

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 diferent 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



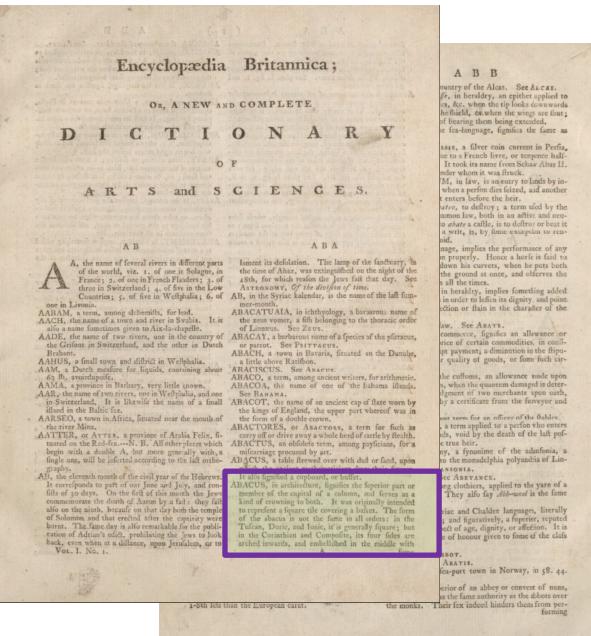
New suite of ML functionalities that can be used to analyse
any other Data Foundry collection

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

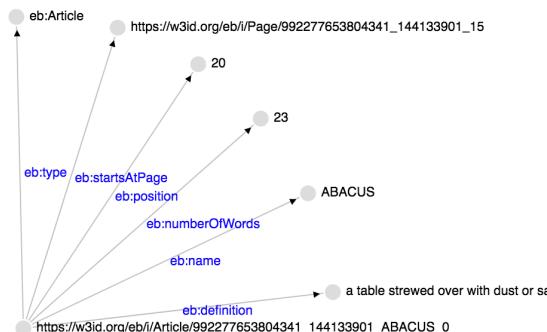
Edition 1, 1771

TERM ABACUS



Cleaned Definition

in architecture, signifies the superior part or member of the capital of a column¹, and serves as a kind of crown covering a basket². The form of the abacus is not the same in all orders: in the Tuscan, Doric, and Ionic, it is of four sides are arched in Awards³, and embellished in the middle with ornamentation, as a rose or other flower⁴; in the Tuscan pedestal⁵; and Palladio calls the plinth above the echinus, or boutin, in the Tufene and Doric orders.



Year	Edition	Volume	Start Page	End Page	Term Type	Definition/Summary	Related Terms
1771	1	1	15	15	Article	a table strewn 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 vado... More	
1	1	1	16	16	Article	logilicus, a right-angled triangle, whose sides forming the right angle contain the numbers from 1 ... More	



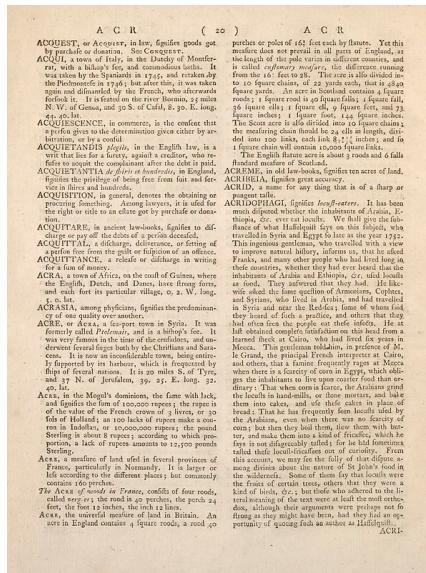
My Journey

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



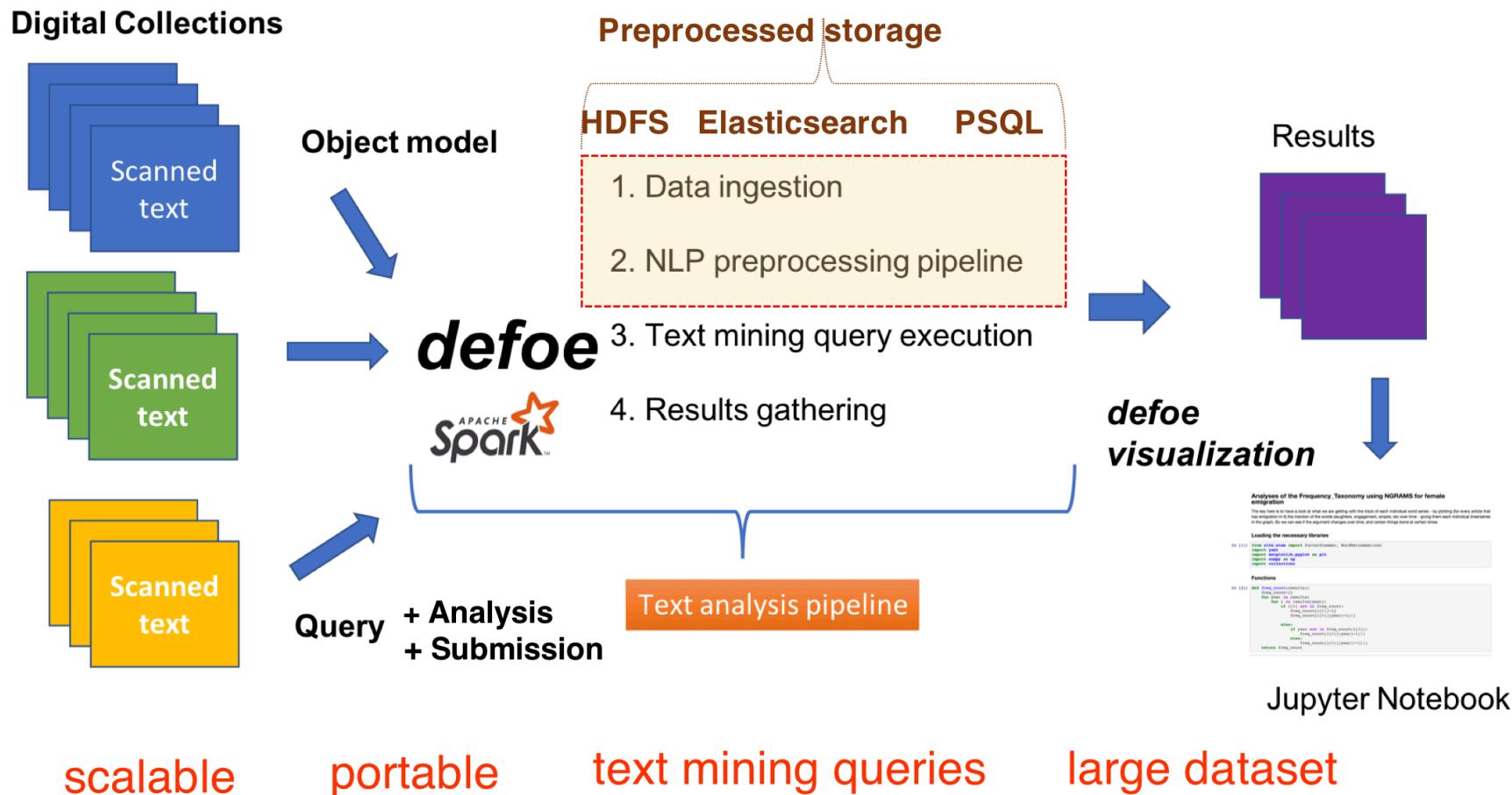
Page 36 of the First Edition
of the Encyclopaedia Britannica

```
    alt
    alt="http://www.loc.gov/standards/alto/v3/alto.xsd">
    <alto>
        <alto:Processing>
            <measurementUnit:pixel></MeasurementUnit>
            <sourceImageInformation>
                <fileName>/data/dfs/c_188936619/l_144133901/188082693.23.pdf</fileName>
            </sourceImageInformation>
            <ocrProcessingInformation>
                <ocrProcessingStep>
                    <processingDateTime>Fri Jul 12 10:36:14 2019</processingDateTime>
                </ocrProcessingStep>
            </ocrProcessingInformation>
        </alto:Processing>
        <description>
            <styling>
                <textStyle ID="font1" FONTCOLOR="#000000" FONTPROPS="serif" FONTMWIDTH="proportional" FONTHHEIGHT="10.000" FONTFAMILY="proportional" FONTCODE="WMMWWMM" FONTSTYLE="normal" />
                <textStyle ID="font1_1" FONTCOLOR="#000000" FONTPROPS="serif" FONTMWIDTH="proportional" FONTHHEIGHT="6.580" FONTFAMILY="proportional" FONTCODE="WMMWWMM" FONTSTYLE="italic" />
            </styling>
        </description>
        <layout>
            <page ID="Page36" PHYSICAL_INCH_NR="36" WIDTH="2466" HEIGHT="3327">
                <printSpace>
                    <TextLine WIDTH="232" HEIGHT="75" ID="p36_11" HPOS="521" VPOS="108">
                        <String ID="p36_11" CONTENT="<P>196. 11° HPOS=521 VPOS=108 WIDTH=42 HEIGHT=75 STYLEREFS="" font1"/>
                        <SP MWIDTH="57" VPPOS="188" HPOS="554"/>
                        <String ID="p36_12" CONTENT="C" HPOS="611" VPOS="108" WIDTH="37" HEIGHT="75" STYLEREFS="" font1"/>
                        <SP MWIDTH="50" VPPOS="188" HPOS="648"/>
                        <String ID="p36_13" CONTENT="11° HPOS=706 VPOS=108" WIDTH="38" HEIGHT="75" STYLEREFS="" font1"/>
                    </TextLine>
                    <TextLine WIDTH="132" HEIGHT="75" ID="p36_12" HPOS="119" VPOS="108">
                        <String ID="p36_12" CONTENT="<P>196. 12° HPOS=119 VPOS=108 WIDTH=19 HEIGHT=75 STYLEREFS="" font1"/>
                        <SP MWIDTH="57" VPPOS="188" HPOS="138"/>
                        <String ID="p36_13" CONTENT="11° HPOS=1200 VPOS=108" WIDTH="51" HEIGHT="75" STYLEREFS="" font1"/>
                    </TextLine>
                    <TextLine WIDTH="243" HEIGHT="75" ID="p36_13" HPOS="171" VPOS="182">
                        <String ID="p36_13" CONTENT="<P>196. 13° HPOS=171 VPOS=182 WIDTH=243 HEIGHT=75 STYLEREFS="" font1"/>
                        <SP MWIDTH="24" VPPOS="182" HPOS="414"/>
                        <String ID="p36_w7" CONTENT="r" HPOS="438" VPOS="182" WIDTH="39" HEIGHT="75" STYLEREFS="" font1"/>
                        <SP MWIDTH="15" VPPOS="182" HPOS="477"/>
                        <String ID="p36_w8" CONTENT="<P>196. 14° HPOS=477 VPOS=182 WIDTH=193 HEIGHT=75 STYLEREFS="" font1"/>
                        <SP MWIDTH="24" VPPOS="182" HPOS="585"/>
                        <String ID="p36_w9" CONTENT="<br>" HPOS="199" VPOS="182" WIDTH="124" HEIGHT="75" STYLEREFS="" font1"/>
                        <String ID="p36_w10" CONTENT="Guifit" HPOS="833" VPOS="182" WIDTH="134" HEIGHT="75" STYLEREFS="" font1"/>
                        <SP MWIDTH="24" VPPOS="182" HPOS="941"/>
                        <String ID="p36_w11" CONTENT="goods" HPOS="1015" VPOS="182" WIDTH="102" HEIGHT="75" STYLEREFS="" font1"/>
                        <SP MWIDTH="24" VPPOS="182" HPOS="1117"/>
                    </TextLine>
                    <TextLine WIDTH="792" HEIGHT="75" ID="p36_14" HPOS="221" VPOS="223">
                        <String ID="p36_14" CONTENT="got" HPOS="1141" VPOS="223" WIDTH="59" HEIGHT="75" STYLEREFS="" font1"/>
                    </TextLine>
                </printSpace>
            </page>
        </layout>
    </alto>
```

