.Filgueira@hw.ac.uk

*frances**: New AI-toolbox to discover automatically insights

New ways to unlock the full value of NLS digital collections

- Objective 1: New facilities to run more complex text analysis queries → ML/NLP techniques.
- Objective 2: Full integration with NLS Data Foundry → Hide the large-scale text mining complexity
- Using the **Encyclopaedia Britannica** as the core dataset

(*) *Frances Wright* (September 6, 1795 – December 13, 1852)

frances: Providing Automatic ML Analysis

frances will provide **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

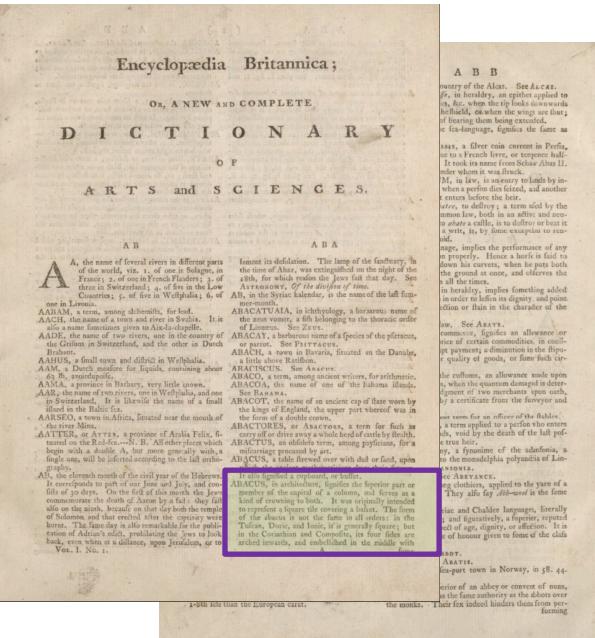
My Journey

- (1) From Semi-Structured EB information (ALTO & METS XML files) →**
Information Extraction → Knowledge Graph → Deep learning
Transformers → **To Augmented EB-Knowledge Graph with advance AI-methods**

- (2) Querying Augment EB Knowledge Graph:**
 - (1) Extracting information already stored in the EB-Knowledge Graph**
 - (2) Processing information stored in the EB-Knowledge Graph **in parallel****
→ Create new data/results

Edition 1, 1771

TERM ABACUS

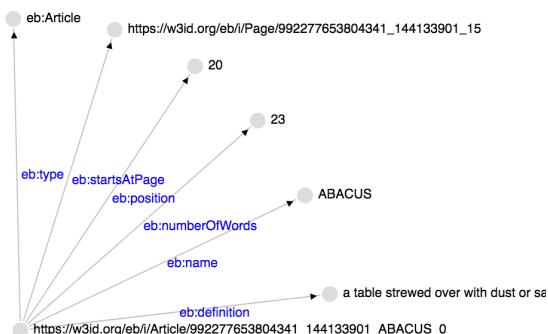
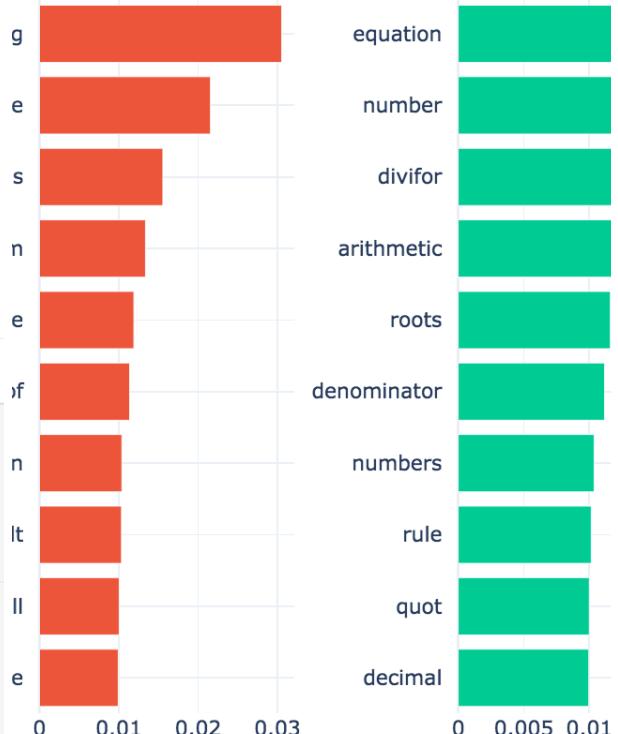


Cleaned Definition [Compute Diff](#) [Reset](#)

in architecture, signifies the superior part or member of the capital of a column, and serves as a kind of crown covering a basket. The form of the abacus is not the same in all orders: in the Tuscan, Doric, and Ionic, it has four sides are arched in Arches, and embellished in the middle with ornament, as a rose or other flower, Scar Tuscan pedestal; and Palladio calls the plinth above the echinus, or boutin, in the Tufean and Doric orders.,

Year	Edition	Volume	Start Page	End Page	Term Type	Definition/Summary	Related Terms
1771	1	1	15	15	Article	a table strewed over with dust or sand, upon which the ancient mathematicians drew their figures, It... More	
1771	1	1	15	16	Article	in architecture, signifies the superior part or member of the capital of a column, and serves as a k... More	
1	1	1	16	16	Article	is also the name of an ancient instrument for facilitating operations in arithmetic. It is vadouly ... More	
1	1	1	16	16	Article	logijlicus, a right-angled triangle, whose sides forming the right angle contain the numbers from 1 ... More	

Topic 11



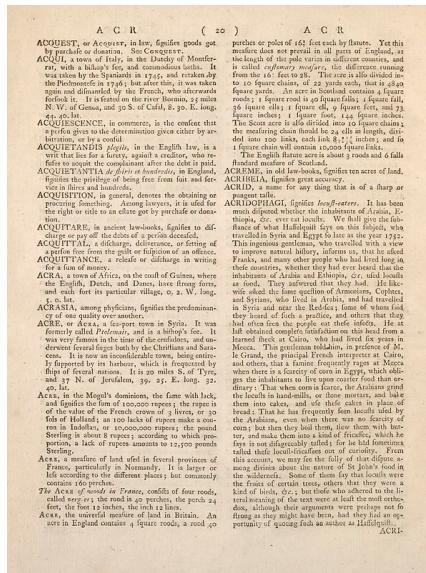
My Journey

Applying advanced AI/NLP methods to historical digital collections

Phase 1: Information Extraction

- Detect, Classify and Extract all the terms of the EB – different Editions
 - 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



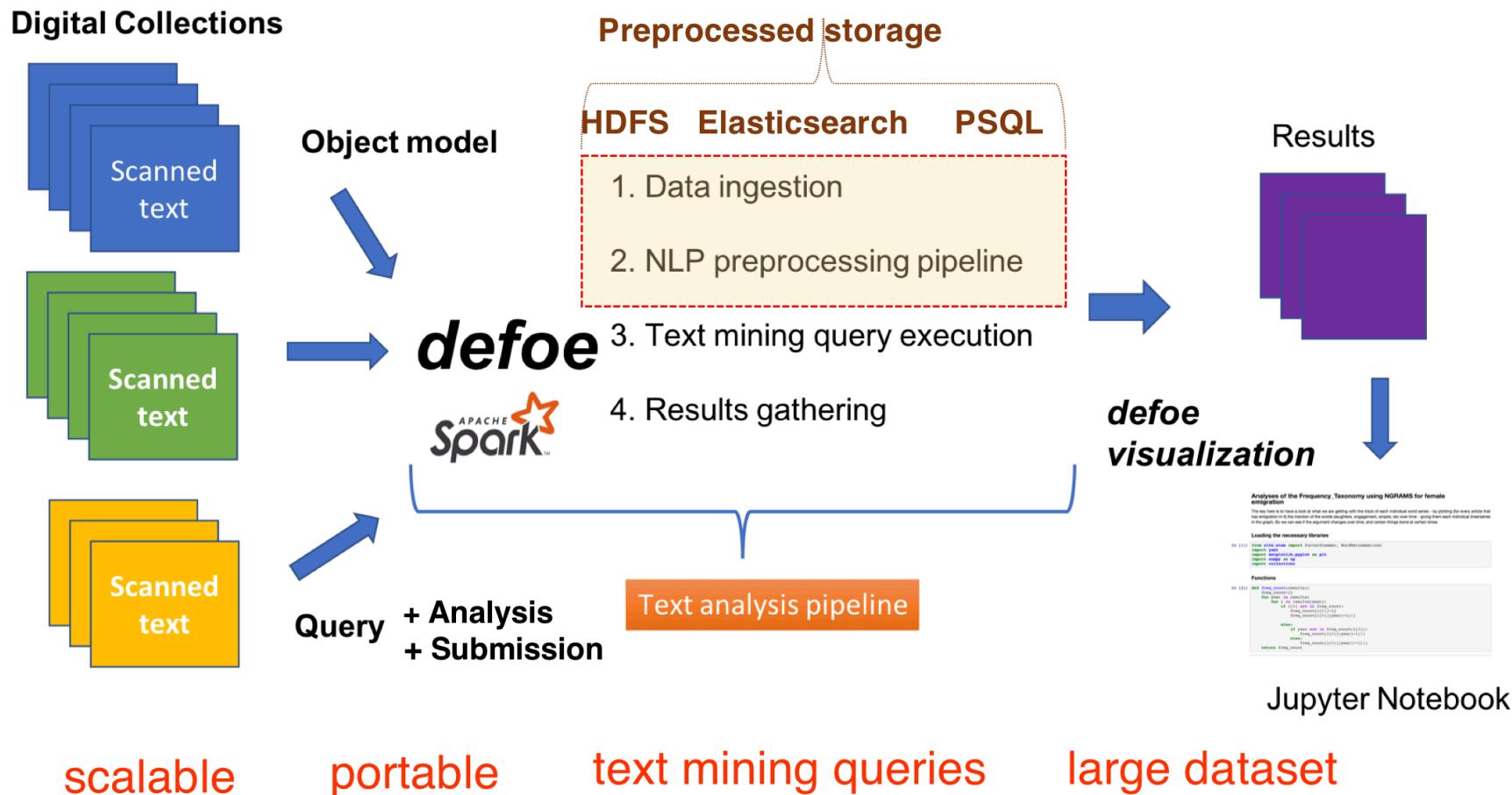
Page 36 of the First Edition
of the Encyclopaedia Britannica

```
xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet href="http://www.loc.gov/standards/alto/v3/alto.xsd">
  <!--Description-->
  <!--MeasurementUnit pixel</MeasurementUnit>
  <!--SourceImageInformation-->
  <!--FileNames-->/data/pdfs/C_188936519/L_144133901/18882693.23.pdf</fileName>
  <!--SourceImagePageNumber-->
  <!--OCRProcessing ID="100cr"-->
  <!--ocrProcessingStep-->
  <!--processingDateTime Fri Jul 12 10:36:14 2019-->
</processingDateTime>
  <!--processingSoftware-->
    <!--softwareCreator>CONTRIBUTORS</softwareCreator>
    <!--softwareName>pdflatoo</softwareName>
    <!--softwareVersion>0.1</softwareVersion>
  </!--processingSoftware-->
  <!--ocrProcessingStep-->
  <!--processingStep-->
</OCRProcessing>
  <!--Description-->
<!--TextStyle-->
  <!--TextStyle ID="font18" FONTNAME="times" FONTSIZE="10.000" FONTTYPE="proportional" FONTWIDTH="proportional" FONTCOLOR="#000000" FONTSTYLE="normal"-->
  <!--TextStyle ID="font11" FONTNAME="times" FONTSIZE="6.580" FONTTYPE="serif" FONTWIDTH="proportional" FONTCOLOR="#000000" FONTSTYLE="italic"-->
<!--TextStyle-->
<!--Layout-->
  <!--Page ID="Page36" PHYSICAL_IMG_NR="36" WIDTH="2466" HEIGHT="3327"-->
    <!--TextLine-->
      <!--TextLine ID="p36_1" WIDTH="232" HEIGHT="75" ID="p36_11" HPOS="512" VPOS="180"-->
        <!--String ID="p36_1d" CONTENT="<P><FONT>100</FONT><B>100</B></P>"-->
        <!--String ID="p36_1s" CONTENT="100" ID="p36_11s" HPOS="512" VPOS="180" WIDTH="42" HEIGHT="75" STYLEREFS="font8"/>
        <!--SP ID="p36_1p" VPOS="180" HPOS="554"/>
        <!--String ID="p36_2" CONTENT="<P><FONT>C</FONT><B>C</B></P>"-->
        <!--String ID="p36_2d" CONTENT="C" ID="p36_2s" HPOS="611" VPOS="188" WIDTH="37" HEIGHT="75" STYLEREFS="font8"/>
        <!--String ID="p36_2p" VPOS="188" HPOS="648"/>
        <!--String ID="p36_3" CONTENT="<P><FONT>HPSO</FONT><B>HPSO</B></P>"-->
        <!--String ID="p36_3d" CONTENT="HPSO" ID="p36_3s" HPOS="705" VPOS="188" WIDTH="38" HEIGHT="75" STYLEREFS="font8"/>
      </TextLine>
      <!--TextLine ID="p36_4" WIDTH="132" HEIGHT="75" ID="p36_42" HPOS="119" VPOS="188"-->
        <!--String ID="p36_4d" CONTENT="<P><FONT>119</FONT><B>119</B></P>"-->
        <!--String ID="p36_4s" HPOS="119" VPOS="188" WIDTH="19" HEIGHT="75" STYLEREFS="font8"/>
        <!--SP ID="p36_4p" VPOS="188" HPOS="138"/>
        <!--String ID="p36_5" CONTENT="<P><FONT>20</FONT><B>20</B></P>"-->
        <!--String ID="p36_5d" CONTENT="20" ID="p36_5s" HPOS="188" VPOS="188" WIDTH="51" HEIGHT="75" STYLEREFS="font8"/>
      </TextLine>
      <!--TextLine ID="p36_6" WIDTH="109" HEIGHT="75" ID="p36_13" HPOS="171" VPOS="182"-->
        <!--String ID="p36_6d" CONTENT="<P><FONT>171</FONT><B>171</B></P>"-->
        <!--SP ID="p36_6p" VPOS="182" HPOS="414"/>
        <!--String ID="p36_7" CONTENT="<P><FONT>0</FONT><B>0</B></P>"-->
        <!--String ID="p36_7d" CONTENT="0" ID="p36_7s" HPOS="492" VPOS="182" WIDTH="39" HEIGHT="75" STYLEREFS="font8"/>
        <!--SP ID="p36_7p" VPOS="182" HPOS="477"/>
        <!--String ID="p36_8" CONTENT="<P><FONT>192</FONT><B>192</B></P>"-->
        <!--String ID="p36_8d" CONTENT="192" ID="p36_8s" HPOS="685" VPOS="182" WIDTH="193" HEIGHT="75" STYLEREFS="font8"/>
        <!--String ID="p36_9" CONTENT="<P><FONT>182</FONT><B>182</B></P>"-->
        <!--String ID="p36_9d" CONTENT="182" ID="p36_9s" HPOS="837" VPOS="189" WIDTH="182" HEIGHT="75" STYLEREFS="font8"/>
        <!--String ID="p36_10" CONTENT="<P><FONT>182</FONT><B>182</B></P>"-->
        <!--String ID="p36_10d" CONTENT="182" ID="p36_10s" HPOS="857" VPOS="182" WIDTH="134" HEIGHT="75" STYLEREFS="font8"/>
        <!--SP ID="p36_10p" VPOS="182" HPOS="991"/>
        <!--String ID="p36_11" CONTENT="<P><FONT>1015</FONT><B>1015</B></P>"-->
        <!--String ID="p36_11d" CONTENT="1015" ID="p36_11s" HPOS="1015" VPOS="182" WIDTH="182" HEIGHT="75" STYLEREFS="font8"/>
        <!--SP ID="p36_11p" VPOS="182" HPOS="1117"/>
        <!--String ID="p36_12" CONTENT="<P><FONT>1141</FONT><B>1141</B></P>"-->
        <!--String ID="p36_12d" CONTENT="1141" ID="p36_12s" HPOS="182" VPOS="182" WIDTH="59" HEIGHT="75" STYLEREFS="font8"/>
      </TextLine>
      <!--TextLine ID="p36_13" WIDTH="792" HEIGHT="75" ID="p36_40" HPOS="221" VPOS="223"-->
    </TextLine>
```

