

# 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

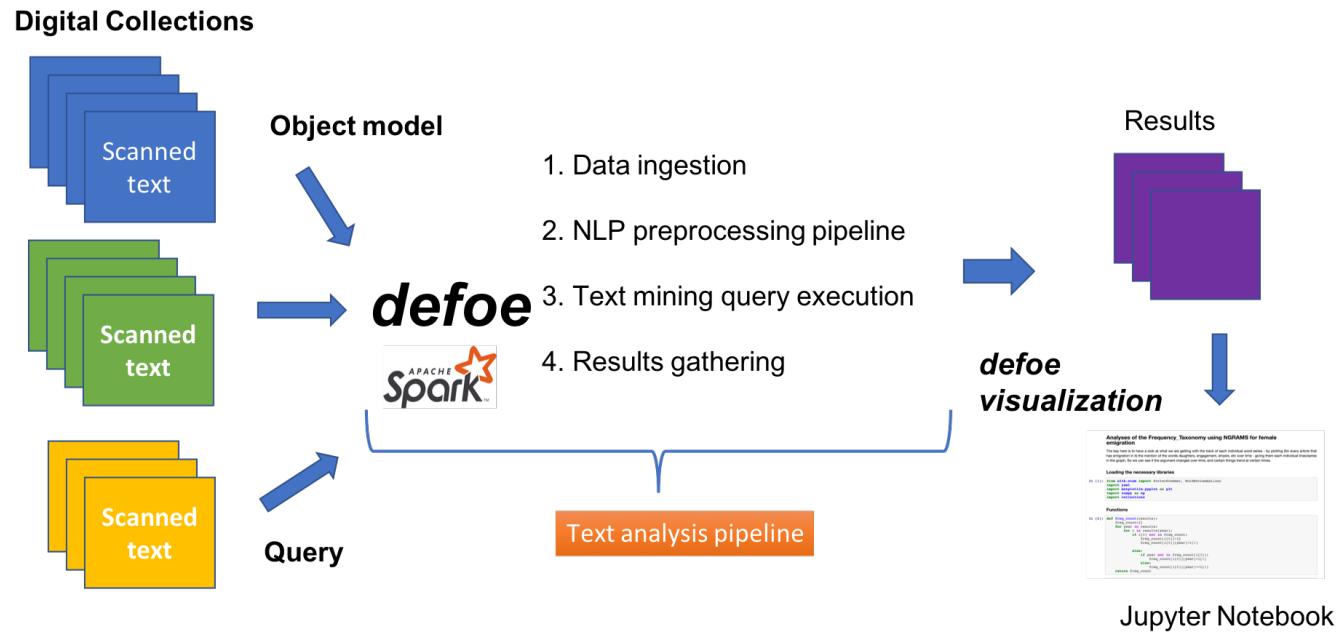
- Objective 1: New facilities to run more complex text analysis queries → ML/NLP techniques.
- Objective 2: Full integration with NLS Data Foundry → Hide the large-scale text mining complexity
- Using the **Encyclopaedia Britannica** as the core dataset

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

## The Journey

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

## Phase 1: Text Mining



METS & ALTO-XML

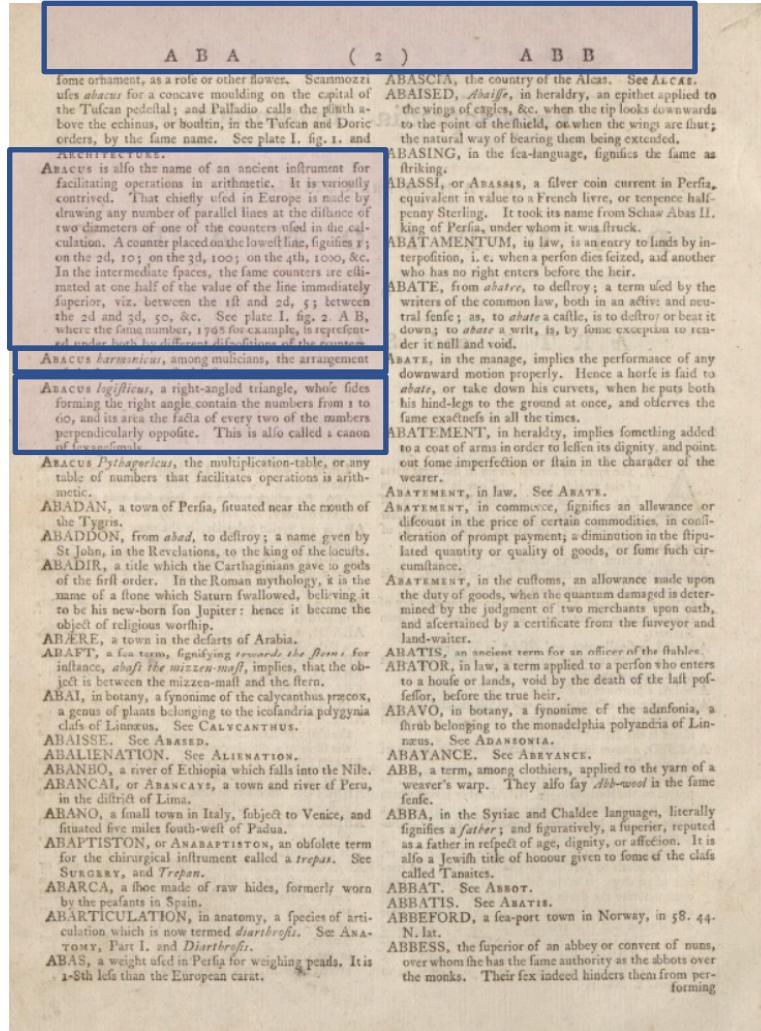
## Phase 1: Text Mining

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

**Extracted Terms Based on Heuristics --`> Pages Layout & Text & Headers → Different heuristics for different EB editions. Using ALTO-XML information.**