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>

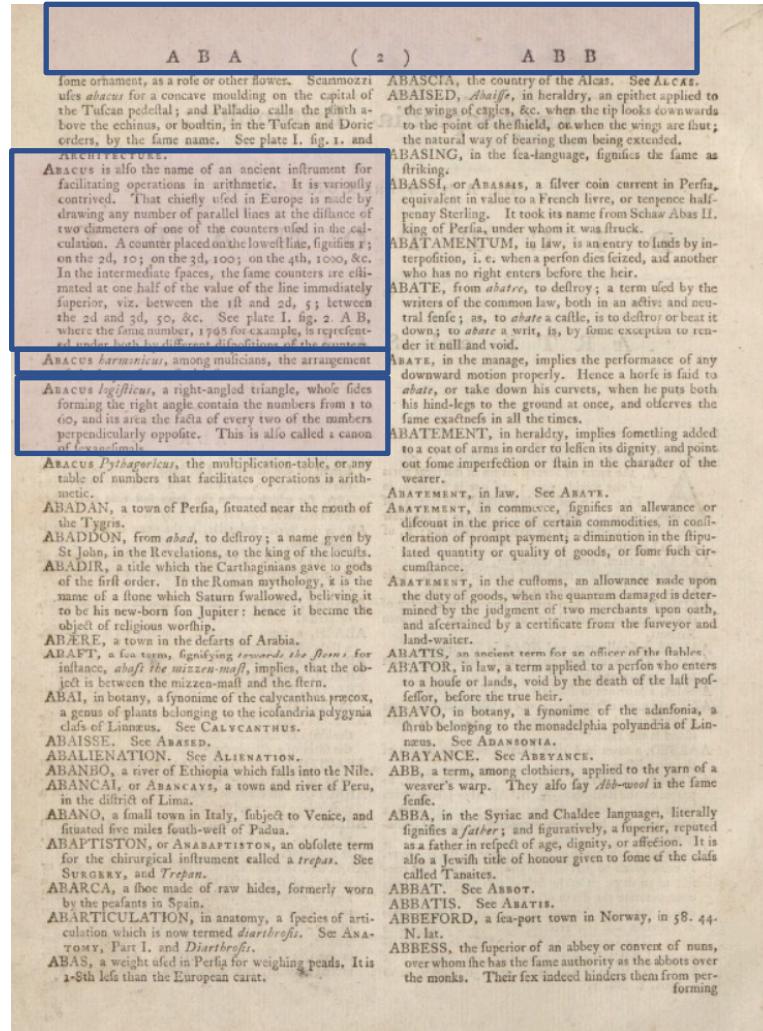
Phase 1: Information Extraction

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.

