

Scalable Pipelines for Lexical & Semantic Text Reuses

Workshop on Text Reuse in the Eighteenth Century

Ananth Mahadevan

Computer Science Department
University of Helsinki

2nd May 2024

Contents

1 Lexical Text Reuses

- Identifying Lexical Reuses
- Extract Transform and Load
- Locke Quotes Task

2 Semantic Text Reuses

- Semantic Pipeline
- Locke Quotes Task

3 Discussion Questions

Lexical Text Reuses

Lexical Pipeline

- ① Identify lexical text reuses
 - BLAST
- ② Extract-Transform-Load (ETL)
 - De-fragment
 - Cluster
 - Identify
- ③ Locke Quotes Task
 - Task-specific data
 - Database interface

Identifying Lexical Reuses

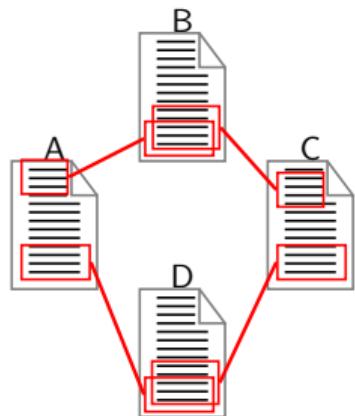
- **BLAST:** Basic Local Alignment Search Tool
 - Bioinformatics algorithm for sequence alignment
 - Encodes OCR text into protein-like sequences
 - Comparing every document against the entire corpus
 - Aligns sequences and returns offsets in document
- Raw Data
 - Collections: **ECCO**, **EEBO-TCP** & British Library **Newspapers**
 - 3.8 Billion pieces of text
 - 6.3 Billion pairs of text reuse

Extract Transform and Loading Stages

Raw BLAST reuses (6.3B)

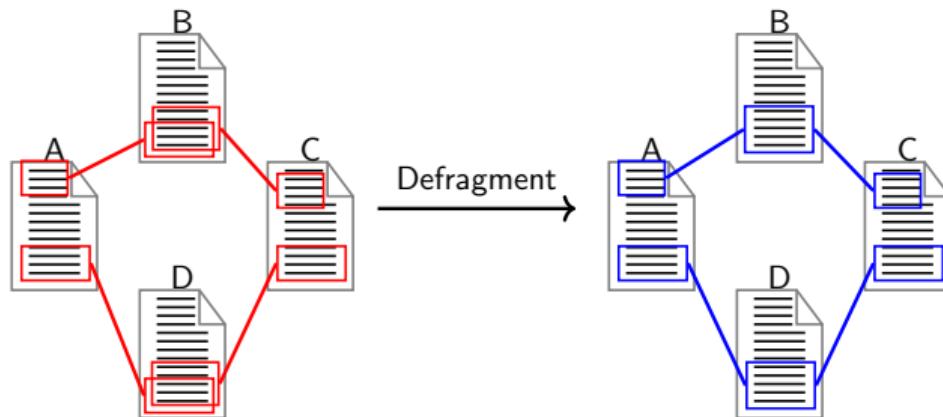
Legend:

- Piece (Red Box)
- Defrag Piece (Blue Box)
- Src Piece (Orange Box)
- Cluster (Purple Box)
- Reuse (Red Line)
- Defrag Reuse (Blue Line)
- Dest Piece (Green Box)
- Reception (Right Arrow)
- Doc (Document Icon)



Extract Transform and Loading Stages

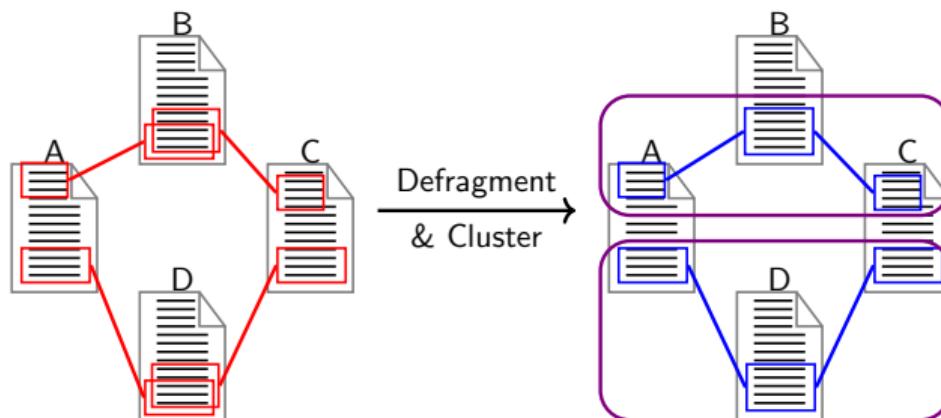
Merge pieces within a document (1.1 billion defragmented pieces)