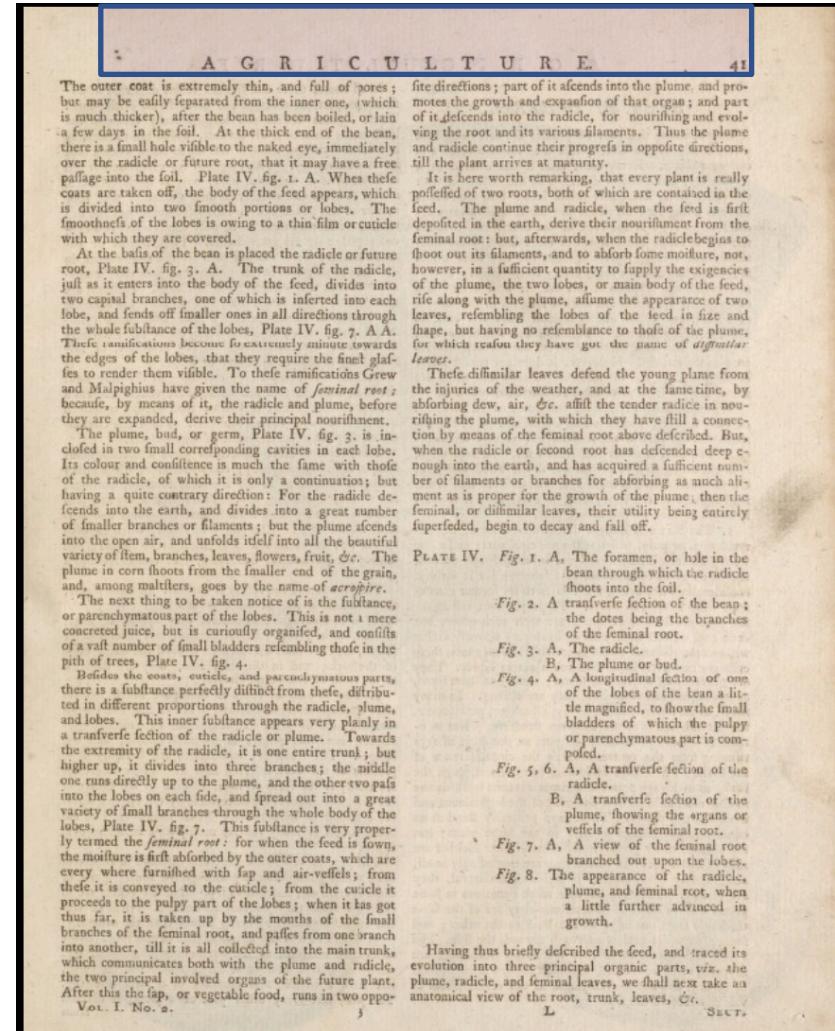
## Phase 1: Text Mining

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: Text Mining

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 + “,” .

AALEN, a bailiwick in the circle of Jaxt, in the kingdom of Württemberg. Its extent is 168 square miles, or 69,120 acres. It is watered by the river Kocher, has some lofty mountains in the southern part, and is most abundantly wooded. It produces but little corn, and neither fruit nor wine, but pastures a competent number of cattle. There are some iron mines worked. Many articles of wood-ware are produced, and some wool and cotton are spun. It contains one city, one market town, and 190 smaller towns and villages, with 17,899 inhabitants.

AALEN, a city, the capital of the bailiwick of the same name. Its chief trade is in woollens and in breweries, and some cotton is spun. It contains 2370 inhabitants. It is in Lat. 48° 47'. 26. N. Long. 10° 7. 27. E.

AALSMEEER, a town in the arrondissement of Amsterdam, in the province of North Holland. It is near the lake of Haarlem; celebrated for its strawberries; contains 1811 inhabitants, employed in making cotton goods.

AALTEN, a town in the arrondissement of Zutphen, and province of Gelderland, in the Netherlands, containing 3524 inhabitants.

AAM, or HAAM, a liquid measure in common use among the Dutch, containing 128 measures called *minglees*, each weighing nearly 36 ounces avoirdupois; whence the Aam contains 288 English, and 148½ pints Paris measure.

AAMADOT, a town of Norway, in the bailiwick of Hedenmark and see of Agderhus. It is situated on the river Glommen, has 2729 inhabitants, and some trade in making woollen and cotton caps.

AAAR, the name of two rivers; one in Switzerland, the other in Westphalia, in Germany. It is also the name of a small island in the Baltic.

AAARASSUS, in *Ancient Geography*, a town of Pisidia, in the Hither Asia, thought to be the Anassus of Ptolemy.

AAARAU, or ARAU, a circle in the canton of Aargau, Switzerland, containing the city from which its name is derived, and in other places, with 2260 houses and 11,893 inhabitants.

AAARAU, the chief city of the canton of the same name, on the banks of the river Aar, over which there is a covered bridge. It is well built, paved, and, at night, lighted. It has a handsome government house, a church, a hospital, 427 dwellings, and 3100 inhabitants, who are very industrious manufacturers. The chief pursuits are making silk ribbons, spinning and weaving cotton, some tanning and cutlery, bleaching, and casting cannon. It is in Lat. 47° 23'. 31. N. Long. 8° 4. 32. E.

AAARBURG, a city in Switzerland, belonging to the circle of Zofingen, and canton of Aargau. It stands at the confluence of the rivers Aar and Bagger, has a strongly fortified castle, the only one in Switzerland, which is the depot for military stores. The city contains 154 houses, and 1000 inhabitants, who make cotton goods and hosiery.

AAARDENBURG, a town in the arrondissement of Middleburg, in the province of Zeeland, in the kingdom of the Netherlands, with 1376 inhabitants.

AAARGA, or ARGOVIA, one of the cantons of Switzerland. It was originally a part of Berne, but by arrangements begun in 1798, and continued in 1803, it was erected into a separate and independent canton. It is bounded on the north by the river Rhine, on the east by Zurich, on the south-east by Zug, on the south by Lucerne, on the south-west by Berne, and on the west by Solothurn and Basle. Its extent is about 600 square miles, and its divisions are into eleven circles, which are again subdivided into forty-eight smaller ones. By the census taken in 1814, the number of inhabitants appeared to be 143,960, and

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.

AAARHUS, one of the sees (*stifts*) into which Denmark is divided. It is in the southmost 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.

AAARHUS, 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.

AAARHUS, 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

AARAU

Aargau  
Aaron

ABERRATION.

Aberration. Aberration.

mended 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 resolved 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, 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  $\gamma$  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 conjecture 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 could affect such an apparent motion as this in the star, 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 passed the zenith, and from thence to the southward, 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 this 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  $\gamma$  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  $\gamma$  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.