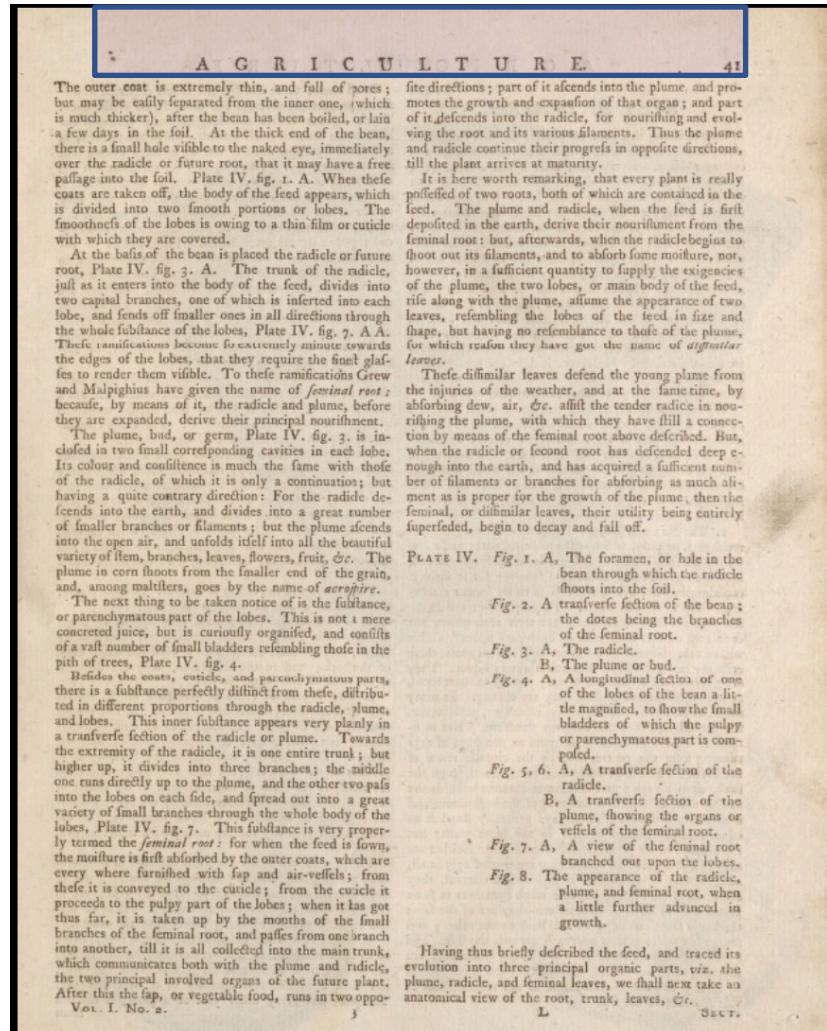
Phase 1: Information Extraction

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

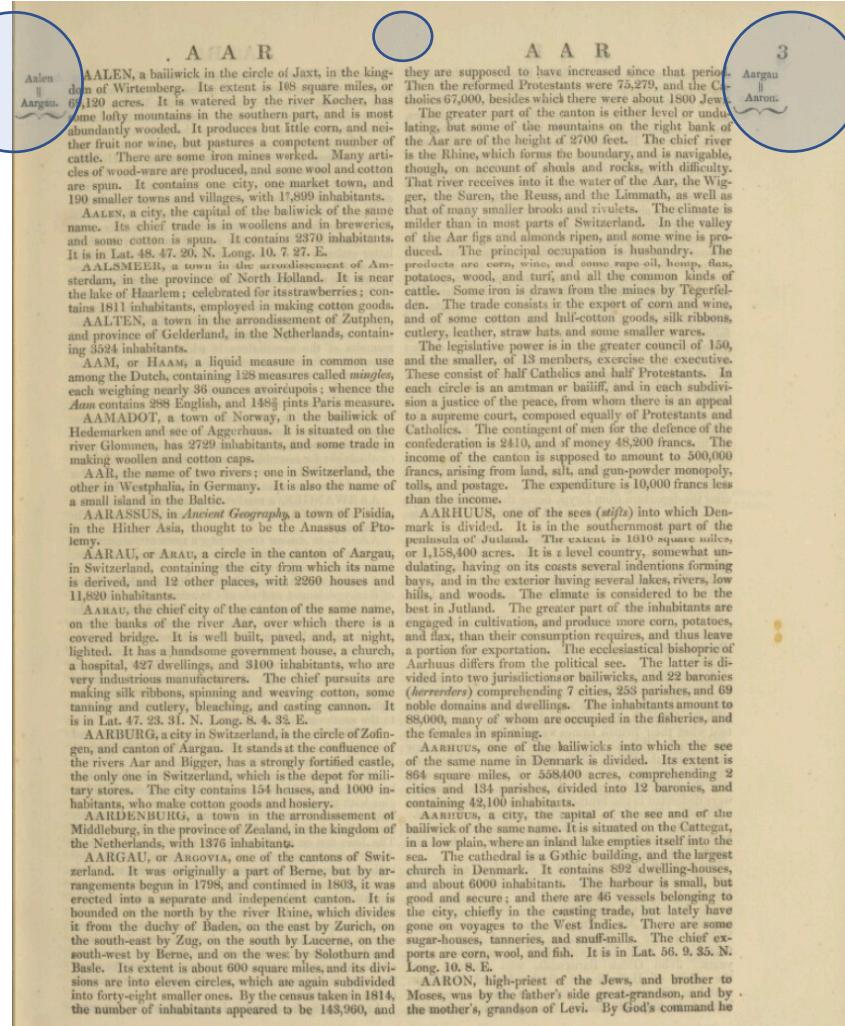
Edition 1 - 1771



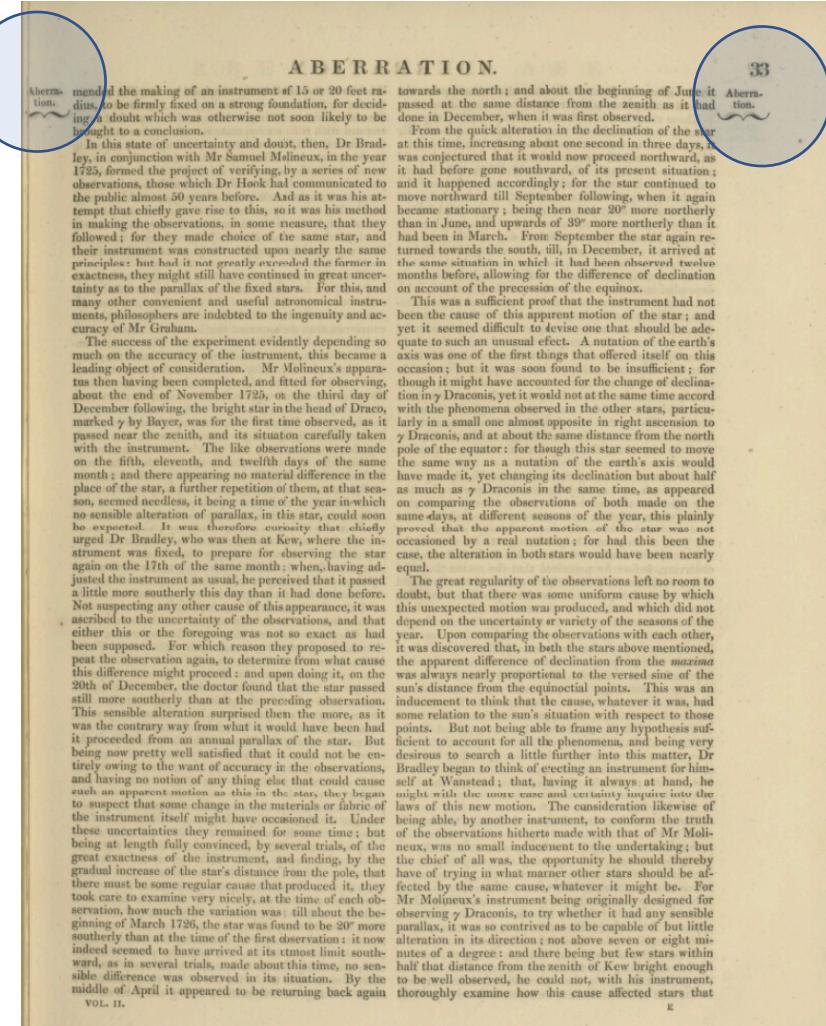
Topic

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 + “,”

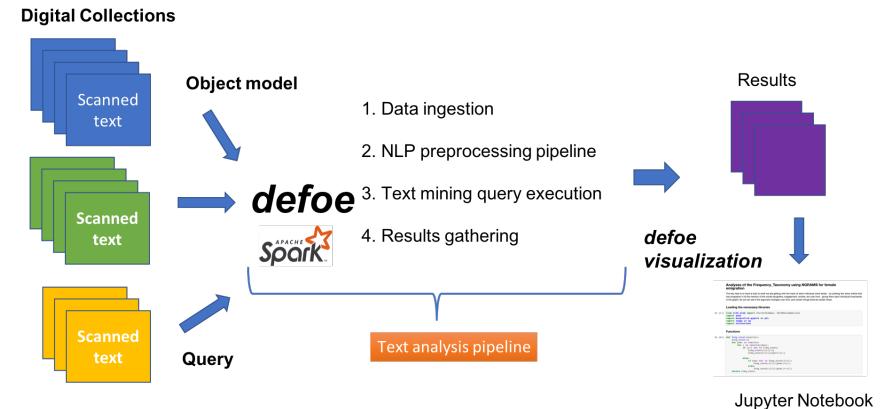


Articles
Edition 7



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

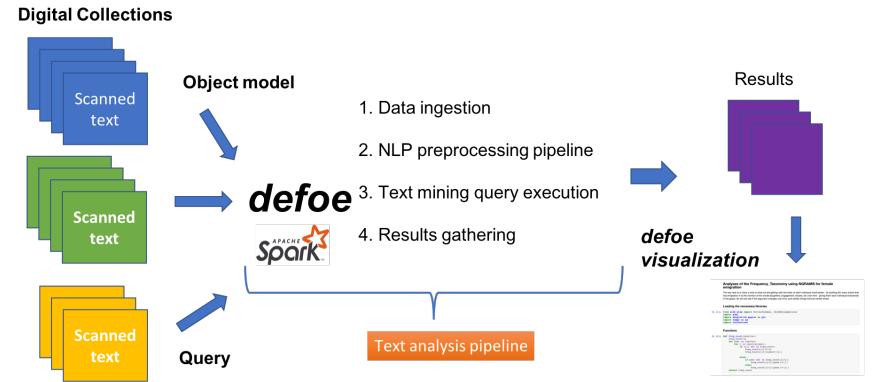
Extracted Term ABACATORES information - Edition 1, 1771, Vol A-B