ALTO-XML of the Page 36 of the First Edition of the Encyclopaedia Britannica

Phase 1: Information Extraction

defoe: scalable toolbox for historical research



<https://github.com/defoe-code/defoe>

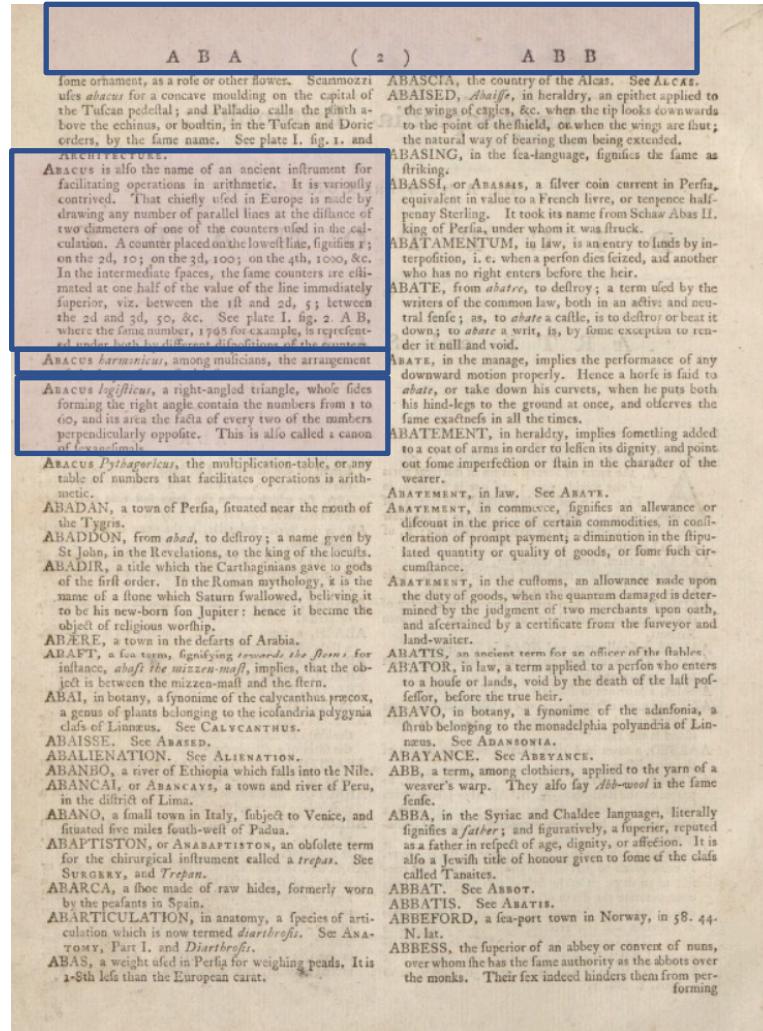
Phase 1: Information Extraction

1.1. Improved **defoe Extract Terms query** → It extracts the Eb Terms by page and classify them between **articles** and **topics (Terms v.1)** → Using ALTO XML

Extracted Terms Based on Heuristics → Pages Layout & Text & Headers →
Different heuristics for different EB editions.

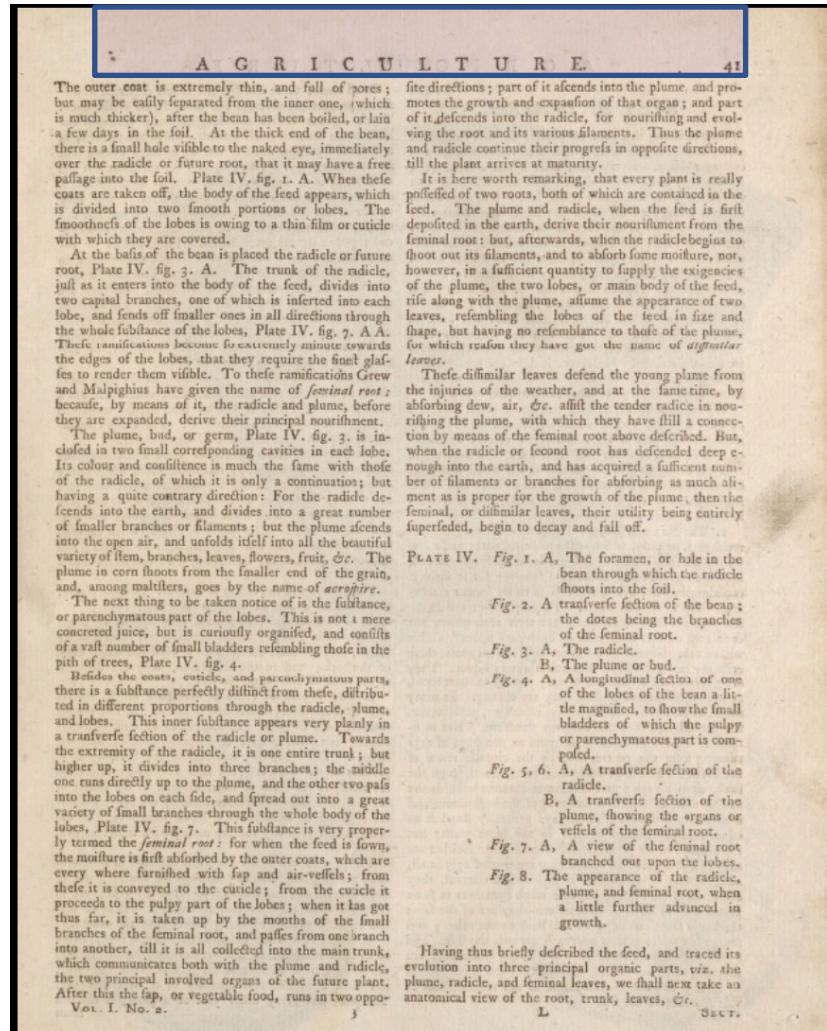
Phase 1: Information Extraction

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

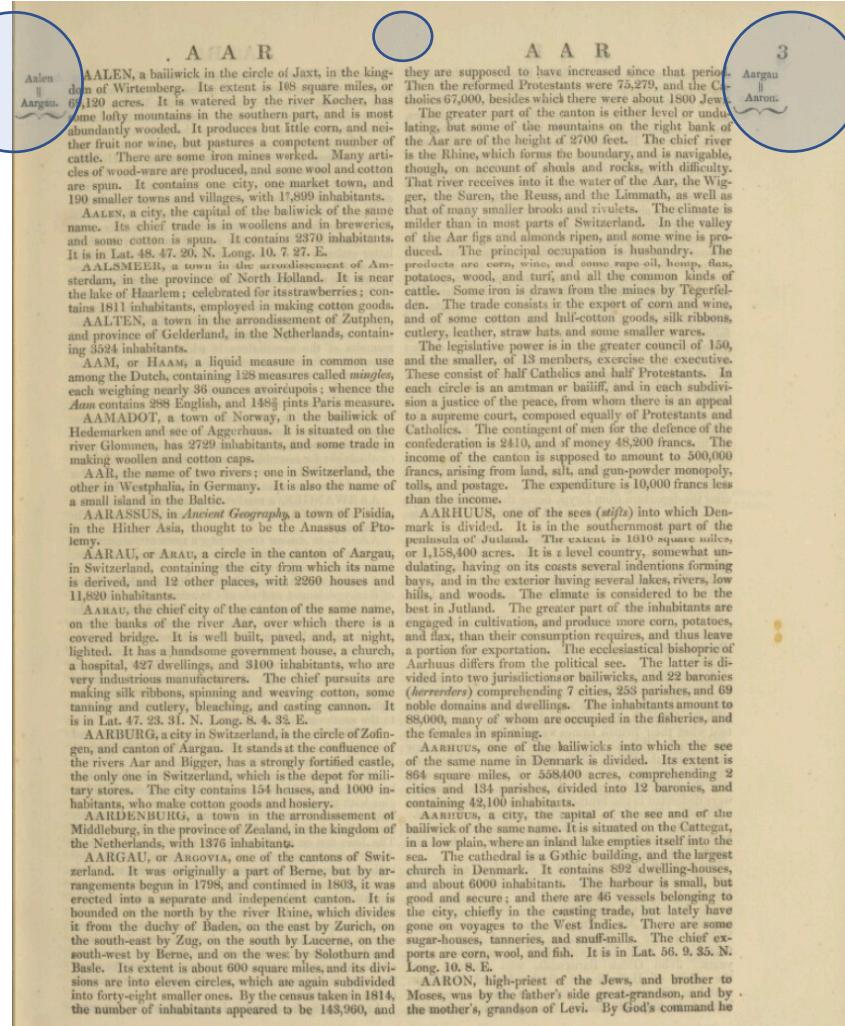
Edition 1 - 1771



Topic

Phase 1: Information Extraction

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 + “,”



they are supposed to have increased since that period. Then the reformed Protestants were 75,279, and the Catholics 67,000, besides which there were about 1800 Jews.