Extract Transform and Loading Stages

Cluster pieces across documents (91 million distinct clusters)

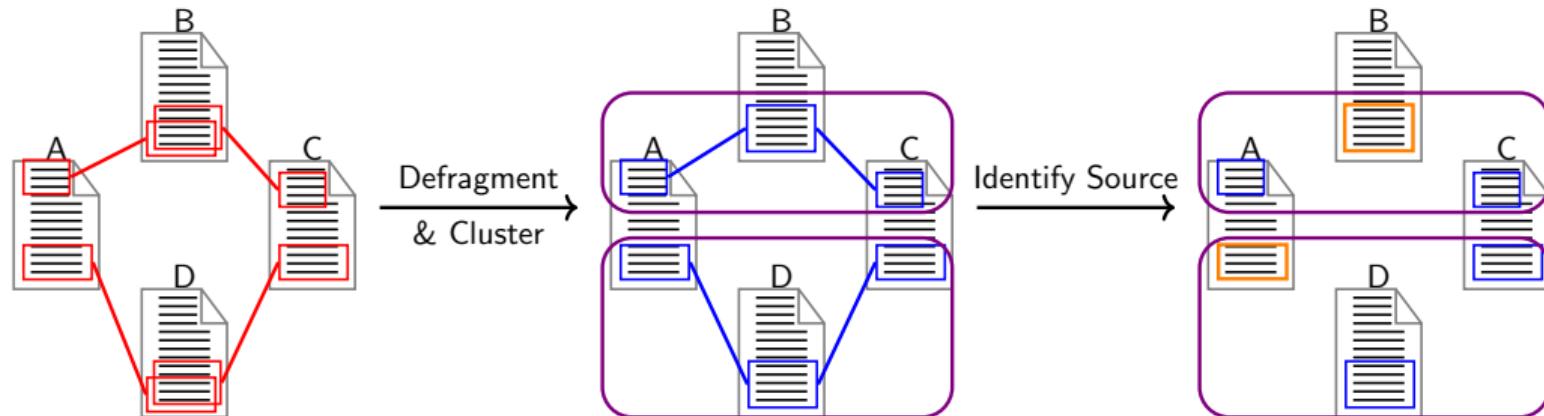
Legend:
Piece (Red Box) Defrag Piece (Blue Box) Src Piece (Orange Box) Cluster (Purple Box)
Reuse (Red Line) Defrag Reuse (Blue Line) Dest Piece (Green Box) Reception (Right Arrow) Doc (Document Icon)



Extract Transform and Loading Stages

Identify source(s) in each cluster

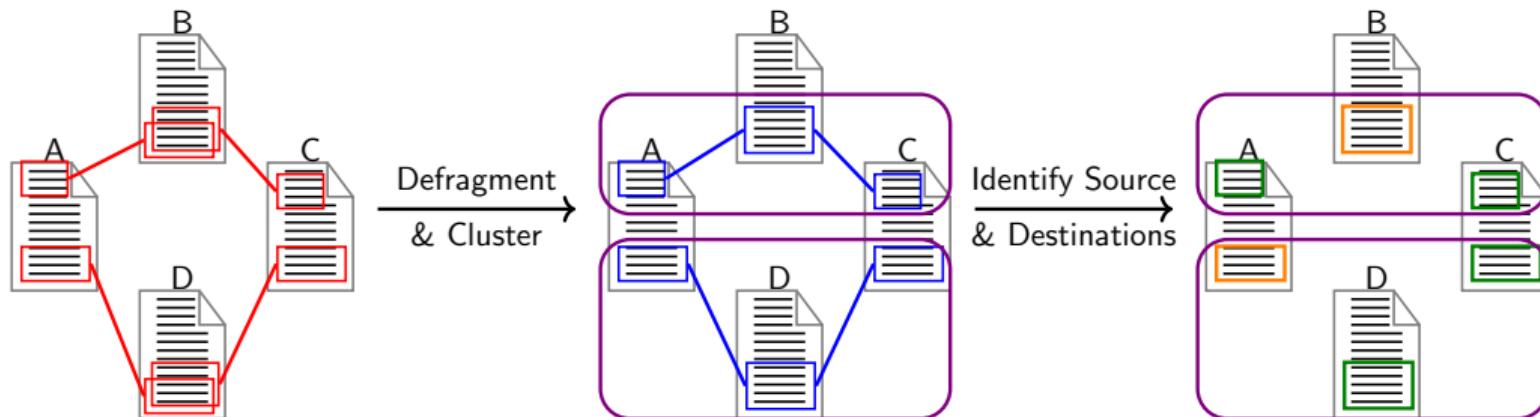
Legend:
Piece (Red Box)
Defrag Piece (Blue Box)
Src Piece (Orange Box)
Cluster (Purple Box)
Reuse (Red Line)
Defrag Reuse (Blue Line)
Dest Piece (Green Box)
Reception (Right Arrow)
Doc (File Icon)