Phase 1: 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**

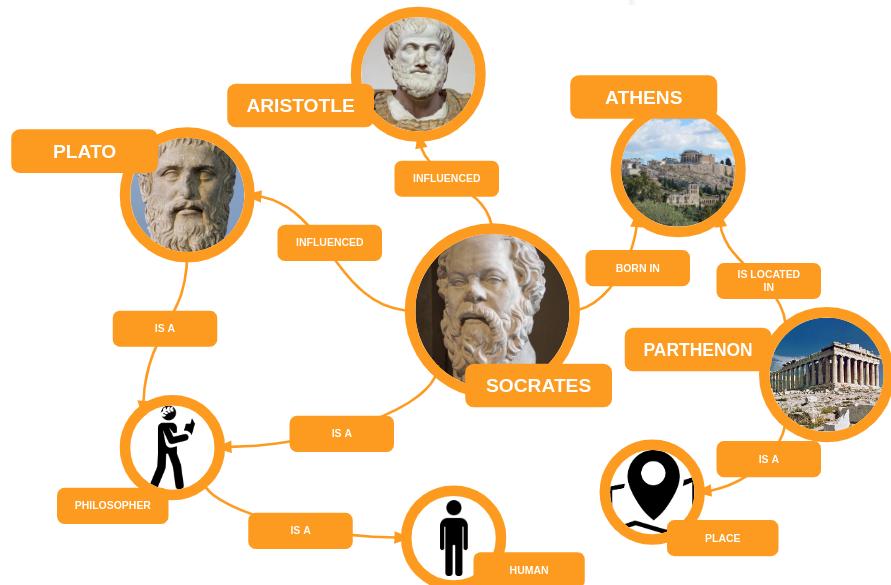
(*) **Also Based on Heuristics** --> Pages Layout & Text & Headers → different heuristics for different EB editions.

Phase 2: Knowledge Graph

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

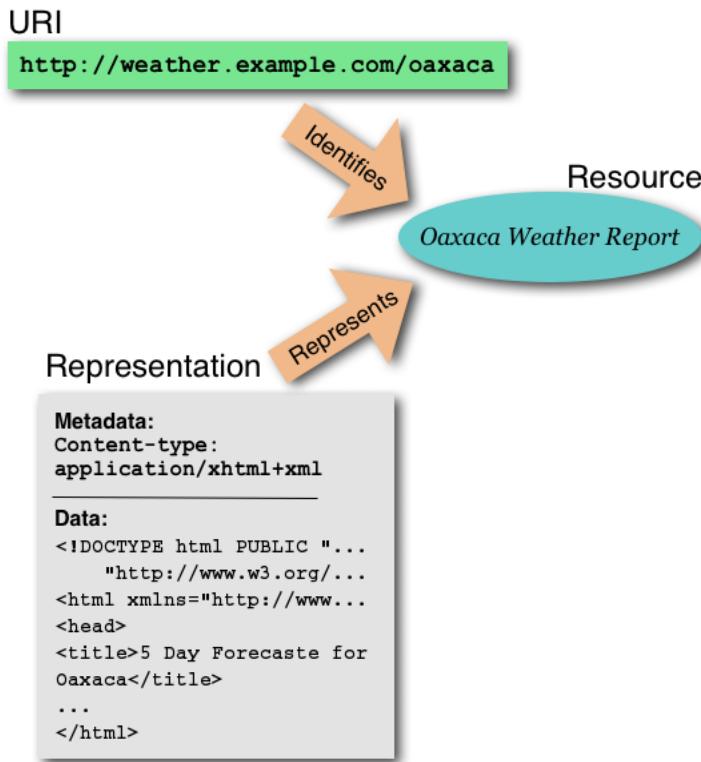
- A **knowledge graph** is a specialized graph or network of the things we want to describe and how they are related
- It is a **semantic** model since we want to capture and generate **meaning** with the model

"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

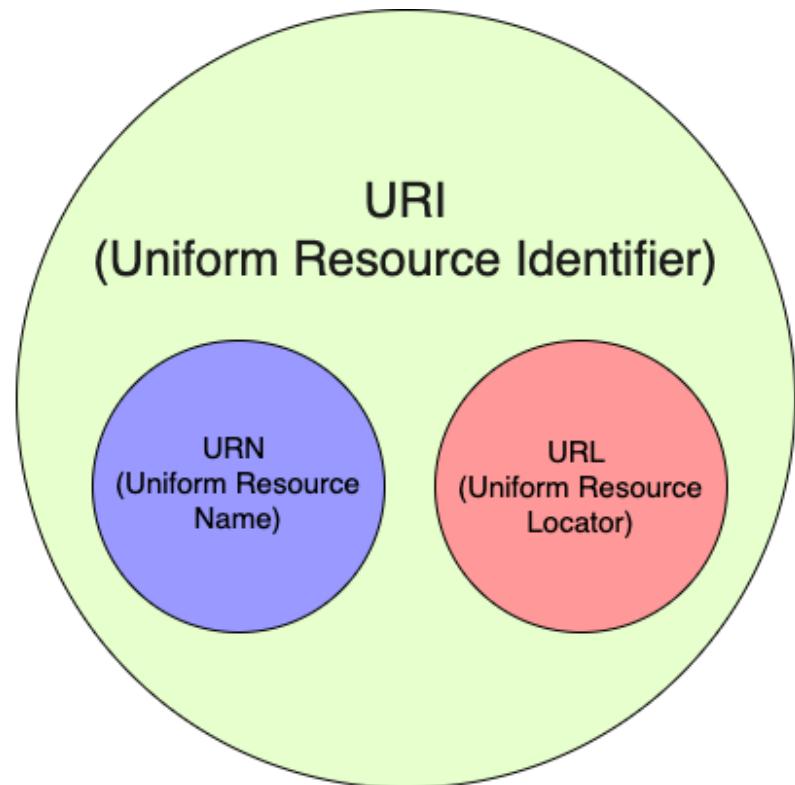


Phase 2: Knowledge Graph

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

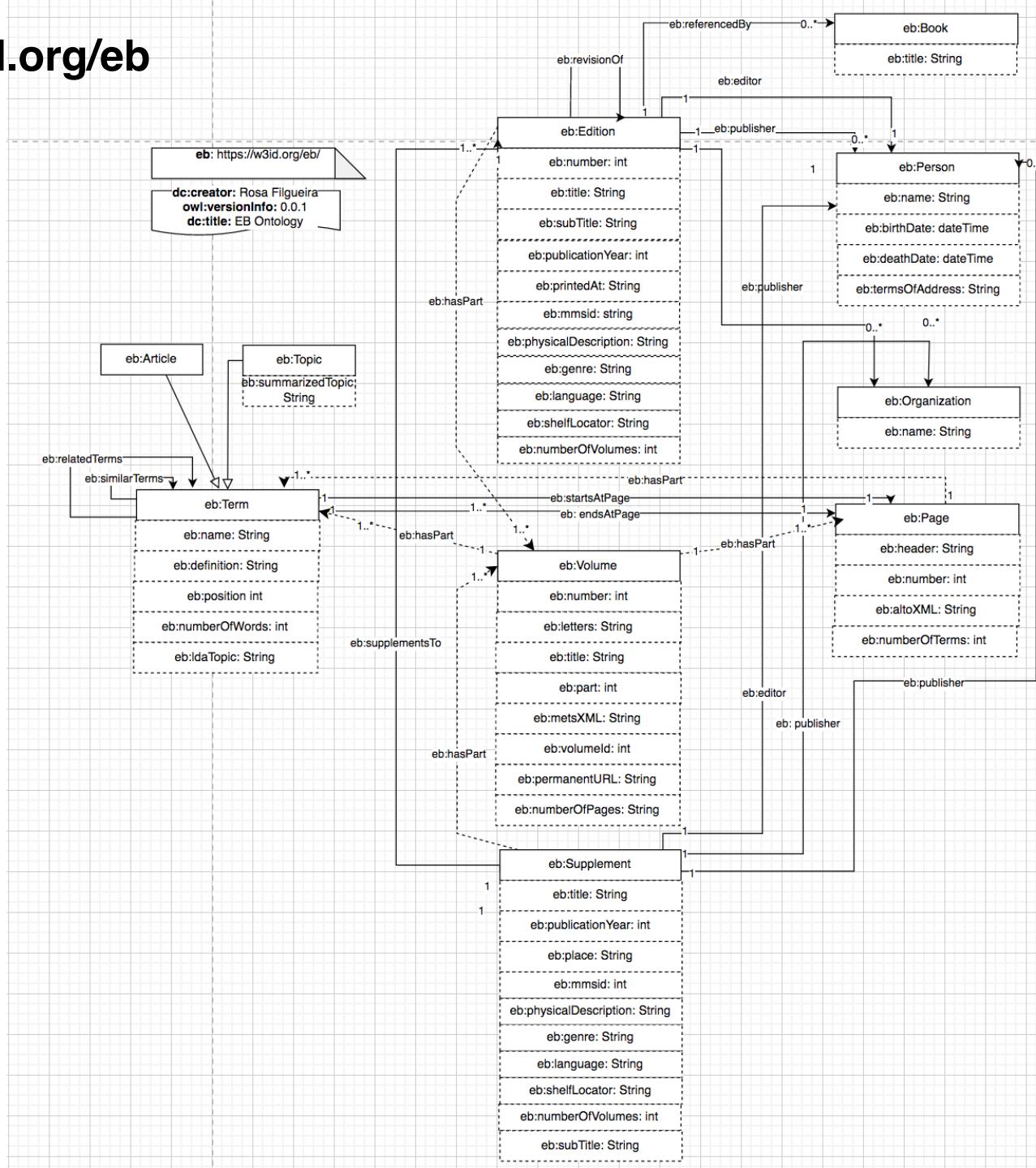
2.1. Create EB Ontology: To explain (give meaning) how our “things” are related with each other.