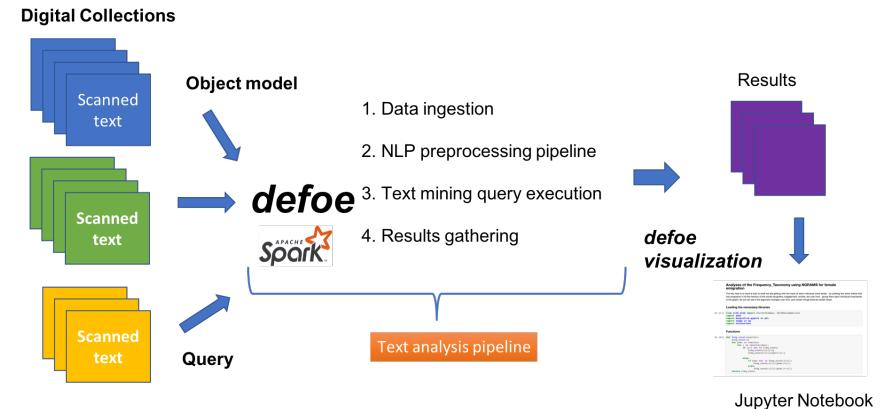
The great regularity of the observations left no room to doubt, but that there was some uniform cause by which this unexpected motion was produced, and which did not depend on the uncertainty or variety of the seasons of the year. Upon comparing the observations with each other, it was discovered that, 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, and he might with the more ease, and convenience 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  $\gamma$  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 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

Articles

Edition 7

Topic

# 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

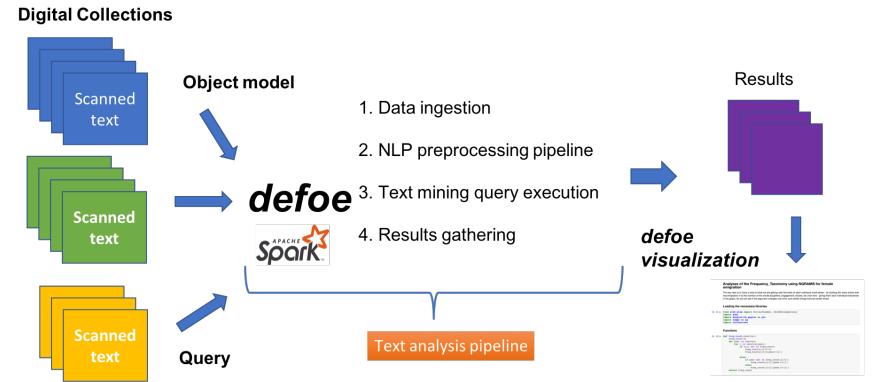
Extracted Term ABACATORES information - Edition 1, 1771, Vol A-B

## Phase 1: Text Mining

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

**Based on METS information**

# Extract collection Metadata (METS)



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

Metadata of some of the Volumes of Edition 3

## Phase 1: Text Mining

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

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

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

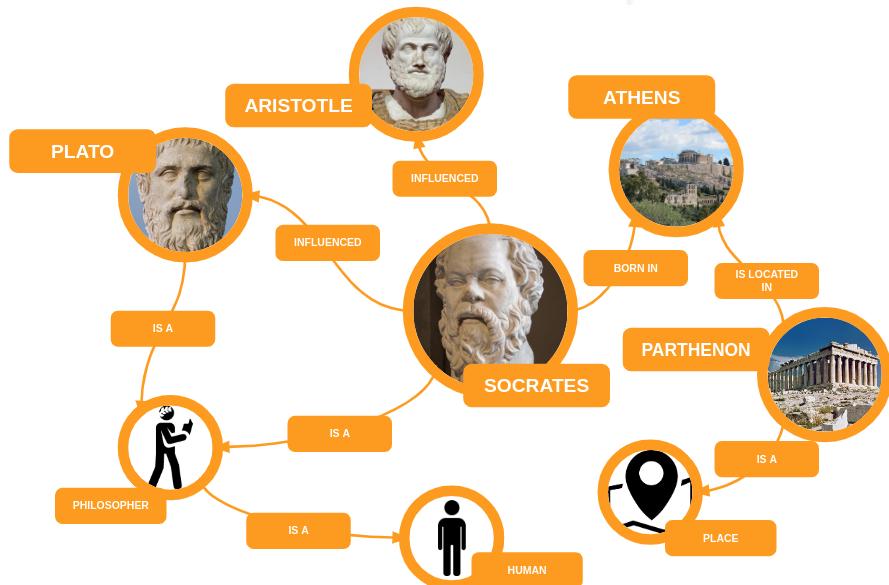
(\*) **Also Based on Heuristics --`>** Pages Layout & Text & Headers → different heuristics for different EB editions.

## 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

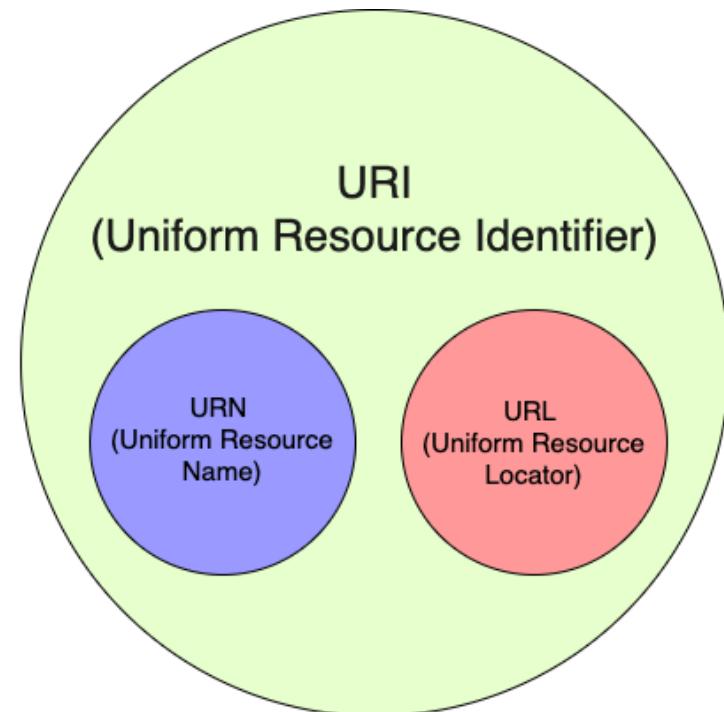
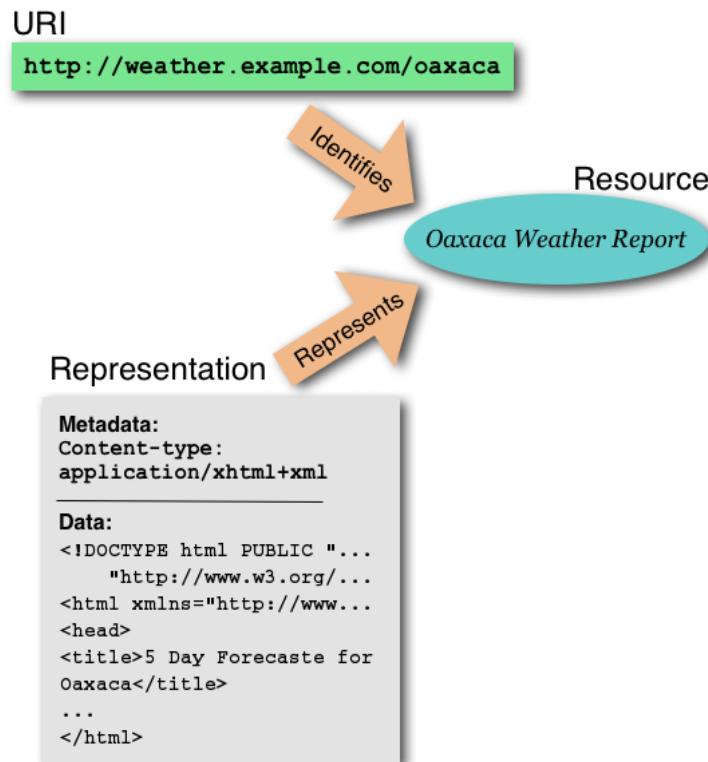
"The application of graph processing and graph DBMSs will grow at 100 percent annually through 2022 to continuously accelerate data preparation and enable more complex and adaptive data science."  
– Gartner's *Top 10 Data and Analytics Technology Trends for 2019*