Extract Transform and Loading Stages

Identify destinations in each cluster

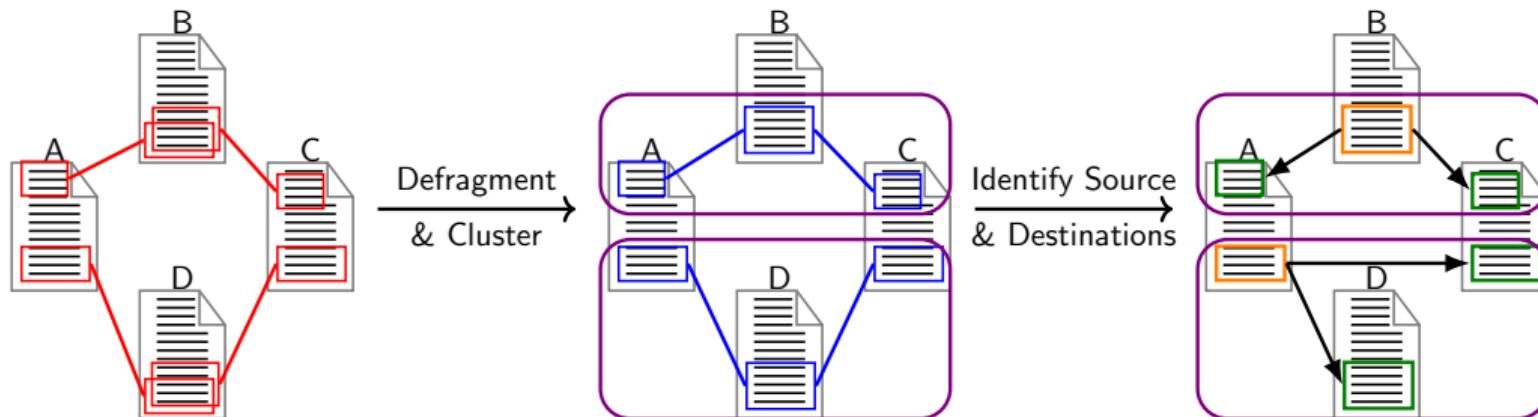
Legend:
Piece (Red Box) Defrag Piece (Blue Box) Src Piece (Orange Box) Cluster (Purple Box)
Reuse (Red Line) Defrag Reuse (Blue Line) Dest Piece (Green Box) Reception (Doc Icon)