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

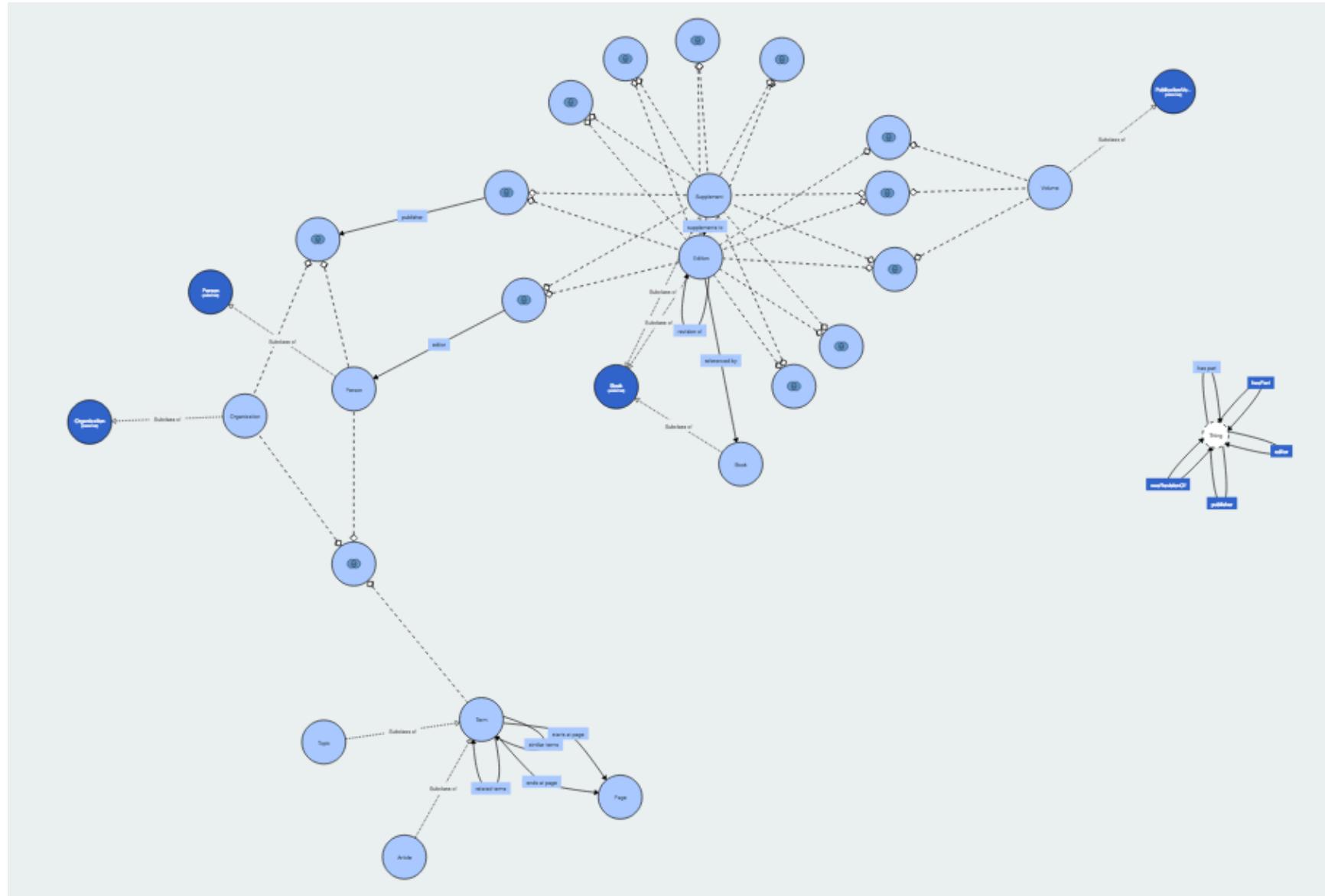
2. 2 Created **EB Knowledge Graph 1.0**: Populated the post-processed information (extracted **Terms & Metadata v.2**) into an RDF triplestore:

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

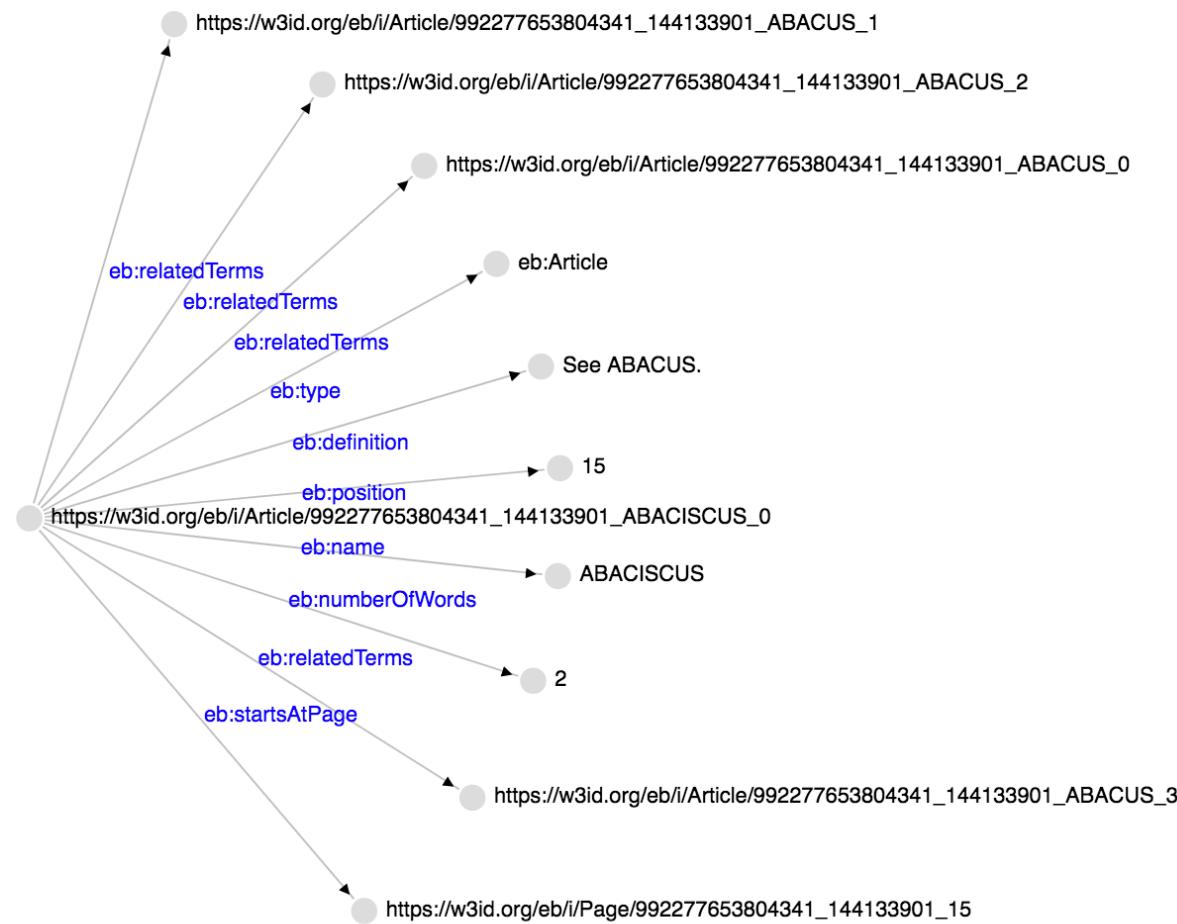
NOTE:

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

EB Knowledge Graph

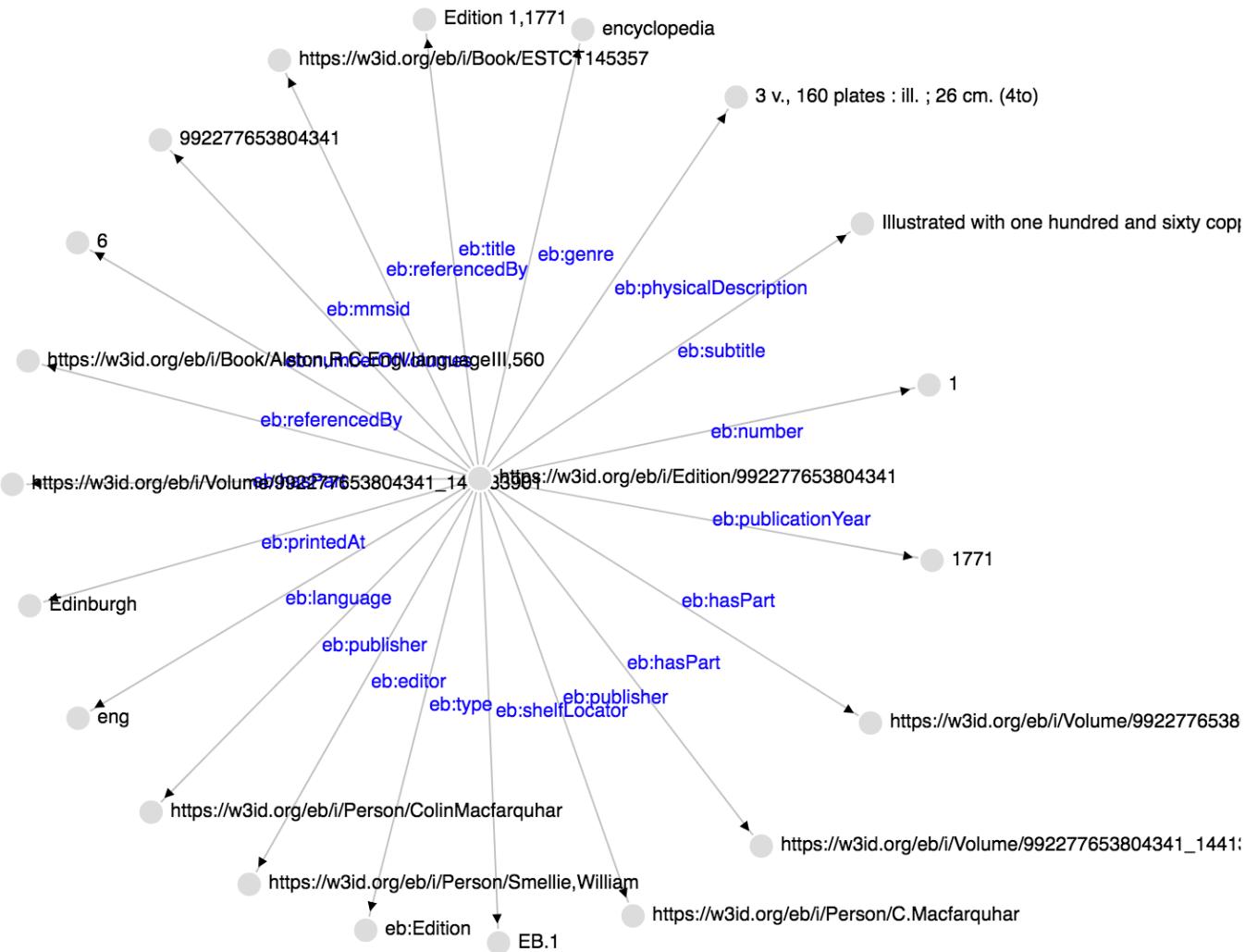


EB Knowledge Graph



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
```

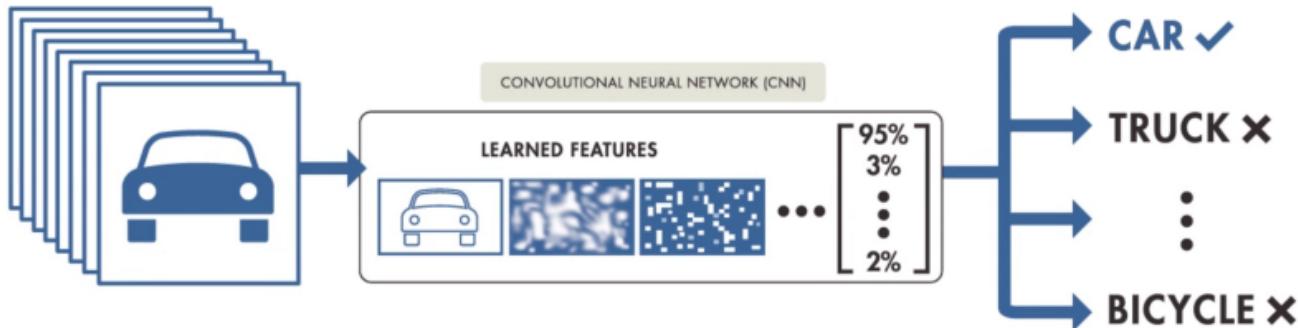


Fuseki RDF triplestore

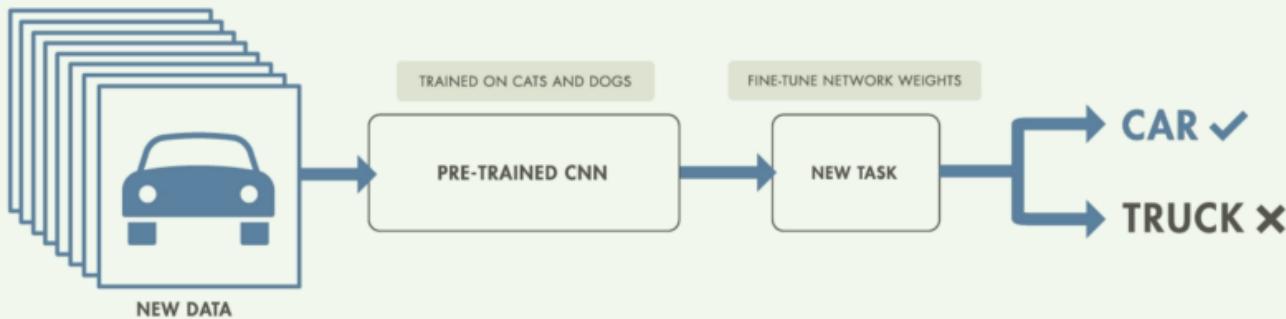
QUERY RESULTS	
Table	Raw Response
Download	
Showing 1 to 8 of 8 entries	
	<input type="text"/> Search: <input type="button"/> Show 50 entries
article	definition
1 https://w3id.org/eb/l/Article/9929192893804340_144850368_PAISLEY_0	"a town of Scotland, in the county of Renfrew, six miles west of Glasgow."
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

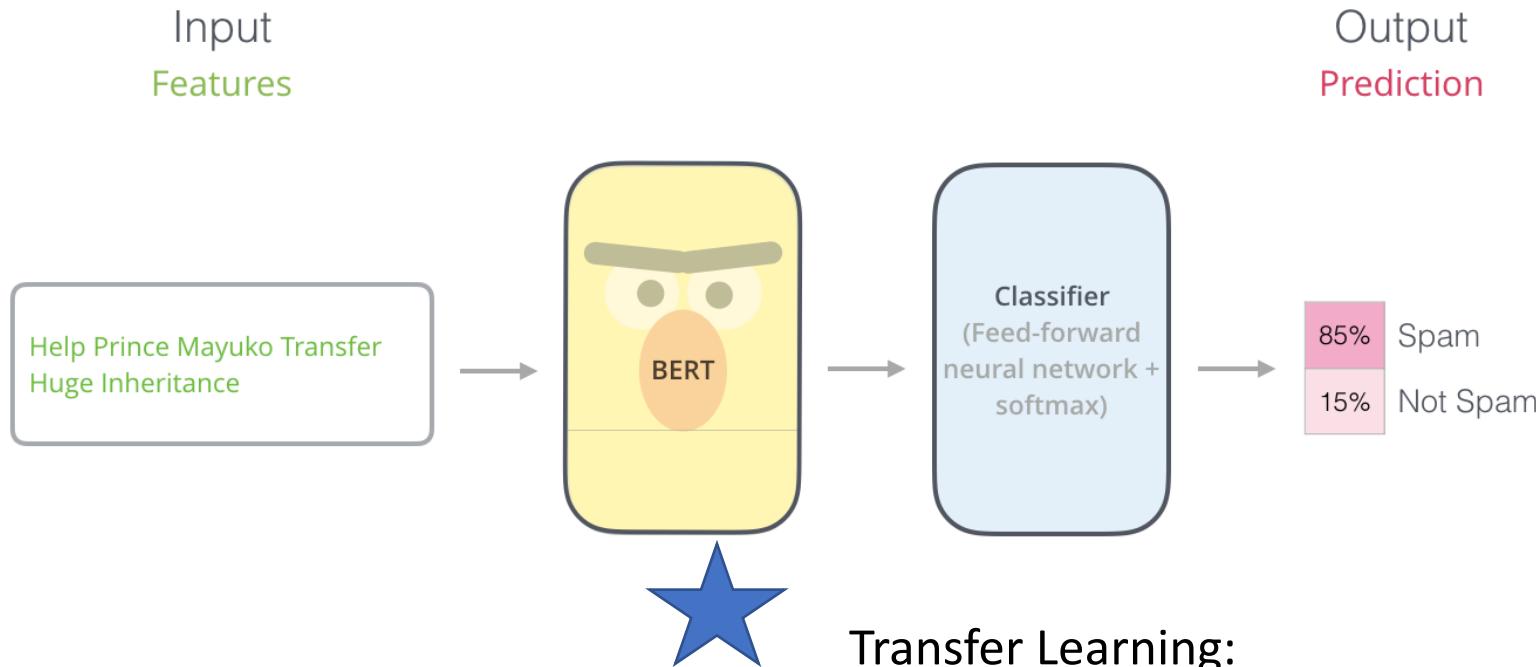
A High-Level Look

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-mnlp-base-v2*](#)
- **term similarity:** Comparing text & semantic similarity
 - transformer: [*all-mnlp-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-mnlp-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 .



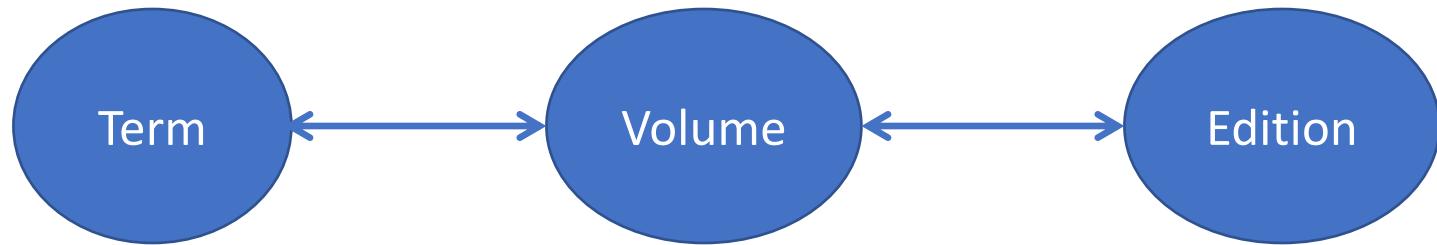
Compute Difference

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.**

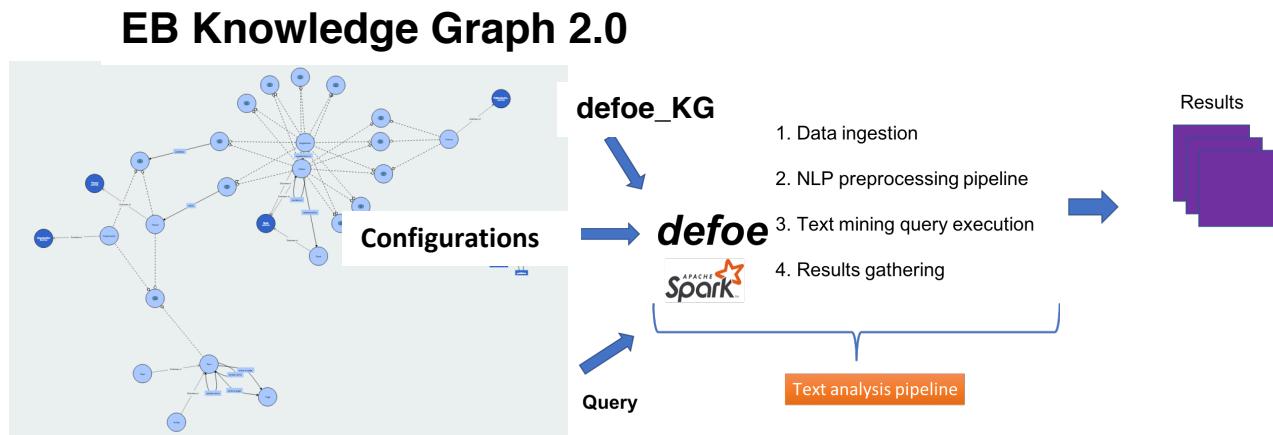


Example: Given a **Term** (e.g., *Edinburgh*), I can get the **Edition** Information (e.g., *Edition Title*)

- **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

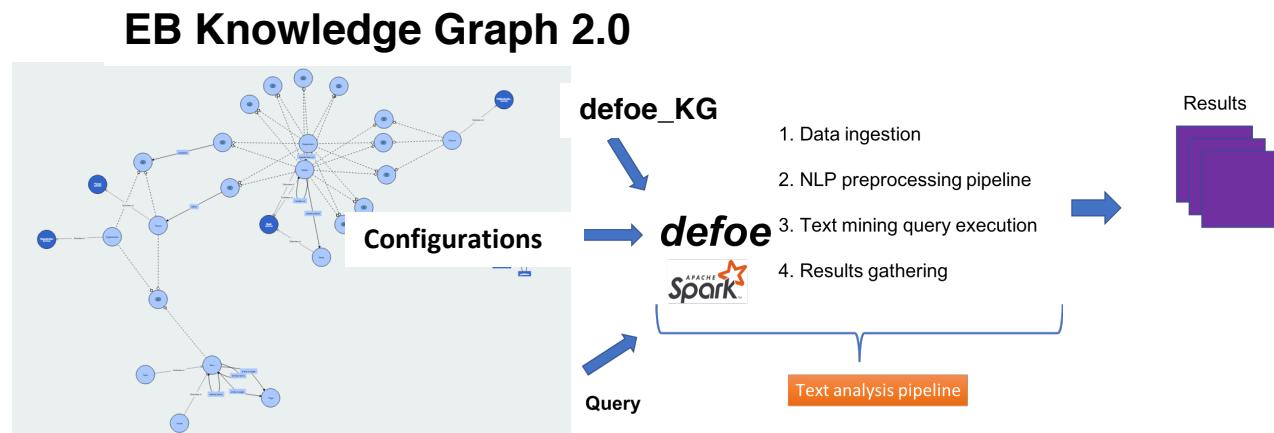
- 4.1. Created a new **KG defoe connector** (based in SPARQL) to run defoe queries using the EB Knowledge Graph as a source of data → **defoe_KG**
- 4.2. Improved **defoe queries to be fully configurable**: different filtering options, target, lexicon, etc.



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.



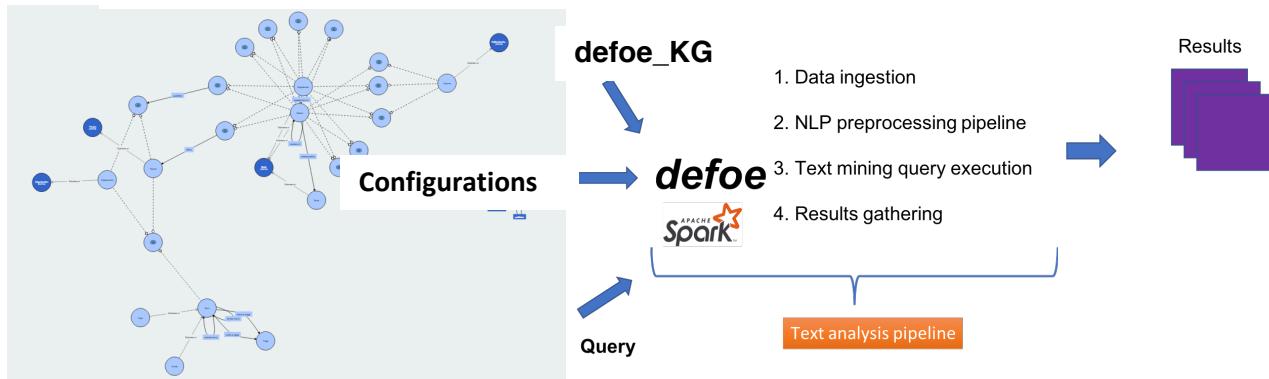
Phase 4: Defoe and Knowledge Graph

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 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
https://w3id.org/eb/i/Article/992277653804341_144133902_EDINBURGH_0	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... More		15_scotland_county_edinburgh_firth	LABEL_0_0.87	Spell Checker Term Similarity Term Evolution
https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0	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 ... More		15_scotland_county_edinburgh_firth	LABEL_0_0.89	Spell Checker Term Similarity Term Evolution

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
https://w3id.org/eb/i/Article/992277653804341_144133902_EDINBURGH_0	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... More		15_scotland_county_edinburgh_firth	LABEL_0_0.87	Spell Checker Term Similarity Term Evolution