The greater part of the canton is either level or undulating, but some of the mountains on the right bank of the Aar are of the height of 2700 feet. The chief river is the Rhine, which forms the boundary, and is navigable, though, on account of shoals and rocks, with difficulty. That river receives into it the water of the Aar, the Wiggen, the Suren, the Reuss, and the Limmat, as well as that of many smaller brooks and rivulets. The climate is milder than in most parts of Switzerland. In the valley of the Aar figs and almonds ripen, and some wine is produced. The principal occupation is husbandry. The products are corn, wine, and some rape-oil, hemp, flax, potatoes, wood, and turf, and all the common kinds of cattle. Some iron is drawn from the mines by Tegerfelden. The trade consists in the export of corn and wine, and of some cotton and half-cotton goods, silk ribbons, cutlery, leather, straw hats, and some smaller wares.

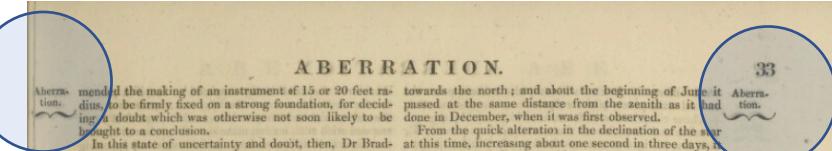
The legislative power is in the greater council of 150, and the smaller, of 13 members, exercise the executive. These consist of half Catholics and half Protestants. In each circle is an amman or bailiff, and in each subdivision a justice of the peace, from whom there is an appeal to a supreme court, composed equally of Protestants and Catholics. The contingent of men for the defence of the confederation is 2410, and of money 48,200 francs. The income of the canton is supposed to amount to 500,000 francs, arising from land, salt, and gun-powder monopoly, tolls, and postage. The expenditure is 10,000 francs less than the income.

AARHUS, one of the sees (*stifts*) into which Denmark is divided. It is in the southemmost part of the peninsula of Jutland. The latter is 1010 miles long, and 1,158,400 acres. It is a level country, somewhat undulating, having on its coasts several indentations forming bays, and in the exterior having several lakes, rivers, low hills and woods. The climate is considered to be the best in Jutland. The greater part of the inhabitants are engaged in cultivation, and produce more corn, potatoes, and flax, than their consumption requires, and thus leave a portion for exportation. The ecclesiastical bishopric of Aarhus differs from the political see. The latter is divided into two jurisdictions or bailiwicks, and 22 baronies (*herrederier*) comprehending 7 cities, 253 parishes, and 69 noble domains and dwellings. The inhabitants amount to 88,000, many of whom are occupied in the fisheries, and the females in spinning.

AARHUS, one of the bailiwicks into which the see of the same name in Denmark is divided. Its extent is 64 square miles, or 558,400 acres, comprehending 2 cities and 134 parishes, divided into 12 baronies, and containing 42,100 inhabitants.

AARHUS, a city, the capital of the see and of the bailiwick of the same name. It is situated on the Cattgat, in a low plain, where an inland lake empties itself into the sea. The cathedral is a Gothic building, and the largest church in Denmark. It contains 892 dwelling-houses, and about 6000 inhabitants. The harbour is small, but good and secure; and there are 46 vessels belonging to the city, chiefly in the coasting trade, but lately have gone on voyages to the West Indies. There are some sugar-houses, tanneries, and snuff-mills. The chief exports are corn, wool, and fish. It is in Lat. 56° 9. 35. N. Long. 10. 8. E.

AARON, high-priest of the Jews, and brother to Moses, was by the father's side great-grandson, and by the mother's, grandson of Levi. By God's command he



3

Aberration. intended the making of an instrument of 15 or 20 feet radius to be firmly fixed on a strong foundation, for deciding a doubt which otherwise not soon likely to be removed by a conclusion.

In this state of uncertainty and doubt, then, Dr Bradley, in conjunction with Mr Samuel Molineux, in the year 1735, formed the project of verifying, by a series of new observations, those which Dr Hook had communicated to the publick almost 50 years before. And as it was his attempt that chiefly gave rise to this, so it was his method in making the observations, in some measure, that they followed; for they made choice of the same star, and their instrument was constructed upon nearly the same principles; but had it not greatly exceeded the former in exactness, they might still have continued in great uncertainty as to the parallax of the fixed stars. For this, and many other convenient and useful astronomical instruments, publick observations are indebted to the ingenuity and accuracy of Mr Graham.

The success of the experiment evidently depending so much on the accuracy of the instrument, this became a leading object of consideration. Mr Molineux's apparatus then having been completed, and fitted for observing about the end of November 1735, on the third day of December following, the bright star in the head of Draco, marked γ by Bayer, was for the first time observed, as it passed near the zenith, and its situation carefully taken with the instrument. Like observations were made on the fifth, eleventh, and twelfth days of the same month; and there appearing no material difference in the place of the star, a further repetition of them, at season, seemed needless, it being a time of the year in which no sensible alteration of parallax, in this star, could soon be expected. It was therefore conjectured that chiefly upon Dr Bradley's plan, then known, the instrument was fixed, to prepare for observing the star again on the 17th of the same month; when, having adjusted the instrument as usual, he perceived that it passed a little more southerly this day than it had done before. Not suspecting any other cause of this appearance, it was ascribed to the uncertainty of the observations, and that either this or the foregoing was not so exact as had been supposed. For which reason they proposed to repeat the observation again, to determine from what cause this difference might proceed; and upon doing it, on the 20th of December, the doctor found that the star passed still more southerly than at the preceding observation. This sensible alteration surprised them the more, as it was the contrary way from what it would have been had it proceeded from an annual parallax of the star. But being now pretty well satisfied that it could not be entirely owing to the want of accuracy in the observations, and having no notion of any thing else that the star could do, such an apparent motion as this in the sky, they began to suspect that some change in the materials or fabric of the instrument itself might have occasioned it. Under these uncertainties they remained for some time; but being at length fully convinced, by several trials, of the great exactness of the instrument, and finding, by the gradual increase of the star's distance from the pole, that there must be some regular cause that produced it, they took care to examine very nicely, at the time of each observation, how much the variation was; till about the beginning of March 1736, the star was found to be 20° more southerly than at the time of the first observation: it now indeed seemed to have arrived at its utmost limit southward, in which time, made about this time, no sensible difference was observed in its situation. By the middle of April it appeared to be returning back again

towards the north; and about the beginning of June it had again returned from the south, as it had done in December, when it was first observed.

From the quick alteration in the declination of the star at this time, increasing about one second in three days, it

was conjectured that it would now proceed northward, as it had before gone southward, of its present situation; and it happened accordingly; for the star continued to move northward till September following, when it again became stationary; being then near 20° more northerly than it had been in March. From September the star again returned towards the south, till, in December, it arrived at the same situation in which it had been observed twelve months before, allowing for the difference of declination on account of the precession of the equinox.

This was a sufficient proof that the instrument had not been the cause of the apparent motion of the star; and yet it seemed difficult to devise one that should be adequate to such an unusual effect. A nutation of the earth's axis was one of the first things that offered itself on this occasion; but it was soon found to be insufficient; for though it might have accounted for the change of declination in Draco, yet it would not at the same time accord with the phenomena observed in the other stars, particularly in a small one almost opposite in right ascension to γ Draconis, and at about the same distance from the north pole of the equator; for though this star seemed to move the same way as a nutation of the earth's axis would have made it, yet changing its declination but about half as much as γ Draconis in the same time, as appeared on comparing the observations of both made on the same days, at different seasons of the year, this plainly proved that the apparent motion of the star was not occasioned by a real nutation; for had this been the case, the alteration in both stars would have been nearly equal.

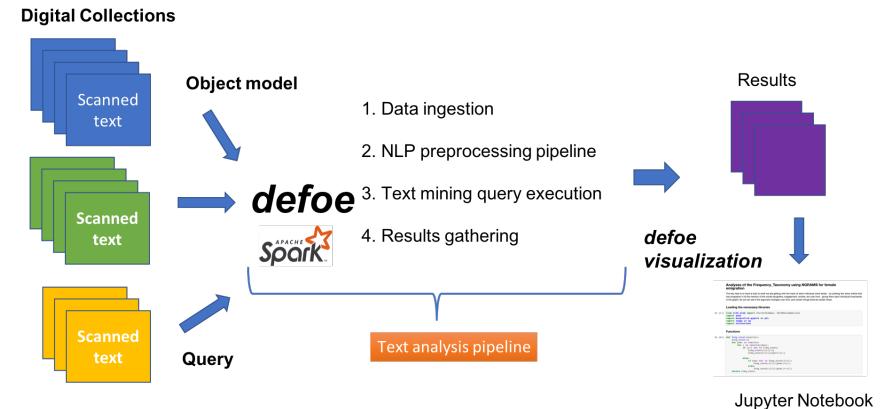
The great regularity of the observations left no room to doubt, but that there was some uniform cause by which this unexpected motion was produced, and which did not depend on the uncertainty or variety of the seasons of the year. Upon comparing the observations with each other, it was discovered that, in both the stars above mentioned, the apparent difference of declination from the *maxima* was always nearly proportional to the versed sine of the sun's distance from the equinoctial points. This was an inducement to think that the cause, whatever it was, had some relation to the sun's situation with respect to those points. But not being able to frame any hypothesis sufficient to account for all the phenomena, and being very desirous to search a little further into this matter, Dr Bradley began to think of erecting an instrument for himself, and of employing it in the same manner as himself, he might with the more ease, and certainty inquire into the laws of this new motion. The consideration likewise of being able, by another instrument, to conform the truth of the observations hitherto made with that of Mr Molineux, was no small inducement to the undertaking; but the chief of all was, the opportunity he should thereby have of trying in what manner other stars should be affected by the same cause, whatever it might be. For Mr Molineux's instrument being originally designed for observing γ Draconis, to try whether it had any sensible parallax, it was so contrived as to be capable of but little alteration in its direction; not above seven or eight minutes of a degree: and there being but few stars within half that distance from the zenith of Kew bright enough to be well observed, he could not, with his instrument, thoroughly examine how this cause affected stars that

VOL. II.

33

Aberration.

Extract EB Terms



```
term
definition
relatedTerms
header
startsAt
endsAt
numberOfTerms
numberOfWords
numberOfPages
positionPage
typeTerm
editionTitle
editionNum
supplementTitle
supplementsTo
year
place
volumeTitle
volumeNum
letters
part
altoXML
Name: 18, dtype: object
```

ABACTORES
or ABACTORS, a term for such as carry offer dr...
[]
EBAA
15
15
22
18
832
18
Article
First edition, 1771, Volume 1, A-B
1
[]
1771
Edinburgh
Encyclopaedia Britannica; or, A dictionary of ...
1
A-B
0
144133901/alto/188082904.34.xml

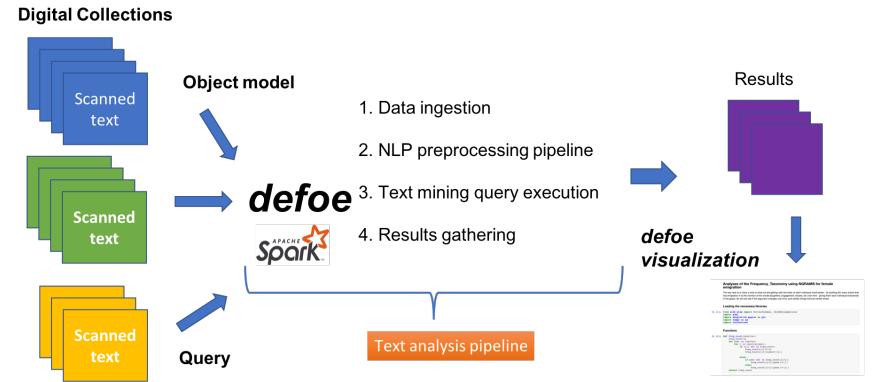
Extracted Term ABACATORES information - Edition 1, 1771, Vol A-B

Phase 1: Information Extraction

1.2 Improved defoe **Metadata Extraction query** → It extracts the metadata per Edition and Volume (**Metadata v.1**) → **Based on METS information**

Note: Supplements have a slightly different metadata information.