## 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

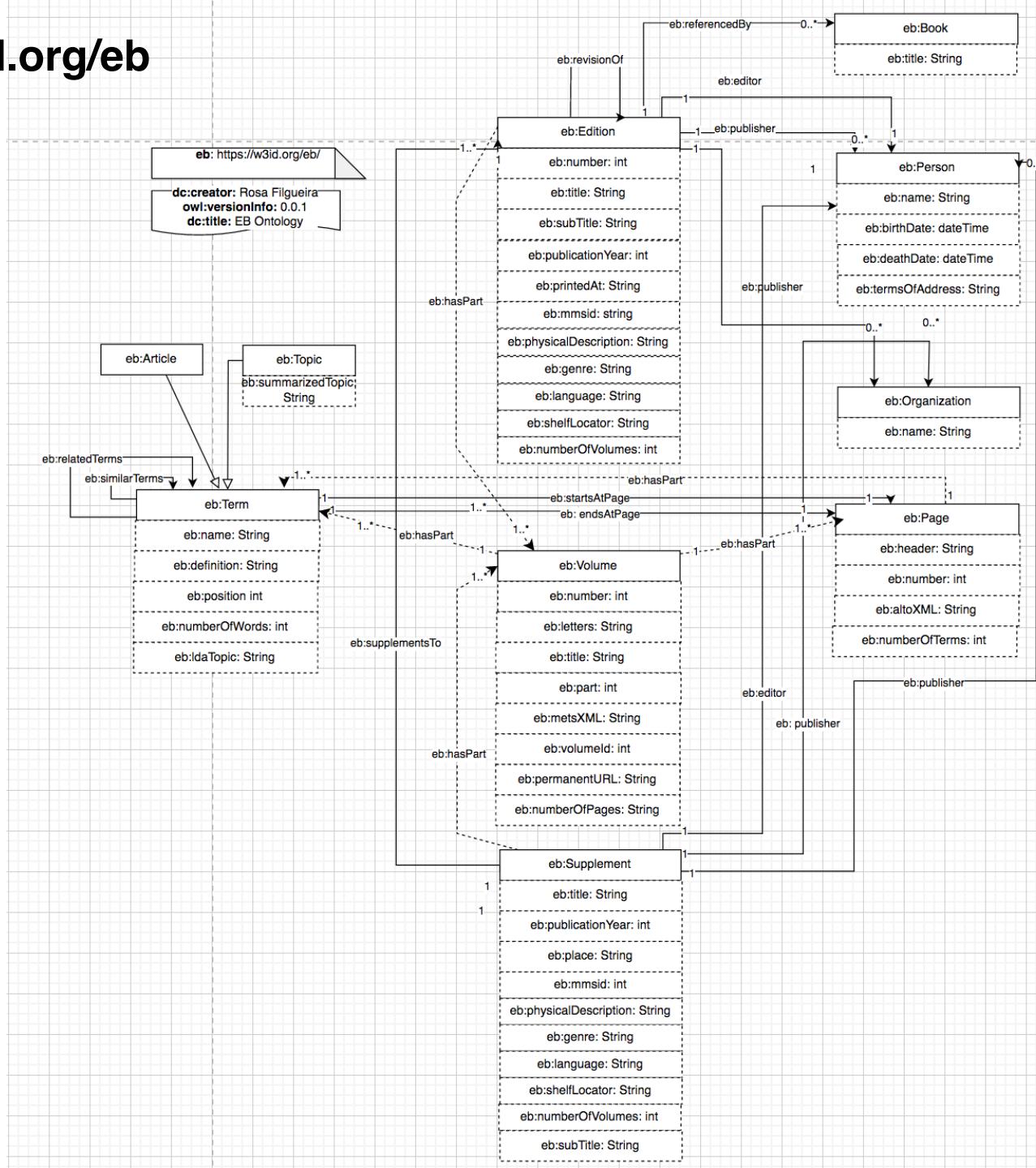
2.1. Create EB Ontology: To explain (give meaning) how our “thing” are related with each other.