https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0	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 ... More		15_scotland_county_edinburgh_firth	LABEL_0_0.89	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**

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_144133902_EDINBURGH_0	1	1771	2	EDINBURGH	the capital city of the kingdom of Scotland, situated W. long. 3°, and N. lat. 56°00'00". We ffiall... More	15_scotland_county_edinburgh_firth	LABEL_0_0.87

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	Similarity Rank	Sentiment_Score
https://w3id.org/eb/i/Article/9929192893804340_144850367_EDINBURGH_0	1	1773	2	EDINBURGH	the capital city of the kingdom of Scotland, situated W. long. 3°, and N. lat. 56°00'00". We (hall ... More	15_scotland_county_edinburgh_firth	0.99541134	LABEL_0_0.89
https://w3id.org/eb/i/Article/992277653804341_144133901_ABERDEEN_0	1	1771	1	ABERDEEN	the name of two cities in Scotland, called the Old and New Towns, situated on the German Ocean, in i... More	15_scotland_county_edinburgh_firth	0.7731124	LABEL_0_0.92
https://w3id.org/eb/i/Article/9929192893804340_144850367_DUMFERMLINE_0	1	1773	2	DUMFERMLINE	a parliament-town of Scotland, situated in the county of Fife	15_scotland_county_edinburgh_firth	0.7610539	POSITIVE_1.00

Interacting with the EB Knowledge Graph

Topic Modelling

Exploring the Encyclopaedia Britannica (1768-1860)

Topic Modelling Search

Enter a **topic modelling name** or just the **number** of a topic to see all the terms within the same topic. All topics modelling names start with a number. If no topic is introduced, it will use the first topic modelling, which name starts with '0'.

We have 127 topics modelling items.

182 terms found for the topic 15_scotland_county_edinburgh_frith

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 182

« 1 2 3 4 5 ... 18 19 »

URI	Edition	Year	Volume	Term	Definition	Sentiment_Score
https://w3id.org/eb/i/Article/9929192893804340_144850368_PAISLEY_0	1	1773	3	PAISLEY	a town of Scotland, in the county of Renfrew, six miles west of Glasgow.	POSITIVE_0.99
https://w3id.org/eb/i/Article/9929192893804340_144850366_ARGYLESHERE_0	1	1773	1	ARGYLESHERE	a county of Scotland, lying westward of Glasgow, and comprehending the counties of Lorn, Cowal, Kna... More	POSITIVE_1.00
https://w3id.org/eb/i/Article/9929192893804340_144850367_LEWIS_0	1	1773	2	LEWIS	the most northerly of any of the western islands of Scotland, lying in 8°W00'00 odd minutes W. long.... More	POSITIVE_0.99

Topic Modelling Visualizations