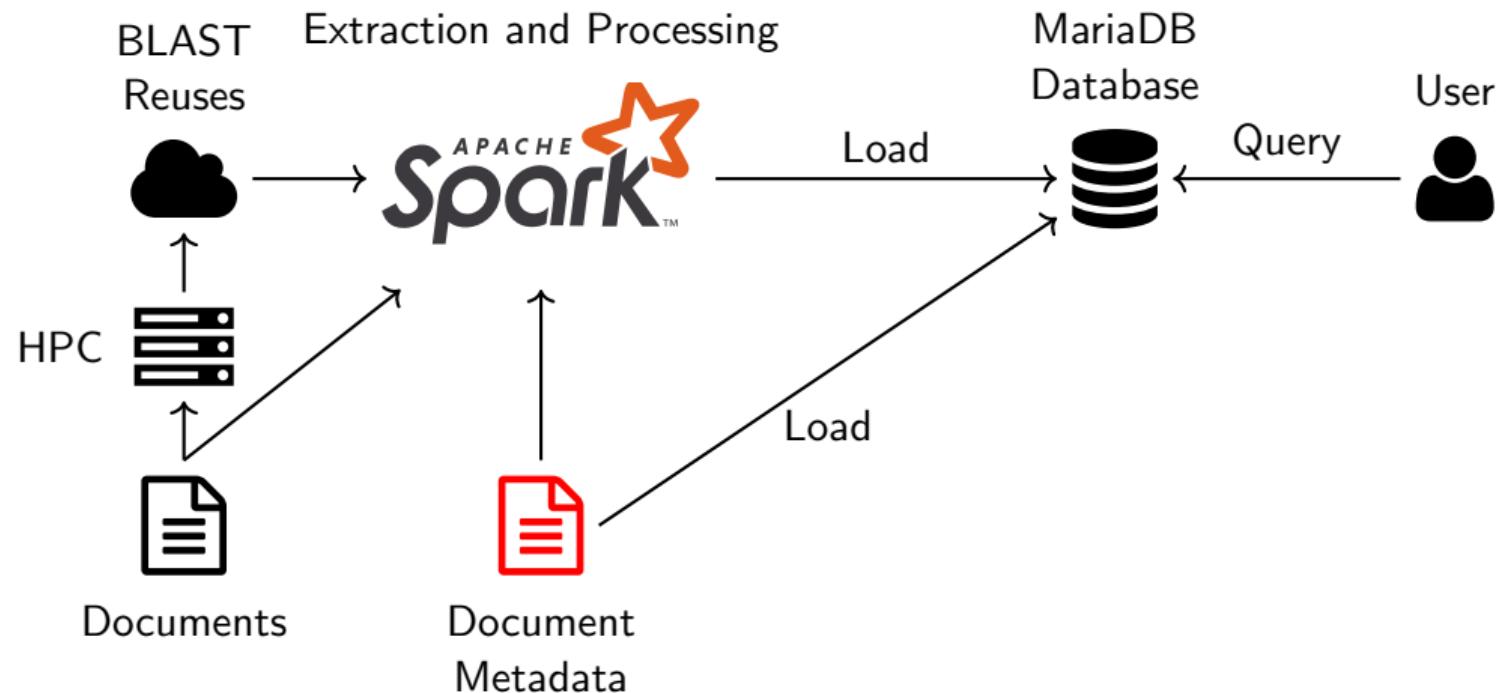
Extract Transform and Loading Stages

Reception edges between source and destinations

Legend:
Piece Defrag Piece Src Piece Cluster
Reuse Defrag Reuse Dest Piece Reception Doc



Implementation



Locke Quotes Task

Document: John Locke: An essay concerning humane understanding

Quote Length: 150 – 200 characters

Quote Metric: Number of reuses in works of different authors

Quantity: Top k quotes

Steps:

- ① Find clusters where source piece belongs to **Document** and is of **Length**
- ② In each cluster find all destination pieces, their authors and filter out John Locke
- ③ Count number of unique destination works for each source piece (**Metric**)
- ④ Return Top- k source pieces as quotes (**Quantity**)

Complex Query

```
1  WITH filtered_clusters AS (
2      SELECT cluster_id, dp.* FROM edition_ids
3      INNER JOIN textreuse_edition_mapping USING(edition_id_i)
4      INNER JOIN defrag_pieces dp USING(trs_id)
5      INNER JOIN earliest_work_and_pieces_by_cluster USING(piece_id)
6      WHERE edition_id = :estc_id AND dp.trs_end - dp.trs_start BETWEEN 150 AND 200
7 ),query_authors AS (
8     SELECT DISTINCT actor_id_i
9     FROM edition_authors
10    INNER JOIN edition_ids USING(edition_id_i)
11    WHERE edition_id = :estc_id AND actor_id_i IS NOT NULL
12 ), cluster_stats AS (
13     SELECT cluster_id,COUNT(DISTINCT work_id_i) as n_works FROM filtered_clusters
14     INNER JOIN non_source_pieces_tmp nsp USING (cluster_id)
15     INNER JOIN defrag_pieces dp ON nsp.piece_id = dp.piece_id
16     INNER JOIN textreuse_edition_mapping tem ON tem.trs_id = dp.trs_id
17     INNER JOIN textreuse_work_mapping twm ON twm.trs_id = dp.trs_id
18     INNER JOIN edition_authors USING(edition_id_i)
19     LEFT JOIN query_authors qa USING(actor_id_i)
20     WHERE qa.actor_id_i IS NULL
21     GROUP BY cluster_id
22 ),results AS (
23     SELECT *,trs_end-trs_start AS piece_length FROM cluster_stats
24     INNER JOIN filtered_clusters fc USING (cluster_id))
25 SELECT * FROM results
26 ORDER BY n_works DESC,piece_length DESC
27 LIMIT 100
```