Extract collection Metadata



	MMSID	editionTitle	editor	editor_date	genre	language	termsOfAddress	numberOfPages	physicalDescription	place	...	permanentURL
14	997902543804341	Third edition, Volume 2, ANG-BAR	None	None	encyclopedia	eng	None	922	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149977338
15	997902543804341	Third edition, Volume 3, BAR-BZO	None	None	encyclopedia	eng	None	856	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149977873
16	997902543804341	Third edition, Volume 4, CAA-CIC	None	None	encyclopedia	eng	None	842	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149978642
17	997902543804341	Third edition, Volume 5, CIC-DIA	None	None	encyclopedia	eng	None	858	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149979156
18	997902543804341	Third edition, Volume 6, DIA-ETH	None	None	encyclopedia	eng	None	850	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149979622
19	997902543804341	Third edition, Volume 7, ETM-GOA	None	None	encyclopedia	eng	None	882	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149981189
20	997902543804341	Third edition, Volume 8, GOB-HYD	None	None	encyclopedia	eng	None	832	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149981670
21	997902543804341	Third edition, Volume 9, Hydrostatics- LES	None	None	encyclopedia	eng	None	872	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149982181
22	997902543804341	Third edition, Volume 10, LES-MEC	None	None	encyclopedia	eng	None	842	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149982692
23	997902543804341	Third edition, Volume 11, Medals- Midwifery	None	None	encyclopedia	eng	None	862	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/149983206
26	997902543804341	Third edition, Volume 1, A-ANG	None	None	encyclopedia	eng	None	894	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/190273291
27	997902543804341	Third edition, Volume 12, MIE-NEG	None	None	encyclopedia	eng	None	870	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/190273372
28	997902543804341	Third edition, Volume 13, NEH-PAS	None	None	encyclopedia	eng	None	874	18v.,plates : ill.,maps,music ; 4to	Edinburgh	...	https://digital.nls.uk/191253798

Extracted Metadata of some of the Volumes of Edition 3

... more metadata columns

Phase 1: Information Extraction

1.3. **New Post-processing python scripts (*)** to improve the previous results:

- Re-classification of **Terms v.1** (articles and topics)
- Join terms split across pages

We get here: **Terms & Metadata v.2**

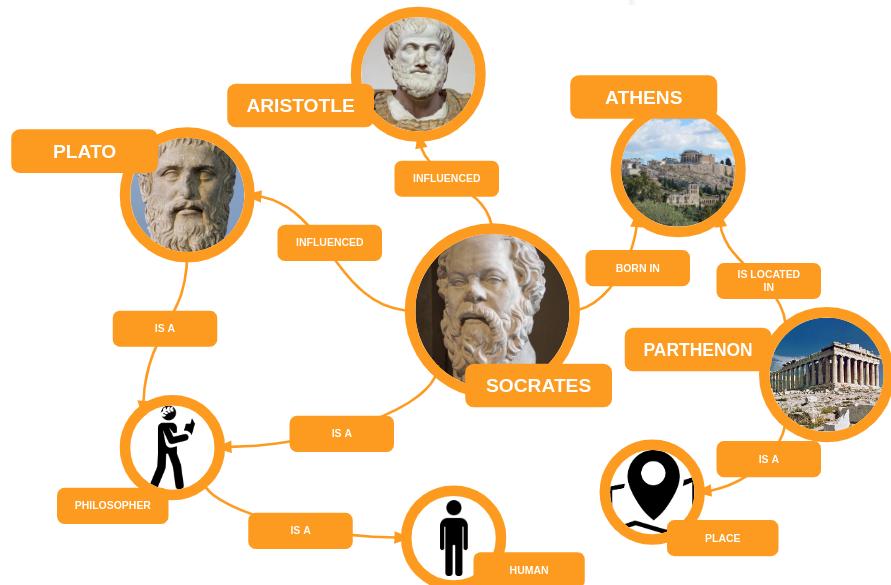
(*) **Also Based on Heuristics** --> Pages Layout & Text & Headers → different heuristics for different EB editions.

Phase 2: Knowledge Graph

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

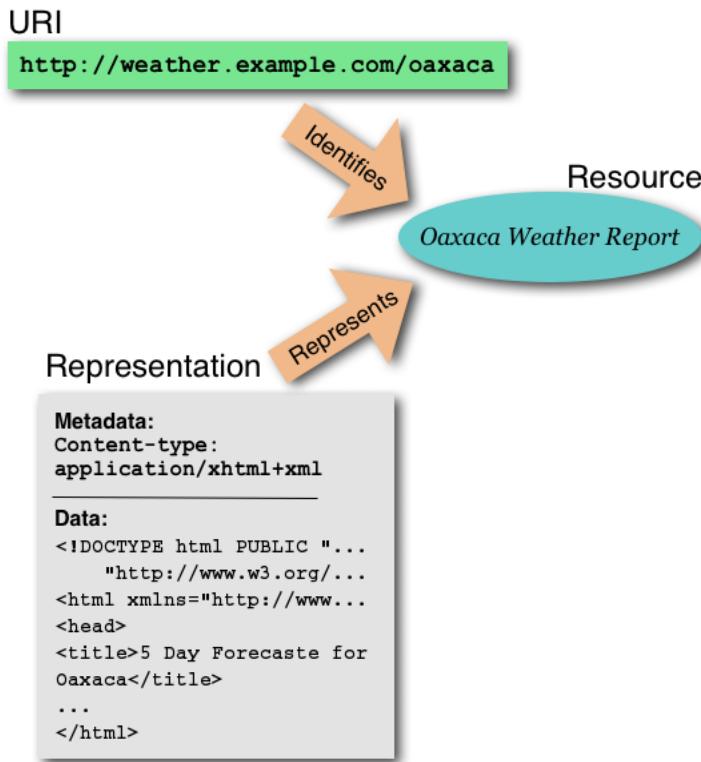
- A **knowledge graph** is a specialized graph or network of the things we want to describe and how they are related
- It is a **semantic** model since we want to capture and generate **meaning** with the model

"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*

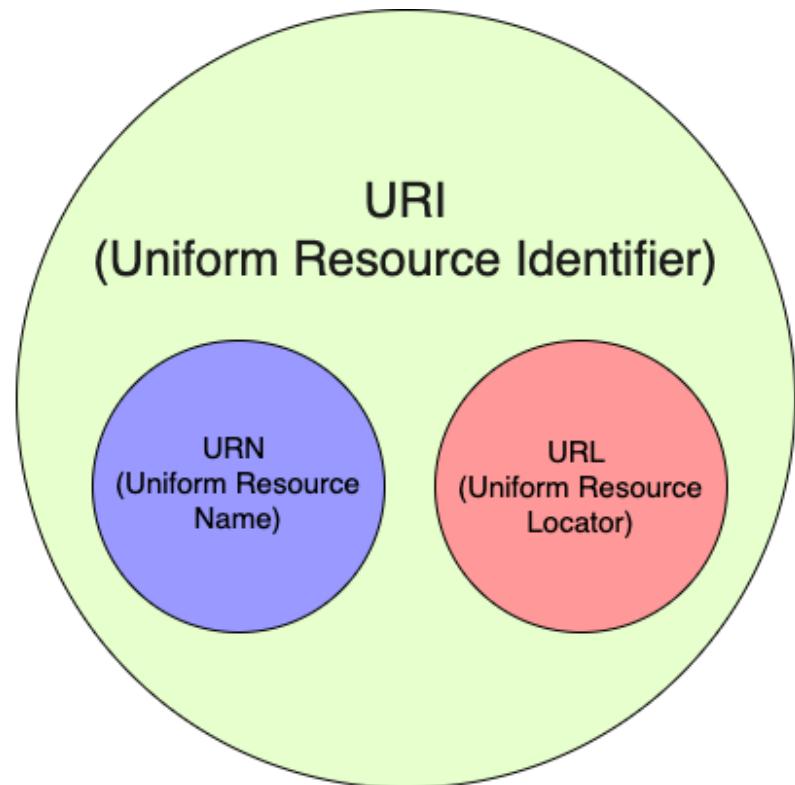


Phase 2: Knowledge Graph

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



- **URI: A Universal Resource Identifier**, is defined to be an ASCII string used to identify “things” on the Knowledge Graph



Phase 2: Knowledge Graph

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

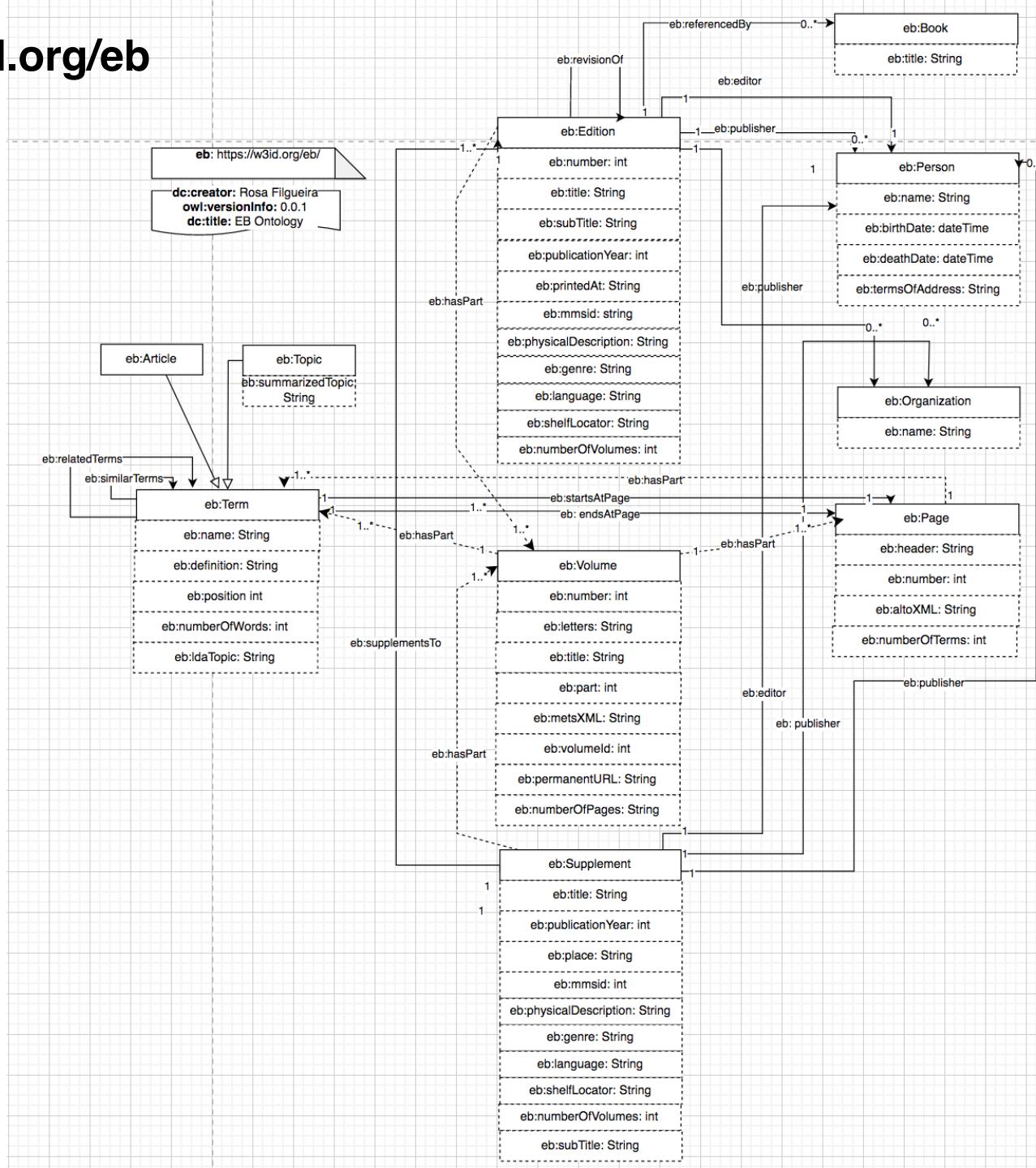
2.1. Create EB Ontology: To explain (give meaning) how our “things” are related with each other.