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 →  
<https://w3id.org/eb/>

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

2.1. Created EB Ontology: The description of this ontology is available online at http: <https://w3id.org/eb/>

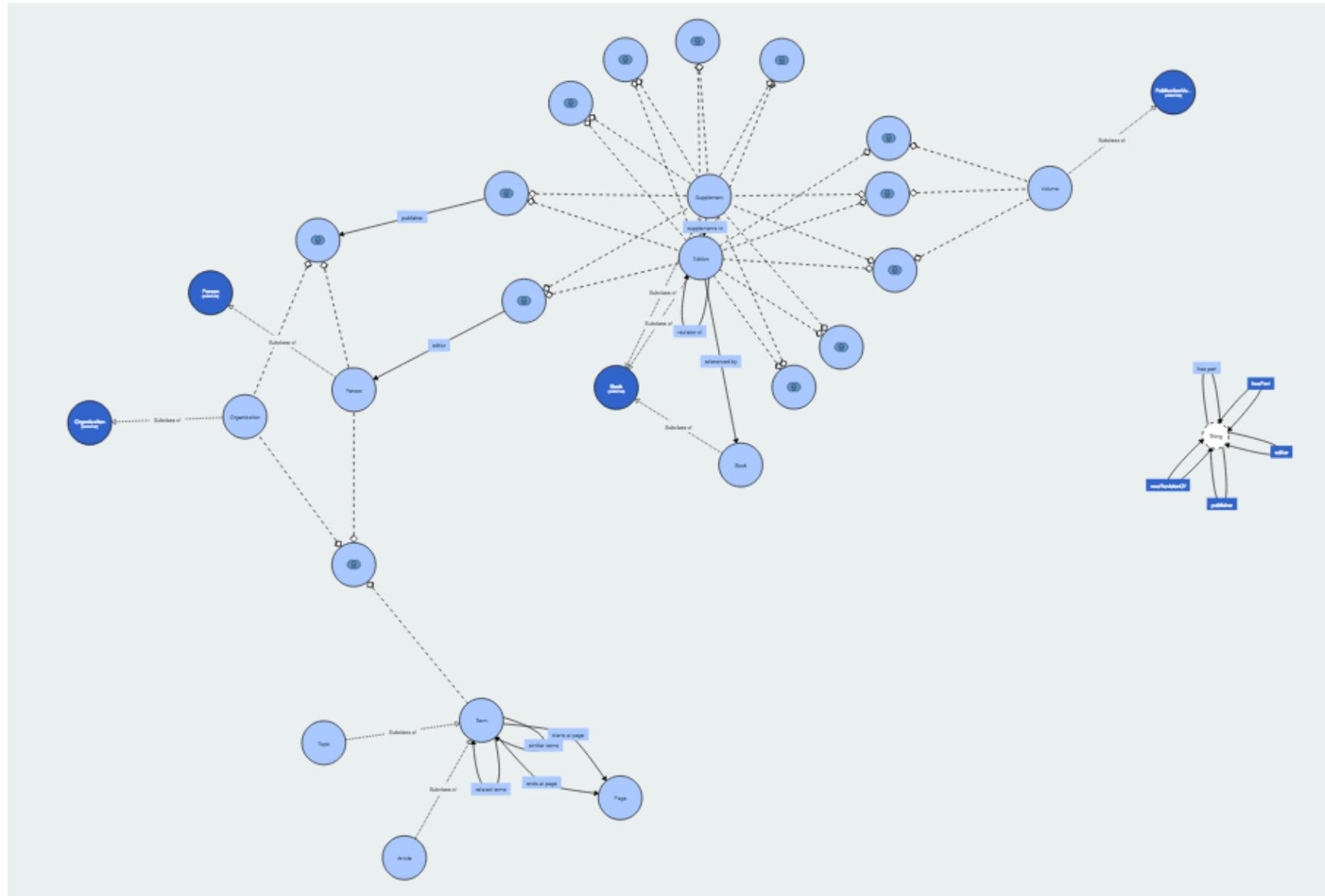
2.2 . Created **EB Knowledge Graph 1.0**: Populated the post-processed information (extracted **Terms & Metadata v.2**) into a RDF triplestore using the EB Ontology

- My RDF EB-data is stored in an Apache Jena FUSEKI SPARQL server – Used this Fuseki-docker image to set up my SPARQL server/end-point