Task Specific Data

Top Quotes Task Data

source_piece_statistics_denorm

- piece_id
- cluster_id
- manifestation_id
- trs_start
- trs_end
- piece_length
- num_reception_edges
- num_different_work_ids
- num_work_ids_different_authors

Task Specific Data

Top Quotes Task Data

source_piece_statistics_denorm
piece_id
cluster_id
manifestation_id
trs_start
trs_end
piece_length
num_reception_edges
num_different_work_ids
num_work_ids_different_authors

IDs

Task Specific Data

Top Quotes Task Data

source_piece_statistics_denorm
piece_id
cluster_id
manifestation_id
trs_start
trs_end
piece_length
num_reception_edges
num_different_work_ids
num_work_ids_different_authors

Offsets

Task Specific Data

Top Quotes Task Data

source_piece_statistics_denorm
piece_id
cluster_id
manifestation_id
trs_start
trs_end
piece_length
num_reception_edges
num_different_work_ids
num_work_ids_different_authors

Metrics

Simple Query

- Document
- Quote Length
- Quote Metric
- Quantity

```
1  SELECT
2      piece_length as length,
3      num_work_ids_different_authors as metric,
4      piece_text(ts.text,trs_start,trs_end) as text
5  FROM
6      source_piece_statistics_denorm spsd
7      INNER JOIN textreuse_ids ti USING(trs_id)
8      INNER JOIN textreuse_sources ts USING(trs_id)
9  WHERE
10     manifestation_id = "A48874"
11     AND piece_length BETWEEN 150 AND 200
12     AND num_work_ids_different_authors >=20
13  ORDER BY
14      num_work_ids_different_authors DESC,
15      trs_start
16  LIMIT 12
```

Query Results

Index	Length	Metric	Text
1	172	42	"everal degrees of pleasure and pain, in all the things that environ and affect us*; and blended them together, in almost all that our Thoughts and Senses have to do with; t"
2	185	42	"n separating carefully *Ideas* one from another, wherein can be found the least difference, thereby to avoid being misled by Similitude, and by affinity to take one thing for another. T"
3	173	40	"rom the other. There are Fishes that have Wings, and are not Strangers to the airy Region: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold a"
4	192	39	"I World, we see no Chasms, or Gaps. All quite down from us, the descent is by easie steps, and a continued series of Things, that in each remove, differ very little one from the other. There a"
5	150	39	
6	194	37	"ion: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. There a"
7	171	37	"e being, in degrees of Perfection, much more remote from the infinite Being of GOD, than we are from the lowest state of Being, and that which approaches nearest to nothin"
8	191	36	"hat the Species of Creatures should also, by gentle degrees, ascend upward from us towards his infinite Perfection, as we see they gradually descend from us downwards: Which if it be probable"
9	165	34	"en both: Amphibious Animals link the Terrestrial and Aquatique together; Seales live at Land and at Sea, and Porpoises have the warm Bloud and Entrails of an Hog, no"
10	176	27	"njoyments of the Creatures can afford us, might be led to seek it in the enjoyment of him, *with whom there is fulness of joy, and at whose right hand are pleasures for evermor"
11	150	23	"nhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. Ther"
12	153	23	"hey gradually descend from us downwards: Which if it be probable, we have reason then to be perswaded, that there are far more Species of Creatures above"