In order to create and publish the EB-Ontology I 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 : <https://github.com/francesNLP/EB-ontology>

<https://w3id.org/eb>



Phase 2: Knowledge Graph

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

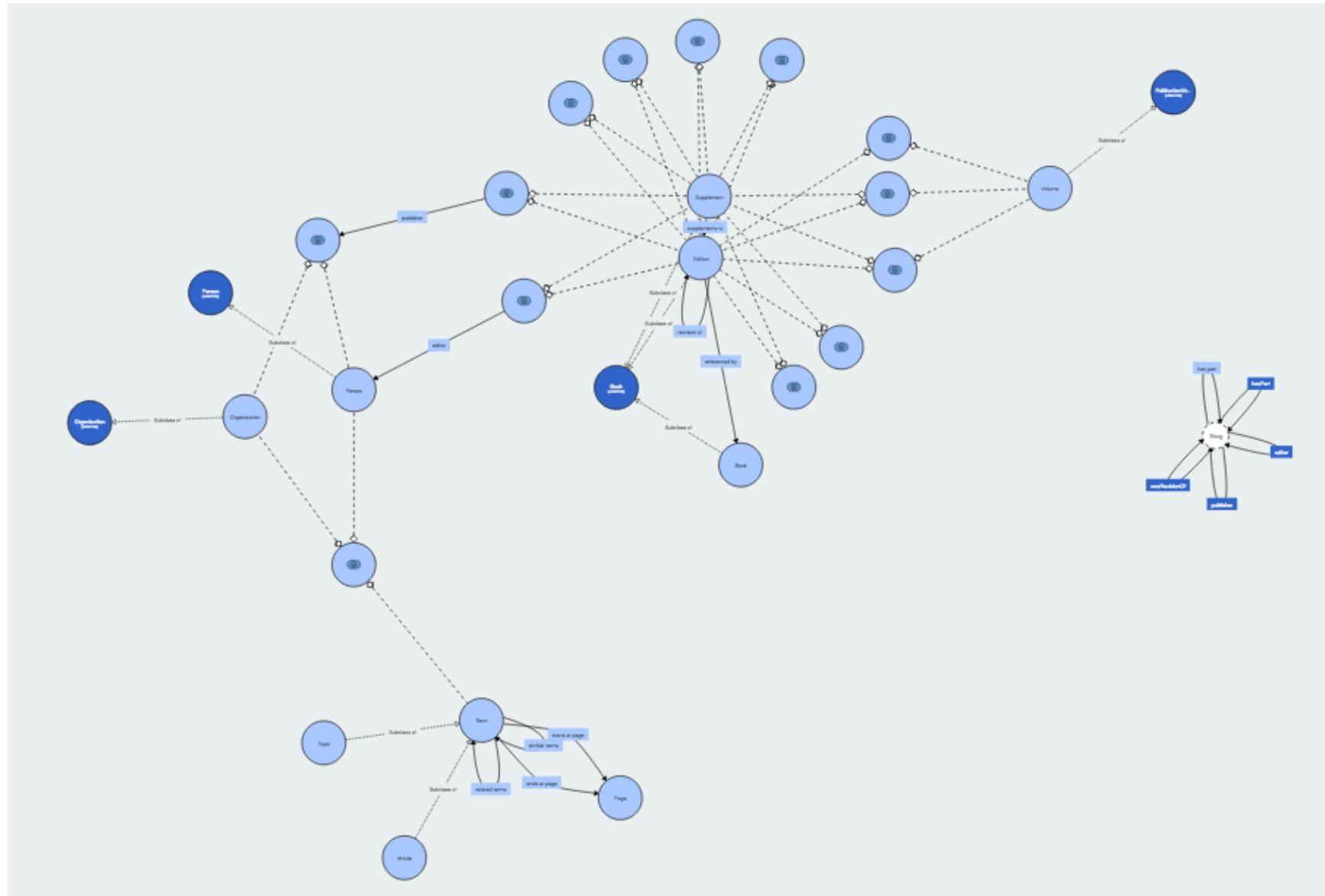
2. 2 Created **EB Knowledge Graph 1.0**: Populated the post-processed information (extracted **Terms & Metadata v.2**) into an RDF triplestore:

- Apache Jena FUSEKI SPARQL server
- Used this [Fuseki-docker](#) image to set up my SPARQL server/end-point

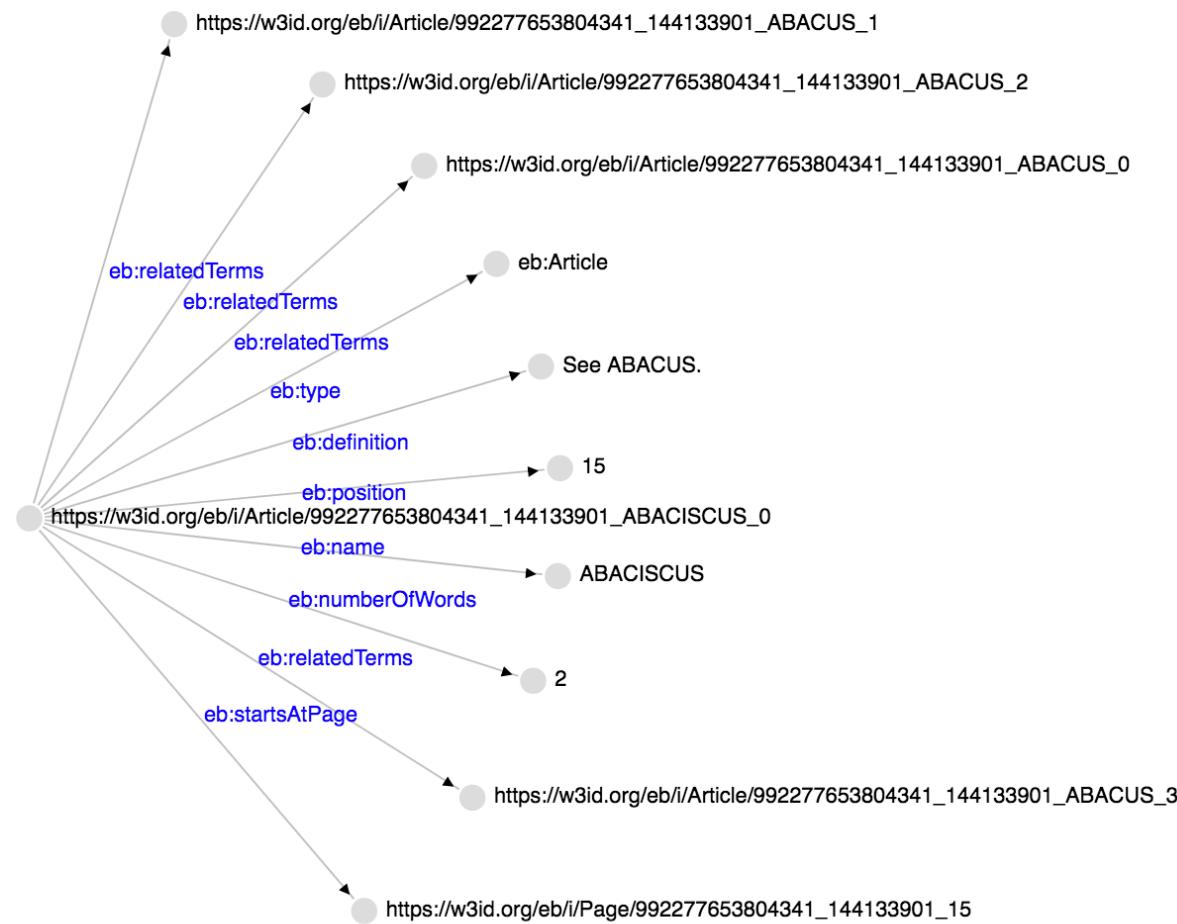
NOTE:

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

EB Knowledge Graph

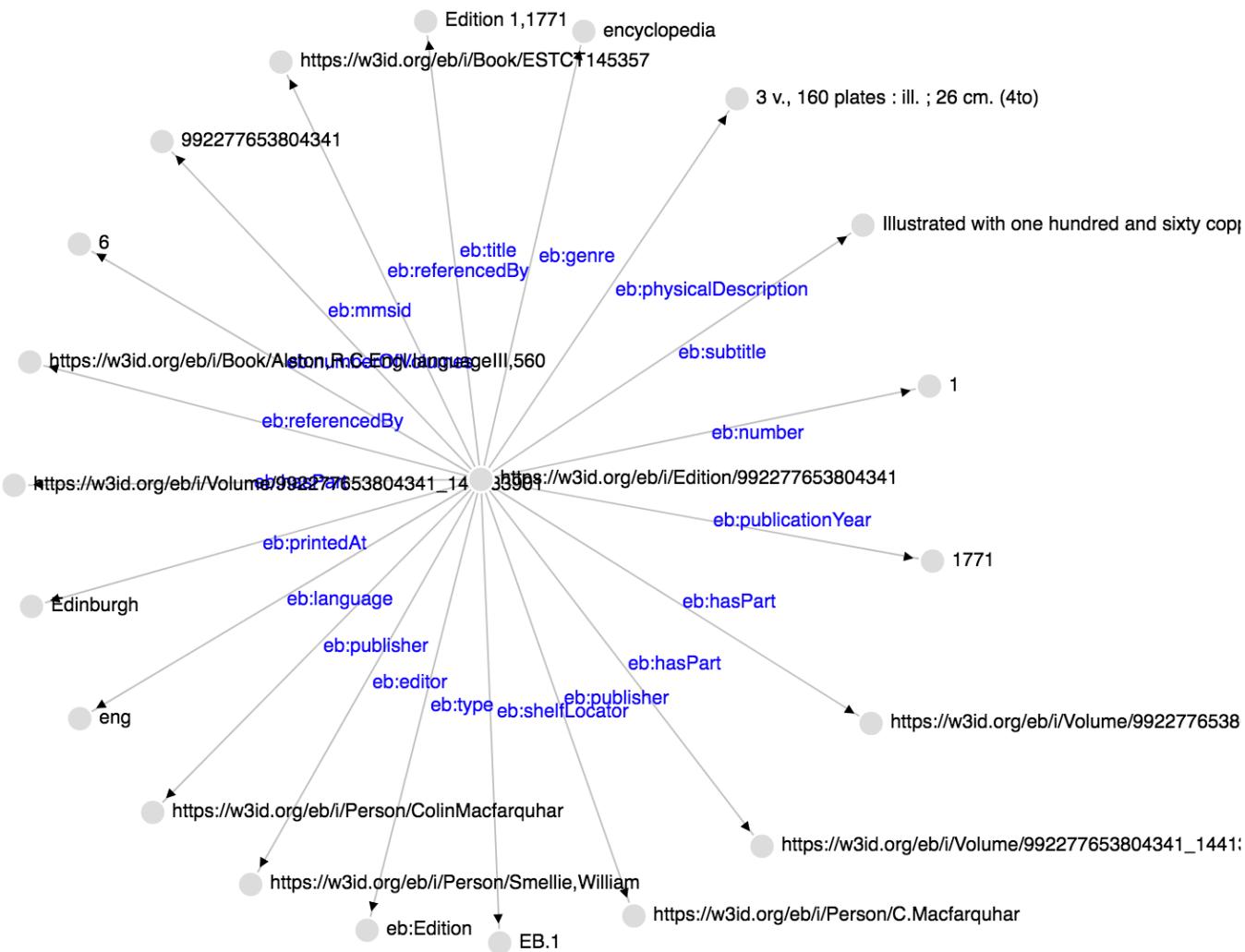


EB Knowledge Graph



Term ABACISCUS, URI: https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_0

EB Knowledge Graph



Edition 1, 1771 – URI: <https://w3id.org/eb/i/Edition/992277653804341>

EB Knowledge Graph – Querying our KG

SPARQL is an **RDF query language**—that is, a semantic query language for databases—able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

```
2 SELECT ?article ?definition
3 WHERE {
4     ?article a eb:Article .
5     ?article eb:definition ?definition
6     FILTER (CONTAINS(?definition, "Scotland"))
7     FILTER (CONTAINS(?definition, "Glasgow"))
8     OPTIONAL{FILTER CONTAINS(?definition, "Edinburgh") }
9 } LIMIT 10
10
```