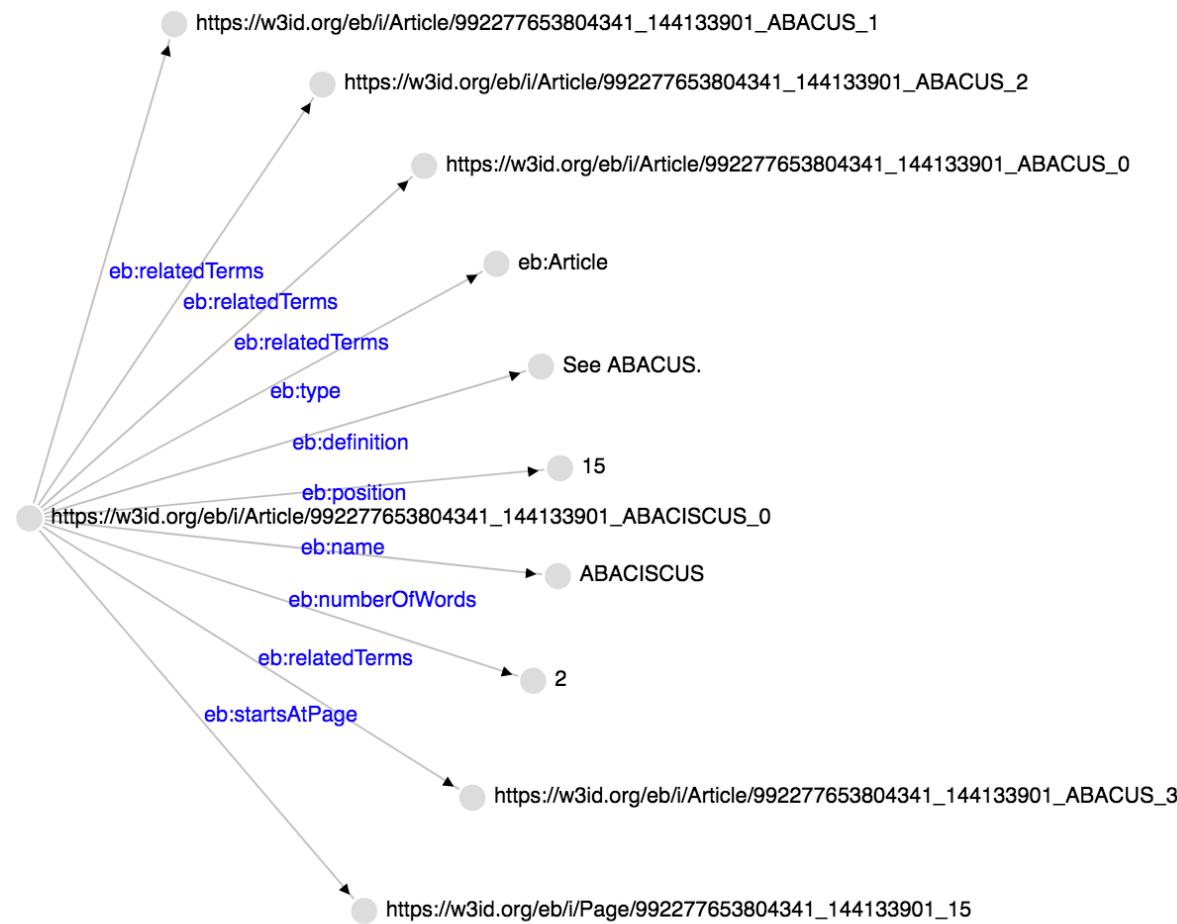
### **NOTE:**

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

# EB Knowledge Graph

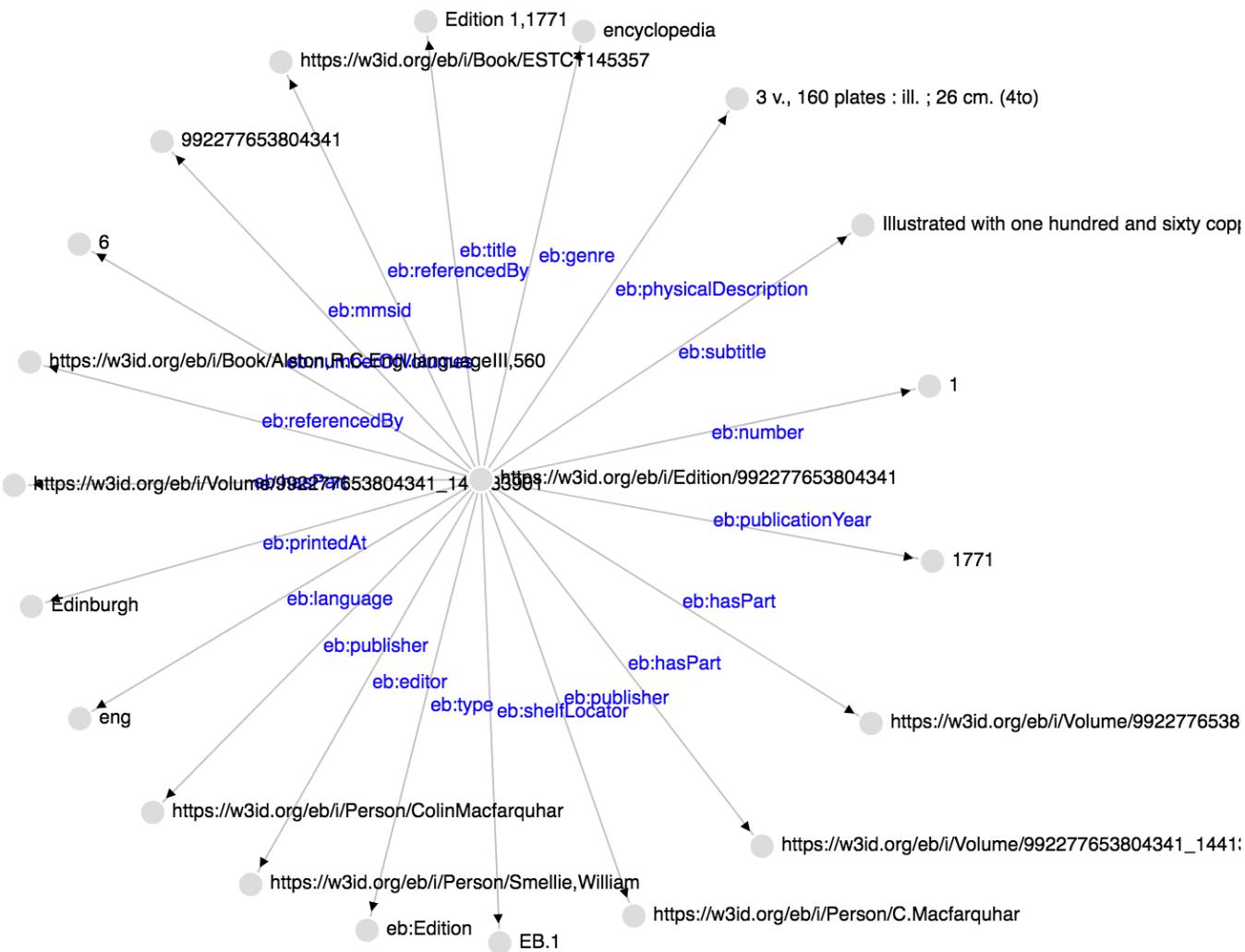


# EB Knowledge Graph



Term **ABACISCUS**, URI: [https://w3id.org/eb/i/Article/992277653804341\\_144133901\\_ABACUS\\_0](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.

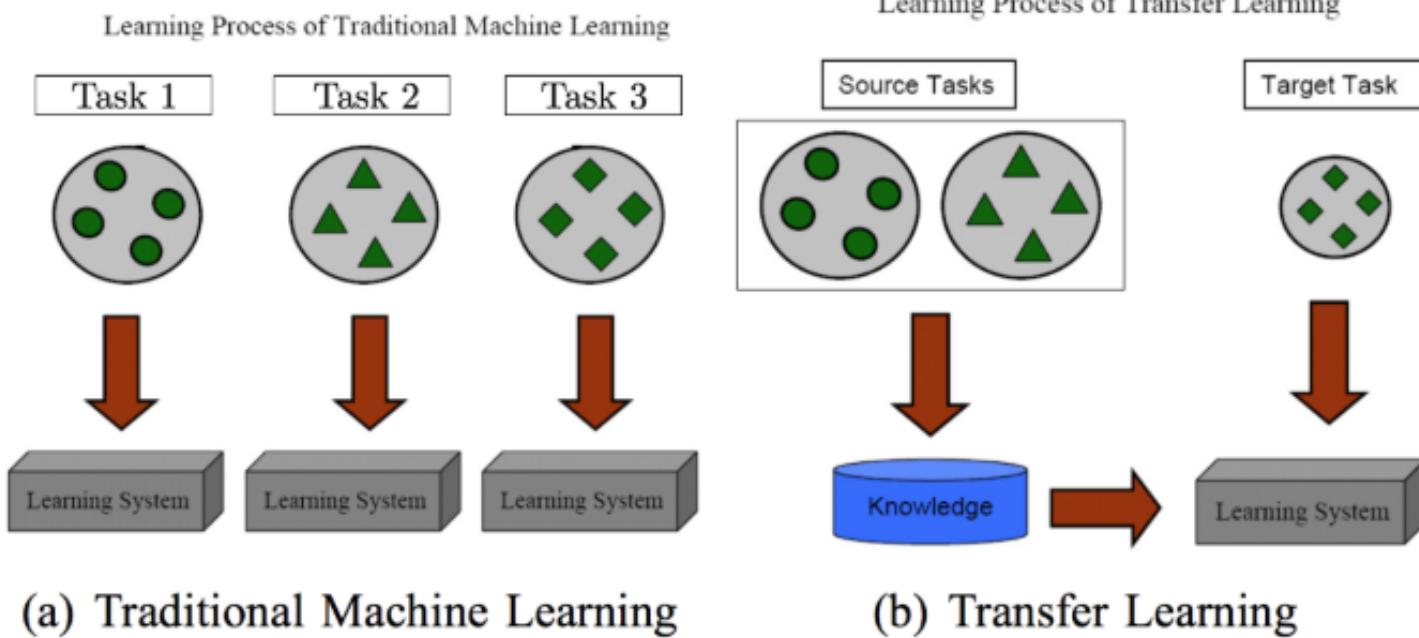
```
2 SELECT ?article ?definition
3 WHERE {
4     ?article a eb:Article .
5     ?article eb:definition ?definition
6     FILTER (CONTAINS(?definition, "Scotland"))
7     FILTER (CONTAINS(?definition, "Glasgow"))
8     OPTIONAL{FILTER CONTAINS(?definition, "Edinburgh") }
9 } LIMIT 10
10
```