We are visualizing the topic model 15_scotland_county_edinburgh_frith. For each topic model, we plot their most common 10 words , along with their c-TF-IDF scores .

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.

URI	Edition	Year	Volume	Term
https://w3id.org/eb/i/Article/9929192893804340_144850368_WIDOW_0	1	1773	3	WIDOW

Original Definition

a woman who has lost her hufbind.

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

a woman who has lost her husband .



Original Definition

a woman who has lost her hufbind.

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

a woman who has lost her huffbisband .

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_144133902_EDINBURGH_0	1	1771	2	EDINBURGH	the capital city of the kingdom of Scotland, situated W. long. 3°, and N. lat. 56°. We ffiall... More	15_scotland_county_edinburgh_frith	LABEL_0_0.87

The evolution of EDINBURGH 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_144850367_EDINBURGH_0	1	1773	2	EDINBURGH	the capital city of the kingdom of Scotland, situated W. long. 3°, and N. lat. 56°. We ffiall... More	15_scotland_county_edinburgh_frith	0.99541134	LABEL_0_0.89

Select Edition

Select Volume

Edition 1 Year 1771

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

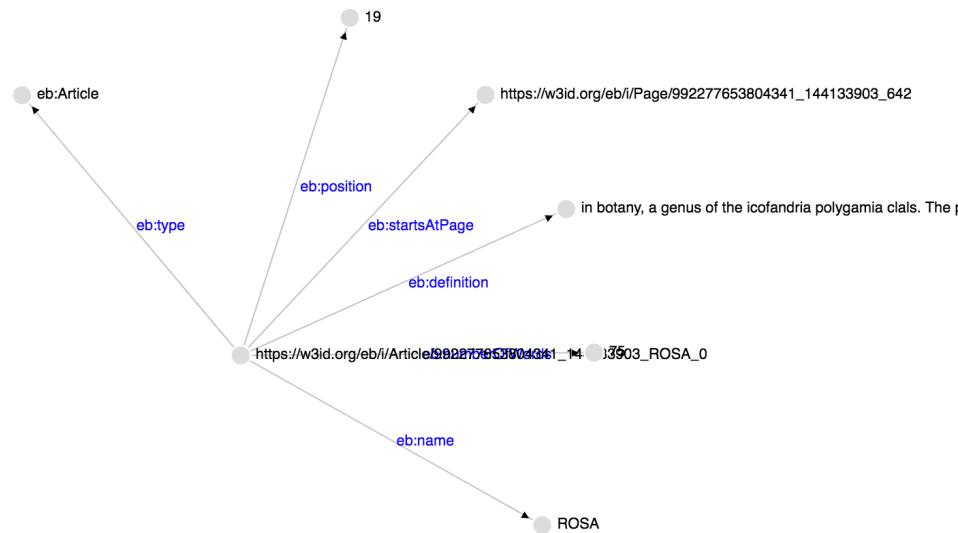
Volume	Details
Volume Number	1
Volume URI	< https://w3id.org/eb/i/Volume/992277653804341_144133901 >
Volume Title	Encyclopaedia Britannica; or, A dictionary of arts and sciences, compiled upon a new plan
Volume Letters	A-B
Volume Permanent URL	https://digital.nls.uk/144133901
Volume Number of Pages	832

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_144133903_ROSA_0>



Interacting with the EB Knowledge Graph

Defoe Queries

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

Defoe query frequency_keysearch_by_year ▾

Preprocess Treatment

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

Select the hit count (only for Freq. query) term ▾

Upload your lexicon file

peace_war.txt

Filtering Options

Introduce Target words/sentences separated by "," (Optional) peace accord,peace agr

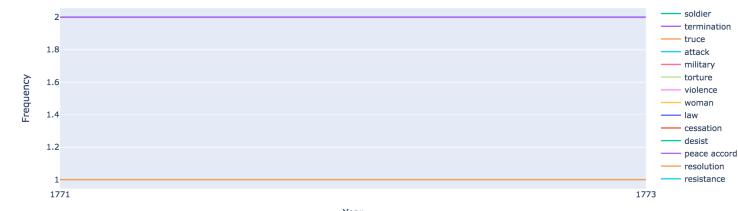
Select the Target filter

- Select terms which contains any of the target words/sentences
- Select terms which contains all the target words/sentences

Start Year (Optional) 1771

End Year (Optional) 1773

Frequency of Lexicon Terms per Years



Normalized Frequency of Lexicon Terms per Years



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

frances: Architecture

NSL Data Foundry Collections