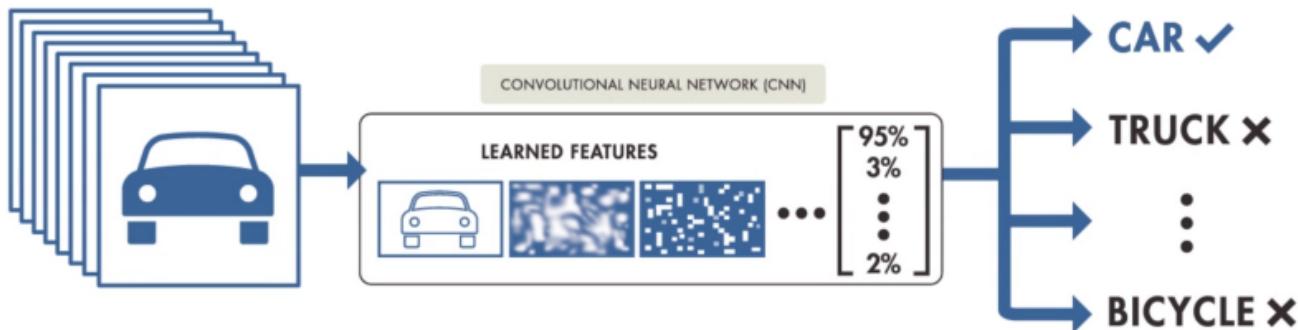


Fuseki RDF triplestore

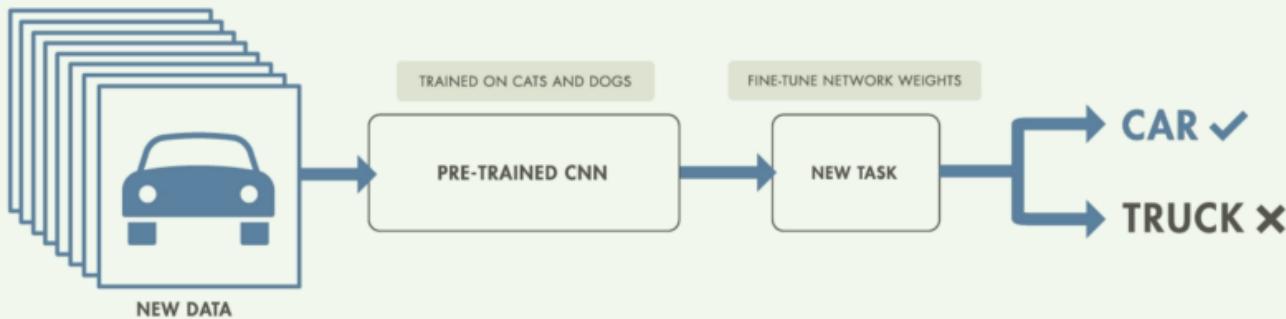
QUERY RESULTS	
Table	Raw Response
Download	
Showing 1 to 8 of 8 entries	
	<input type="text"/> Search: <input type="button"/> Show 50 entries
article	definition
1 < https://w3id.org/eb/l/Article/9929192893804340_144850368_PAISLEY_0 >	"a town of Scotland, in the county of Renfrew, six miles west of Glasgow."
2 < https://w3id.org/eb/l/Article/9929192893804340_144850366_ARGYLESHIRE_0 >	"a county of Scotland, lying westward of Glasgow, and comprehending the countries of Lorn, Cowal, Knapdale, Kintyre, together with the islands Mull, Jura, Iona, & Canna. It gives the title of duke to the noble family of Campbell."
3 < https://w3id.org/eb/l/Article/9929192893804340_144850367_INVERARY_0 >	"a parliament town of Scotland, in the county of Argyle, of which it is the capital, situated in Lochay, forty five miles north-west of Glasgow : W. long. 5° 0', N. lat 36° 28'."
4 < https://w3id.org/eb/l/Article/9929192893804340_144850367_HAMILTON_0 >	"a town of Scotland, in the county of Clydesdale, situated on the river Clyde, eleven miles south-east of Glasgow : W. long. 3° 0', N. lat. 55° 0"

Phase 3: Augmented Knowledge Graph with Deep Transfer Learning

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



Transformers

build passing

license Apache-2.0

website online

release v2.0.0

State-of-the-art Natural Language Processing for PyTorch and TensorFlow 2.0

😊 Transformers provides thousands of pretrained models to perform tasks on texts such as classification, information extraction, question answering, summarization, translation, text generation, etc in 100+ languages. Its aim is to make cutting-edge NLP easier to use for everyone.

😊 Transformers provides APIs to quickly download and use those pretrained models on a given text, fine-tune them on your own datasets then share them with the community on our [model hub](#). At the same time, each python module defining an architecture can be used as a standalone and modified to enable quick research experiments.

😊 Transformers is backed by the two most popular deep learning libraries, [PyTorch](#) and [TensorFlow](#), with a seamless integration between them, allowing you to train your models with one then load it for inference with the other.

Phase 3: Augmented Knowledge Graph with Deep Transfer Learning



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

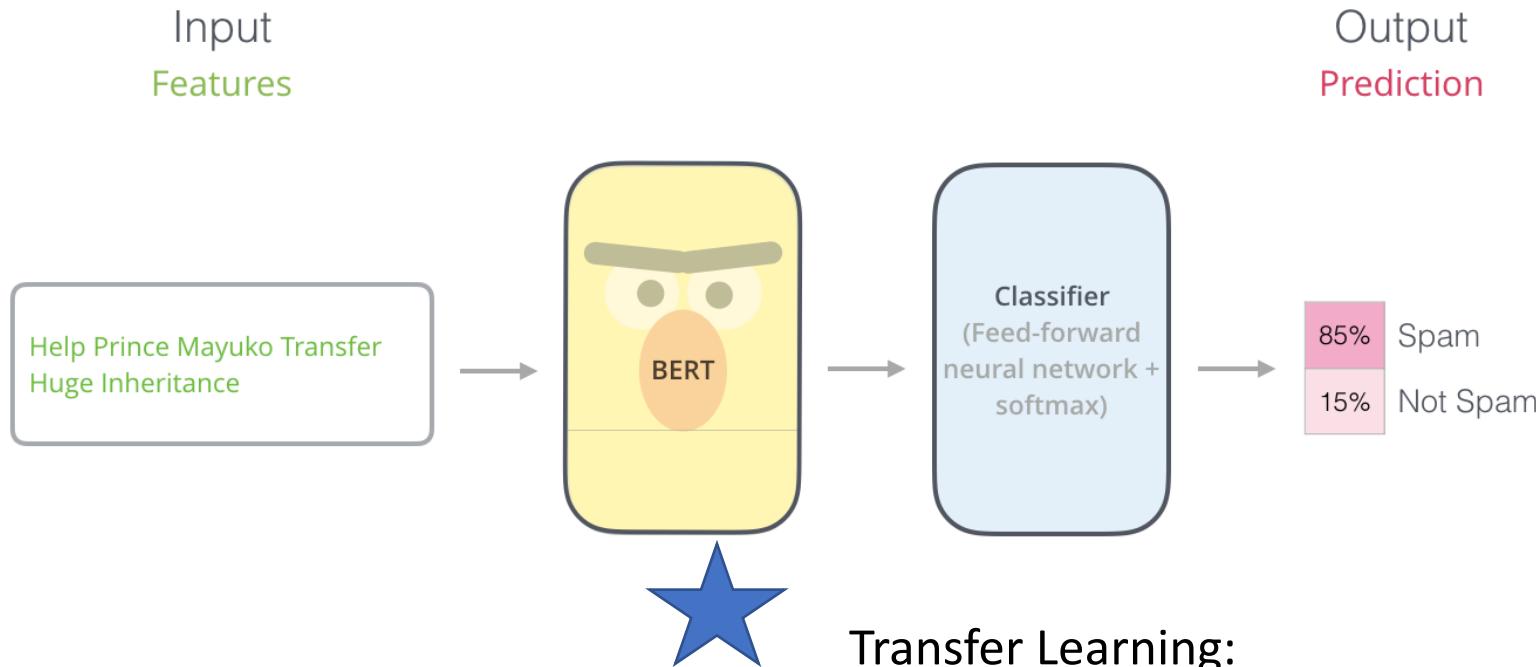
A High-Level Look

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 Transferring Learning

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



Transfer Learning:
We have not built the
model/transformer

Phase 3: Augmented Knowledge Graph with Deep Transferring Learning

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*](#) + *ElmoslstmChecker*
- **term evolution:** Checking how a term has changed over the years
 - transformer: [*all-mpnet-base-v2*](#)
- **summarization:** Summarizing the text of a topic term (XLNET)
 - transformer: [*XLNeT*](#)

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

Phase 3: Augmented Knowledge Graph with Deep Transferring Learning

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
- **topic modelling:** Clustering terms into topics
- **term similarity:** Comparing text & semantic similarity
- **spelling checking:** Finding misspelling/ocr errors and fixing them
- **term evolution:** Checking how a term has changed over the years
- **summarization:** Summarizing the text of a topic term

Example: Spelling Checking → *Lewis* Term

Original Definition

the mort northerly of any of the w eftern islands of Scotland, lying in 8\u00b0 odd minutes W. long, and between 58\u00b0 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 .



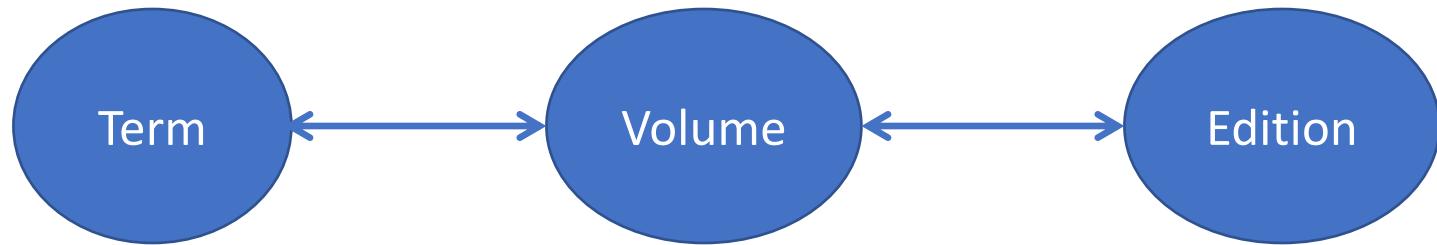
Compute Difference

the morst northerly of any of the w eftern islands of Scotland, lying in 8\u00b0 odd minutes W. long, and between 58\u00b0 and 59 0 odd minutes N. latnd .

Phase 3: Augmented Knowledge Graph with Deep Transferring Learning

Two types of “queries” to run against the EB Knowledge Graph 2.0 :

- **Type 1: Extracting information from the EB Knowledge → SPARQL → We “just” navigate through the KG to get the desired information.**

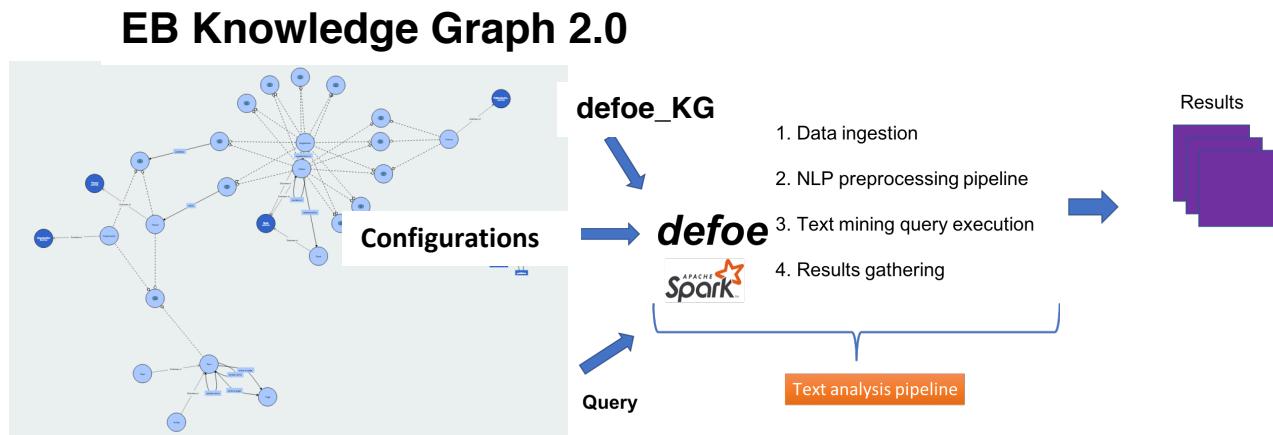


Example: Given a **Term** (e.g., *Edinburgh*), I can get the **Edition** Information (e.g., *Edition Title*)

- **Type 2: Processing information from the EB Knowledge → defoe → We are going to process further the definitions from the selected terms.**
 - But for doing this we needed to do some work on defoe first → Phase 4

Phase 4: Defoe and Knowledge Graph

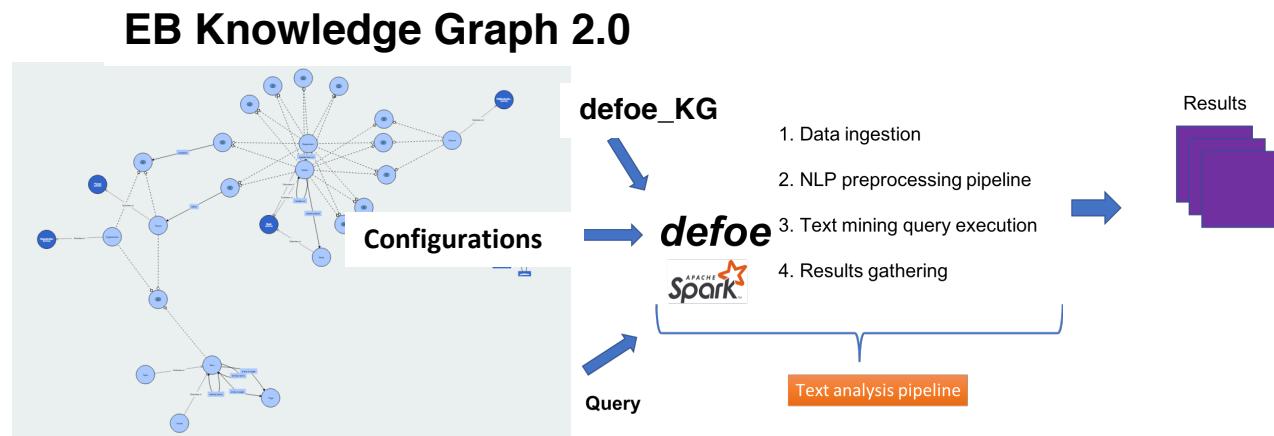
- 4.1. Created a new **KG defoe connector** (based in SPARQL) to run defoe queries using the EB Knowledge Graph as a source of data → **defoe_KG**
- 4.2. Improved **defoe queries to be fully configurable**: different filtering options, target, lexicon, etc.