Fuseki RDF triplestore

QUERY RESULTS	
article	definition
1 < <a href="https://w3id.org/eb/l/Article/9929192893804340_144850368_PAISLEY_0">https://w3id.org/eb/l/Article/9929192893804340_144850368_PAISLEY_0</a> >	"a town of Scotland, in the county of Renfrew, six miles west of Glasgow."
2 < <a href="https://w3id.org/eb/l/Article/9929192893804340_144850366_ARGYLESHIRE_0">https://w3id.org/eb/l/Article/9929192893804340_144850366_ARGYLESHIRE_0</a> >	"a county of Scotland, lying westward of Glasgow, and comprehending the countries of Lorn, Cowal, Knapdale, Kintyre, together with the islands Mull, Jura, Iona, Islay. It gives the title of duke to the noble family of Campbell."
3 < <a href="https://w3id.org/eb/l/Article/9929192893804340_144850367_INVERARY_0">https://w3id.org/eb/l/Article/9929192893804340_144850367_INVERARY_0</a> >	"a parliament town of Scotland, in the county of Argyll, 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 < <a href="https://w3id.org/eb/l/Article/9929192893804340_144850367_HAMILTON_0">https://w3id.org/eb/l/Article/9929192893804340_144850367_HAMILTON_0</a> >	"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



## 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:

- **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\*](#)

## 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 .



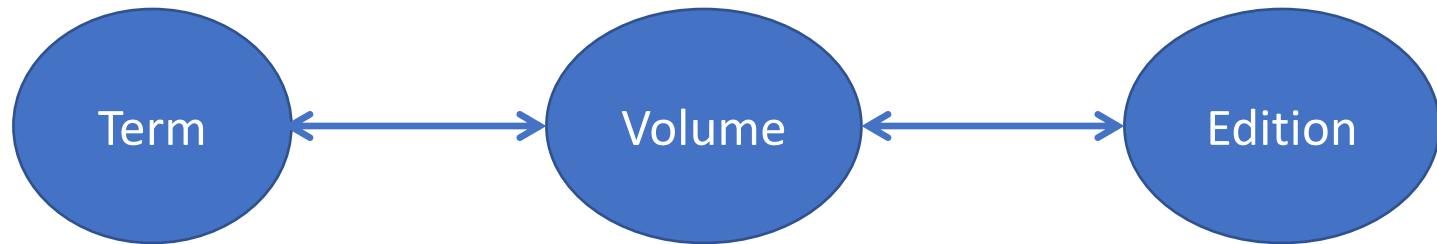
Compute Difference

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” 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.**

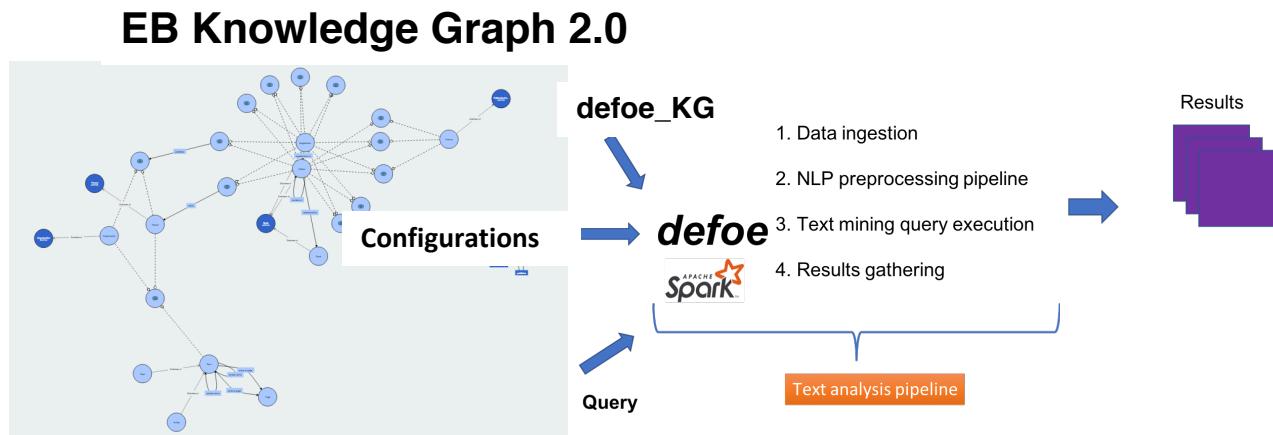


Example: Given a **Term** ( e.g., *Edinburgh*), I can get the **Edition** Information (e.g., *Edition Title*)