Phase 4: Defoe and Knowledge Graph

4.3 defoe text mining queries:

- 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.



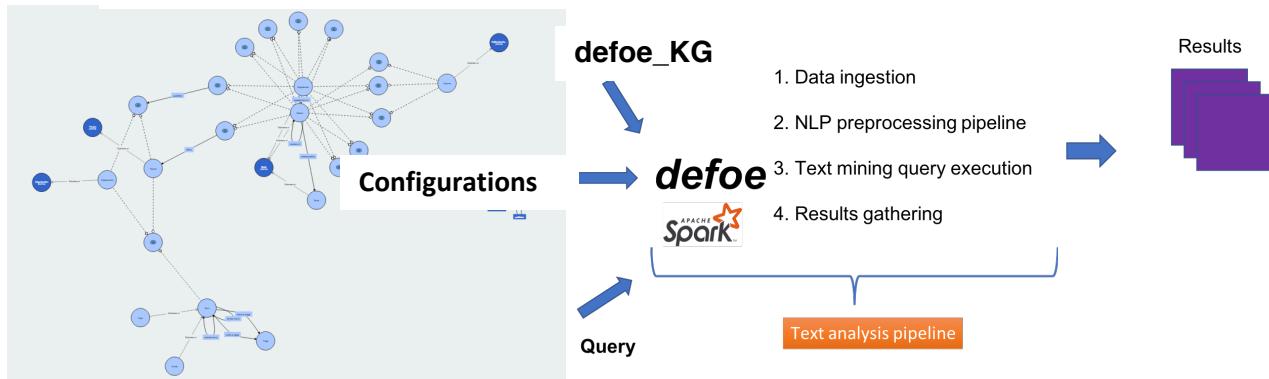
Phase 4: Defoe and Knowledge Graph

4.4 Created **new defoe ML/NLP queries** (using also deep learning transformers)

- sentiment analysis: Calculate the **sentiment** analyses of selected terms.
- topic modelling: Calculate the **topic** of selected terms. Filtering options.
- spelling checker: Check the **spelling** of selected terms. Filtering options.

(TO-DO)

EB Knowledge Graph 2.0



Phase 5: Flask Web-Application

5.1 Web-User Interface:

- So, 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
- Functionalities:
 - Term Search
 - Term Similarity
 - Topic Modelling
 - Spelling Checker
 - Defoe queries
 - EB Details
 - Knowledge Graph Resources Visualizations

5.2 Flask + JQuery + JavaScript + HTML + CSS + Web-services + Plotly

Summary

1. EB ALTO & METS + **defoe** → Extracted **Terms & Metadata (v.1)**
2. Extracted terms and metadata (v.1) + **postprocessing scripts**:
 - 2.1 Final set of **Terms & Metadata (v.2)**
 - 2.2 **EB ontology**
3. Terms & Metadata (v.2) + EB ontology → **EB Knowledge Graph 1.0**
4. EB Knowledge Graph 1.0 + Transformers → **EB Knowledge Graph 2.0**
5. EB Knowledge Graph 2.0 + defoe → **defoe_KG**
6. **defoe_KG + configurable defoe queries**
7. EB Knowledge Graph 2.0 + defoe_KG + defoe queries (text mining/ NLP)
8. EB Knowledge Graph Queries:
 - 8.1 **Type 1: SPARQL queries**: Extracting the EB KG data
 - 8.2 **Type 2: defoe queries**: Processing the EB KG data
9. Flask-web app: User-Interaction with the EB KG
 - 9.1 **Abstractions to SPARQL and defoe queries & download of results & visualizations**

Interacting with the EB Knowledge Graph

Term Search Term Similarity Topic Modelling Spell Checker Term Evolution EB Details Visualization of Resources Text Mining Platform

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 no term is introduced, it will search for the first term in the Encyclopaedia.

Results for **ABACUS**.

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 8

« 1 2 3 4 »

URI	Year	Edition	Volume	Start Page	End Page	Term Type	Definition/Summary	Related Terms	Topic Modelling	Sentiment_Score	Advanced Options
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_0	1771	1	1	15	15	Article	a table strewed over with dust or sand, upon which the ancient mathematicians drew their figures, It... More	-1_other_we_their_these	POSITIVE_0.99	Spell Checker Term Similarity Term Evolution	
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1	1771	1	1	15	16	Article	in architeflure, signifies the superior part or member of the capital of a column, and serves as a k... More	11_building_architecture_columns_beam	POSITIVE_1.00	Spell Checker Term Similarity Term Evolution	

« 1 2 3 4 »

Interacting with the EB Knowledge Graph

Type 1

Type 2

Term Search

Term Similarity

Topic Modelling

Spell Checker

Term Evolution

EB Details

Visualization of Resources

Text Mining Platform

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 no term is introduced, it will search for the first term in the Encyclopaedia.

Results for **ABACUS**.

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 8

« 1 2 3 4 »

URI	Year	Edition	Volume	Start Page	End Page	Term Type	Definition/Summary	Related Terms	Topic Modelling	Sentiment_Score	Advanced Options
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_0	1771	1	1	15	15	Article	a table strewed over with dust or sand, upon which the ancient mathematicians drew their figures, It... More	-1_other_we_their_these	-1_building_architecture_columns_beam	POSITIVE_0.99	Spell Checker Term Similarity Term Evolution
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1	1771	1	1	15	16	Article	in architecture, signifies the superior part or member of the capital of a column, and serves as a k... More	11_building_architecture_columns_beam	POSITIVE_1.00	Spell Checker Term Similarity Term Evolution	

Note: Two types of EB-KG queries

- **Type 1:** Extracting information from EB-KG – Navigating across the KG – **SPARQL queries**
- **Type 2:** Processing information from EB-KG -- Processing in parallel further the Term's definitions (stored in KG) – **DEFOE queries**

« 1 2 3 4 »

Interacting with the EB Knowledge Graph

Term Similarity

Enter an **URI of a term or some text** that you would like to search similar terms for. If no term is introduced, it will search for similar terms of the first term in the Encyclopaedia.

Details of this resource

URI	Edition	Year	Volume	Term	Original Definition	Topic Modelling	Sentiment_Score
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1	1	1771	1	ABACUS	in architeflure, signifies the superior part or member of the capital of a column, and serves as a k... More	11_building_architecture_columns_beam	POSITIVE_1.00

The first **20** most similar results. The results are sorted by **similitud rank**.

Note that if you click over an **URI** in this table, it will search for **similar terms** of that resource in this 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. 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 - 10 records in total 20

« 1 2 »

URI	Edition	Year	Volume	Term	Definition	Topic Modelling	Similitud Rank	Sentiment_Score
https://w3id.org/eb/i/Article/9929192893804340_144850366_ABACUS_1	1	1773	1	ABACUS	in architedture, signifies the superior part or member of the capital of a column, and serves as a k... More	11_building_architecture_columns_beam	0.97762585	POSITIVE_1.00
https://w3id.org/eb/i/Article/992277653804341_144133901_FUSAROLE_2	1	1771	2	FUSAROLE	in archite&ure, More	11_building_architecture_columns_beam	0.67097366	POSITIVE_0.99

Interacting with the EB Knowledge Graph

Topic Modelling

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 no topic is introduced, it will use the first topic modelling, which name starts with '0'.

We have 127 topics modelling items.

206 terms found for the topic 11_building_architecture_columns_beam

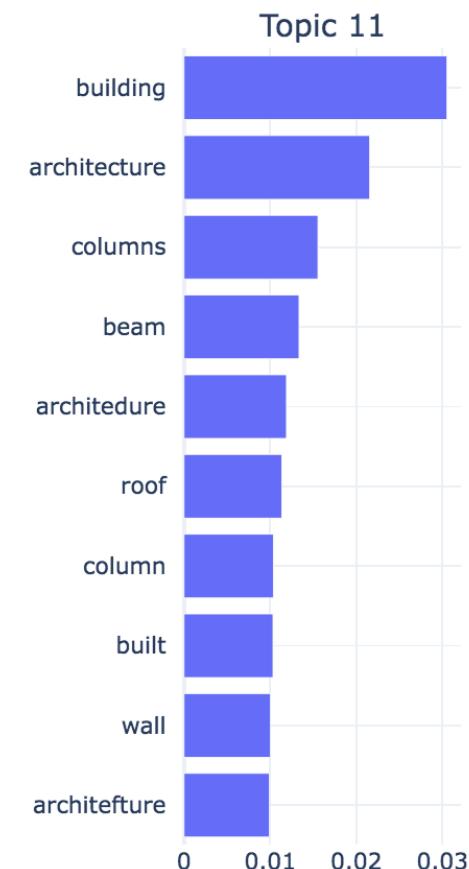
Note that if you click over an **URI** in this table, it will take you to the [Visualization of Resources page](#). However, if you instead click over a **term**, it will take you to the [Term Search page](#), showing all the searching results for that term.

displaying 1 - 10 records in total 206

« 1 2 3 4 5 ... 20 21 »

URI	Edition	Year	Volume	Term	Definition	Sentiment_Score
https://w3id.org/eb/i/Article/9929192893804340_144850368_WAINSCOT_0	1	1773	3	WAINSCOT	in building, the timber-work that serves to line the walls of a room, being usually made in pannels,... More	NEGATIVE_0.99
https://w3id.org/eb/i/Article/992277653804341_144133902_HOUSE_0	1	1771	2	HOUSE	a habitation, or place built with conveniences for dwelling in. See Architecture. House, in aitrolo... More	POSITIVE_0.99
https://w3id.org/eb/i/Article/9929192893804340_144850367_CALOTTE_1	1	1773	2	CALOTTE	in architedure, a round cavity or depprefre, in form of a cap or cup, lathed and plastered, used to... More	POSITIVE_0.99
https://w3id.org/eb/i/Article/9929192893804340_144850366_BINN_0	1	1773	1	BINN	in country affairs, a place boarded up to put	NEGATIVE_1.00
https://w3id.org/eb/i/Article/992277653804341_144133902_DECORUM_0	1	1771	2	DECORUM	in. architedure, is the fuitableness of a building, and the several parts and ornaments thereof, to... More	POSITIVE_1.00
https://w3id.org/eb/i/Article/9929192893804340_144850368_PROFILE_0	1	1773	3	PROFILE	in architedufe. tte draught of a building, fortification, be. wherein 1 are expfeffed the several n... More	NEGATIVE_0.96

Topic Word Scores



Interacting with the EB Knowledge Graph

Spell Checker

Enter an **URI of a term** that you would like to check its spelling. If no term is introduced, it will check the spelling of the first term in the Encyclopaedia.

Place an URI	<input type="button" value="🔍"/>			
URI	Edition	Year	Volume	Term
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1	1	1771	1	ABACUS

Original 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^{is} generally square; but in the Corinthian and Composite, its four sides are arched in Aroids, and embellished in the middle with ornament, as a rose or other flower, Scamozzi uses abacus for a concave moulding on the capital of the Tuscan pedestal; and Palladio calls the plinth above the echinus, or boulton, 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 Composite , its four sides are arched in Aroids , and embellished in the middle withornament , as a rose or other flower , Scamozzi uses abacus for a concave moulding on the capital of the Tuscan pedestal ; and Palladio calls the plinth above the echinus , or boultin , in the Tufean and Doric orders , by the same name . See plate I. fig . i. and



Cleaned Definition

in architefcture, 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^{is} generally square; but in the Corinthian and Compofite, its four sides are arched in Aroids, and embellished in the middle withornament, as a rose or other flower, Scamozzi uses abacus for a concave moulding on the capital of the Tuscan pedefistal; and Palladio calls the plinth above the echinus, or boultin, in the Tufean and Doric orders, by the same name. See plate I. fig. i. and

Interacting with the EB Knowledge Graph

Term Evolution

Enter an **URI of a term** that you would like to see its evolution.

Details of this resource

URI	Edition	Year	Volume	Term	Original Definition	Topic Modelling	Sentiment_Score
https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1	1	1771	1	ABACUS	in architeflure, signifies the superior part or member of the capital of a column, and serves as a k... More	11_building_architecture_columns_beam	POSITIVE_1.00

The evolution of **ABACUS** can be seen in this table.

URI	Edition	Year	Volume	Term	Definition	Topic Modelling	Similitud Rank	Sentiment_Score
https://w3id.org/eb/i/Article/9929192893804340_144850366_ABACUS_1	1	1773	1	ABACUS	in architedture, signifies the superior part or member of the capital of a column, and serves as a k... More	11_building_architecture_columns_beam	0.97762585	POSITIVE_1.00

Select Edition

Edition 1 Year 1771

Select Volume

1 A-B



Returned details for Edition 1 Year 1771 and Volume 1 A-B

Edition	Details
Year	1771
Edition Number	1
Edition URI	< https://w3id.org/eb/i/Edition/992277653804341 >
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	< https://w3id.org/eb/i/Volume/992277653804341_144133901 >
Volume Title	Encyclopaedia Britannica; or, A dictionary of arts and sciences, compiled upon a new plan
Volume Letters	A-B
Volume Permanent URL	https://digital.nls.uk/144133901
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

Interacting with the EB Knowledge Graph

Visualization of Resources

Visualization RDF graph for the resource <https://w3id.org/eb/i/Article/992277653804341_144133901_ABACUS_1>



Interacting with the EB Knowledge Graph

Text Mining Platform

Select your **text mining query** that you would to run. Depending on the selected query, different configurations will be available for you.

- 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

[Configure your query](#)

Interacting with the EB Knowledge Graph

Example 1

Text Mining Platform

The **frequency_keysearch_by_year** query counts number of terms or times in which appear keywords or keysentences and groups the results by years .

Preprocess Treatment

- None
- Normalize
- Normalize & Numbers
- Normalize & Lemmatize
- Normalize & Stemming

Upload your lexicon file with your words or sentences to be used in the query. One word/sentence per line.

[Browse...](#) peace_war.txt

Select the hit count

- Term: Count the number of terms in which any of your lexicon words or sentences has been found
- Word: Count the number of times in which any of your lexicon words or sentences has been found

Target words or sentences allows for filtering further the terms in your search

Introduce Target words/sentences separated by " " (Optional).

peace accord, peace agreement,
ceasefire, war

You can also filter your search by restricting the search to a period of time. Specifying the "start_year" it will select only terms from that period onwards. Specifying an "end_year" it will select terms from that year backwards. If not period time is specified, it will search across all the years.

Start Year (Optional) 1771 End Year (Optional) 1773

Click here when you are ready to submit your query.

[Submit Query](#)

Click here, if you would like to start again, and select a different query.

[Reset Configurations](#)

Select the target filter

- OR: Select terms which contains **any** of the target words/sentences
- AND: Select terms which contains **all** the target words/sentences

Interacting with the EB Knowledge Graph

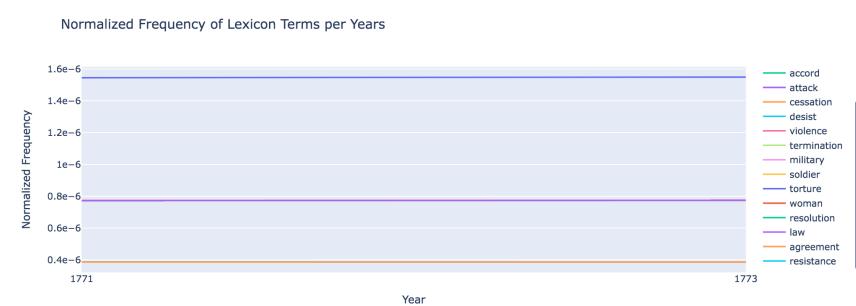
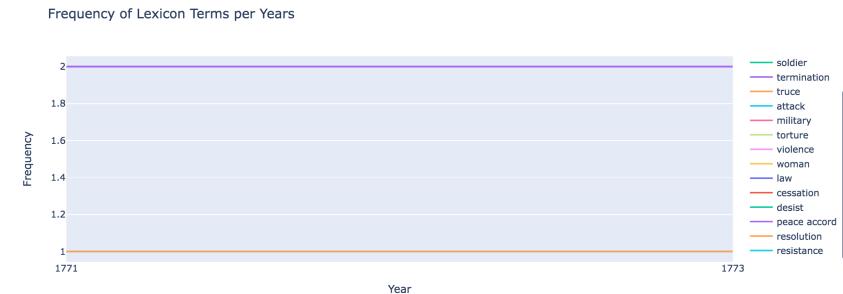
Results of frequency_keysearch_by_year query

Here we can see the frequency of your lexicon words/sentences taking into account the target words/sentences (if any) and other configurations, sorted by years.

Year	Keyword : Frequency
1771	<ul style="list-style-type: none">• accord : 1• agreement : 1• soldier : 1• termination : 1• truce : 2• attack : 1• military : 1• torture : 1• violence : 1• woman : 1• law : 2• cessation : 1• desist : 1• peace accord : 2• resolution : 1
1773	<ul style="list-style-type: none">• peace accord : 2• resolution : 1• law : 2• cessation : 1• desist : 1• truce : 2• attack : 1• military : 1• resistance : 1• torture : 1• violence : 1• woman : 1• accord : 1• agreement : 1• soldier : 1• termination : 1

[Download Results](#)

Example 1



Interacting with the EB Knowledge Graph

Text Mining Platform

Example 2

The **frequency_keysearch_by_year** query counts number of terms or times in which appear keywords or keysentences and groups the results by years .

Preprocess Treatment

- None
- Normalize
- Normalize & Numbers
- Normalize & Lemmatize
- Normalize & Stemming

Upload your lexicon file with your words or sentences to be used in the query. One word/sentence per line.

[Browse...](#) sport.txt

Select the hit count

- Term: Count the number of terms in which any of your lexicon words or sentences has been found
- Word: Count the number of times in which any of your lexicon words or sentences has been found

You can also filter your search by restricting the search to a period of time. Specifying the "start_year" it will select only terms from that period onwards. Specifying an "end_year" it will select terms from that year backwards. If no period time is specified, it will search across all the years.

Start Year (Optional)

End Year (Optional)

Interacting with the EB Knowledge Graph

Results of frequency_keysearch_by_year query

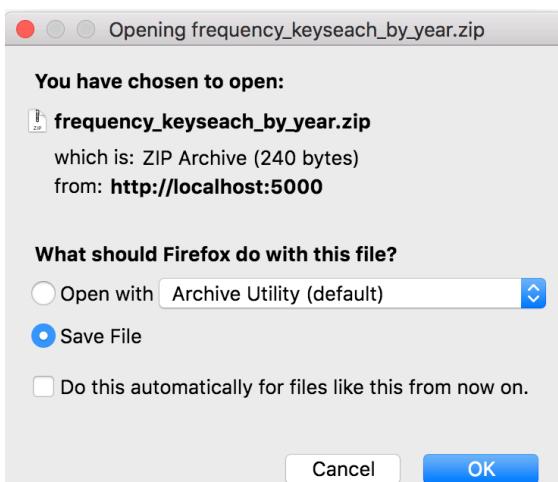
Your configurations:

- preprocess: normalize
- target_filter: or
- start_year: 1771
- end_year: 1773
- hit_count: word
- data : sport.txt

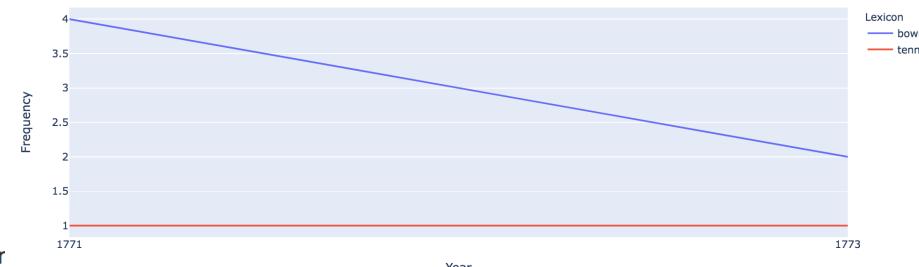
Here we can see the frequency of your lexicon words/sentences taking into account

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

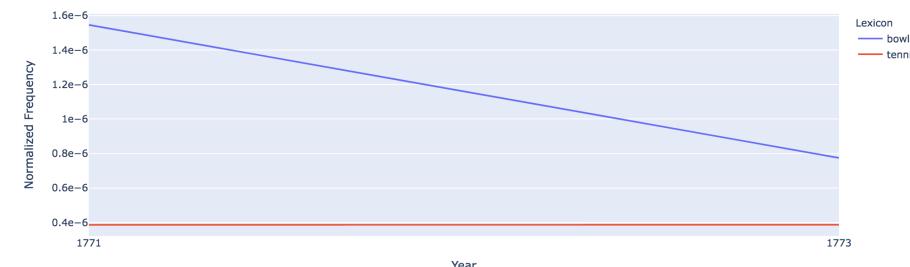
[Download Results](#)



Frequency of Lexicon Terms per Years



Normalized Frequency of Lexicon Terms per Years



```
frequency_keysearch_by_year.yml
1 1771:
2   -- bowls
3     - 4
4     -- tennis
5       - 1
6 1773:
7   -- bowls
8     - 2
9     -- tennis
10    - 1
```

Example 2

Interacting with the EB Knowledge Graph

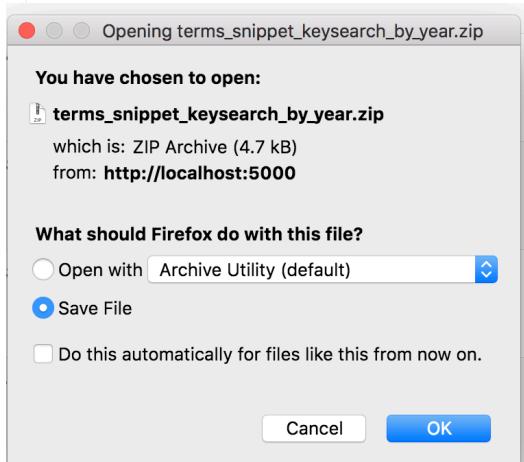
Results of terms_snippet_keysearch_by_year query

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 bellow 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
https://w3id.org/eb/i/Article/992277653804341_144133902_CANDLESTICK_0	• bowls	Spell Checker Term Similarity Term Evolution
https://w3id.org/eb/i/Article/992277653804341_144133902_CANDYING_0	• bowls	Spell Checker Term Similarity Term Evolution
https://w3id.org/eb/i/Article/992277653804341_144133901_BOWL_0		
https://w3id.org/eb/i/Article/992277653804341_144133901_BOWLING_0		
https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWLING_0		
https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWL_0		



```
terms_snippet_keysearch_by_year.yml x
1 [terms_details:
2   - '1773'
3     -- archive_filename: 144850366-mets.xml
4       edition: '1'
5       header: BOWL
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 fird thing to be observed in
11        - bowling is the right chufing your bowl which
12          term: BOWLING
13          title: Edition 1,1773
14          url: https://w3id.org/eb/i/Article/9929192893804340_144850366_BOWLING_0
15          volume: '1'
16        - archive_filename: 144850366-mets.xml
17          edition: '1'
18          header: BOWL
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            url: 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            url: https://w3id.org/eb/i/Topic/9929192893804340_144850368_METAPHYSICS_0
43            volume: '3'
44            - '1773'
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
```

Example 3

frances: Architecture

NSL Data Foundry Collections