- **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

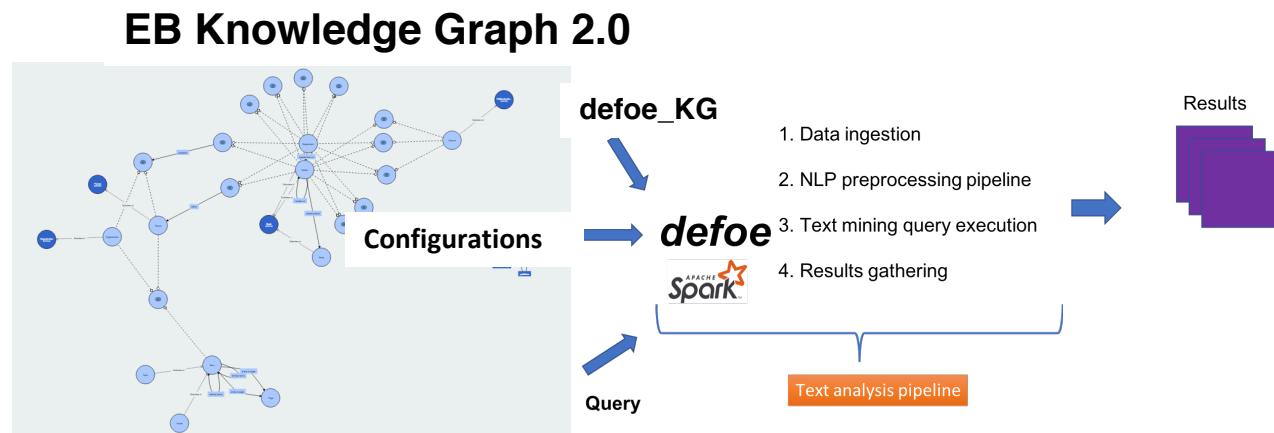
- 4.1. Created a new **KG defoe connector** (based in SPARQL) to run defoe queries using the EB Knowledge Graph as a source of data → **defoe\_KG**
- 4.2. Improved **defoe queries to be fully configurable**: different filtering options, target, lexicon, etc.



## 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 snippet of definition** in which appear keywords or keysentences and group by years. Several filtering options, including the snippet size.
- publication normalization: **Extract 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.

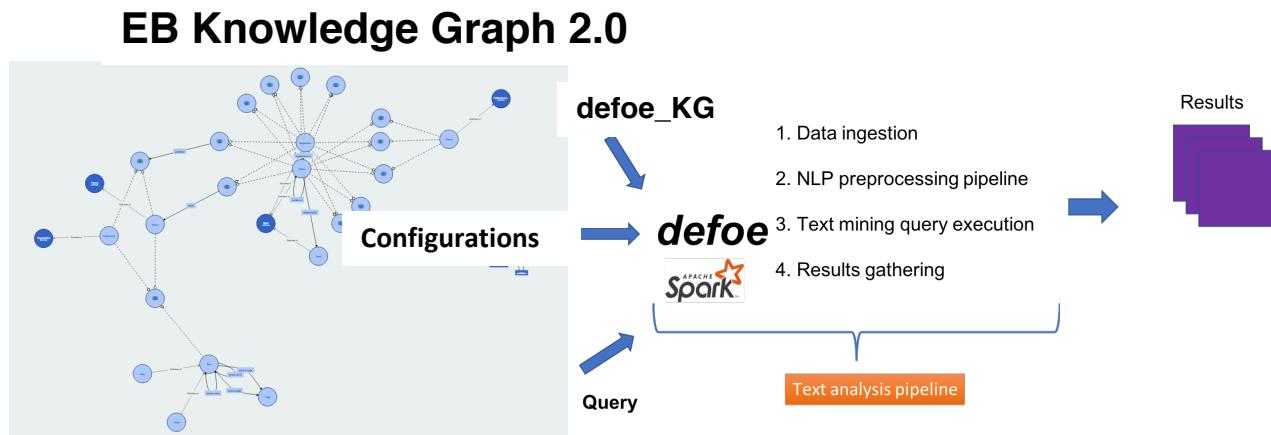


## Phase 4: Defoe and Knowledge Graph

#### 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)



## 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 from them → Abstraction to SPARQL & DEFOE
- The web-app run both types of queries and visualize 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**:
  1. Final set of **Terms & Metadata (v.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:
  1. **Type 1: SPARQL queries:** Extracting the EB KG data
  2. **Type 2: defoe queries:** Processing the EB KG data
9. Flask-web app: **Abstractions to SPARQL** and **defoe queries**

# Interacting with the EB Knowledge Graph

Term Search   Term Similarity   Topic Modelling   Spell Checker   Term Evolution   Defoe Queries   EB Details   Visualization of Resources

## 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 **EDINBURGH**.

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** if will take you to the **Topic Modelling page**, listing all the terms belonging to that particular topic model.

displaying 1 - 2 records in total 2

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_EDINBURGH_0">https://w3id.org/eb/i/Article/992277653804341_144133902_EDINBURGH_0</a>	1771	1	2	408	409	Article	the capital city of the kingdom of Scotland, situated W. long. 3° 0', and N. lat. 56° 0' 0". We ffiall... <a href="#">More</a>		<a href="#">15_scotland_county_edinburgh_firth</a>	LABEL_0_0.87	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a> <a href="#">Term Evolution</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0">https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0</a>	1773	1	2	414	415	Article	the capital city of the kingdom of Scotland, situated W. long. 3° 0', and N. lat. 56° 0' 0". We (hall ... <a href="#">More</a>		<a href="#">15_scotland_county_edinburgh_firth</a>	LABEL_0_0.89	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a> <a href="#">Term Evolution</a>

FLASK Web-Application

# Interacting with the EB Knowledge Graph

## Type 1

Term Search

Term Similarity

Topic Modelling

Spell Checker

Term Evolution

## Type 2

Defoe Queries

## Type 1

EB Details

Visualization of Resources

## 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 **EDINBURGH**.

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 2

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_EDINBURGH_0">https://w3id.org/eb/i/Article/992277653804341_144133902_EDINBURGH_0</a>	1771	1	2	408	409	Article	the capital city of the kingdom of Scotland, situated W. long. 3° 0', and N. lat. 56° 0' 0". We ffiall... <a href="#">More</a>		<a href="#">15_scotland_county_edinburgh_firth</a>	LABEL_0_0.87	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a> <a href="#">Term Evolution</a>
<a href="https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0">https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0</a>	1773	1	2	414	415	Article	the capital city of the kingdom of Scotland, situated W. long. 3° 0', and N. lat. 56° 0' 0". We (hall ... <a href="#">More</a>		<a href="#">15_scotland_county_edinburgh_firth</a>	LABEL_0_0.89	<a href="#">Spell Checker</a> <a href="#">Term Similarity</a> <a href="#">Term Evolution</a>

### Note: Two types of EB-KG queries

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

# frances: Architecture

## NSL Data Foundry Collections