Query Results

Index	Length	Metric	Text
1	172	42	"everal degrees of pleasure and pain, in all the things that environ and affect us*; and blended them together, in almost all that our Thoughts and Senses have to do with; t"
2	185	42	"n separating carefully *Ideas* one from another, wherein can be found the least difference, thereby to avoid being misled by Similitude, and by affinity to take one thing for another. T"
3	173	40	"rom the other. There are Fishes that have Wings, and are not Strangers to the airy Region: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold a"
4	192	39	"I World, we see no Chasms, or Gaps. All quite down from us, the descent is by easie steps, and a continued series of Things, that in each remove, differ very little one from the other. There a"
5	150	39	
6	194	37	"ion: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. There a"
7	171	37	"e being, in degrees of Perfection, much more remote from the infinite Being of GOD, than we are from the lowest state of Being, and that which approaches nearest to nothin"
8	191	36	"hat the Species of Creatures should also, by gentle degrees, ascend upward from us towards his infinite Perfection, as we see they gradually descend from us downwards: Which if it be probable"
9	165	34	"en both: Amphibious Animals link the Terrestrial and Aquatique together; Seales live at Land and at Sea, and Porpoises have the warm Bloud and Entrails of an Hog, no"
10	176	27	"njoyments of the Creatures can afford us, might be led to seek it in the enjoyment of him, *with whom there is fulness of joy, and at whose right hand are pleasures for evermor"
11	150	23	"nhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. Ther"
12	153	23	"hey gradually descend from us downwards: Which if it be probable, we have reason then to be perswaded, that there are far more Species of Creatures above"

Query Results

Index	Length	Metric	Text
1	172	42	"everal degrees of pleasure and pain, in all the things that environ and affect us*; and blended them together, in almost all that our Thoughts and Senses have to do with; t"
2	185	42	"n separating carefully *Ideas* one from another, wherein can be found the least difference, thereby to avoid being misled by Similitude, and by affinity to take one thing for another. T"
3	173	40	"rom the other. There are Fishes that have Wings, and are not Strangers to the airy Region: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold a"
4	192	39	"I World, we see no Chasms, or Gaps. All quite down from us, the descent is by easie steps, and a continued series of Things, that in each remove, differ very little one from the other. There a"
5	150	39	": Which if it be probable, we have reason then to be perswaded, that there are far more Species of Creatures above us, than there are beneath: we bein"
6	194	37	"ion: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. There a"
7	171	37	"e being, in degrees of Perfection, much more remote from the infinite Being of GOD, than we are from the lowest state of Being, and that which approaches nearest to nothin"
8	191	36	"hat the Species of Creatures should also, by gentle degrees, ascend upward from us towards his infinite Perfection, as we see they gradually descend from us downwards: Which if it be probable"
9	165	34	"en both: Amphibious Animals link the Terrestrial and Aquatique together; Seales live at Land and at Sea, and Porpoises have the warm Bloud and Entrails of an Hog, no"
10	176	27	"njoyments of the Creatures can afford us, might be led to seek it in the enjoyment of him, *with whom there is fulness of joy, and at whose right hand are pleasures for evermor"
11	150	23	"nhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. Ther"
12	153	23	"hey gradually descend from us downwards: Which if it be probable, we have reason then to be perswaded, that there are far more Species of Creatures above"

Query Results

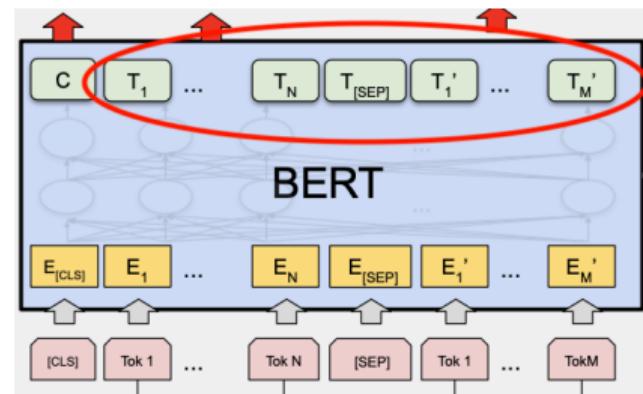
Index	Length	Metric	Text
1	172	42	"everal degrees of pleasure and pain, in all the things that environ and affect us*; and blended them together, in almost all that our Thoughts and Senses have to do with; t"
2	185	42	"n separating carefully *Ideas* one from another, wherein can be found the least difference, thereby to avoid being misled by Similitude, and by affinity to take one thing for another. T"
3	173	40	"rom the other. There are Fishes that have Wings, and are not Strangers to the airy Region: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold a"
4	192	39	"I World, we see no Chasms, or Gaps. All quite down from us, the descent is by easie steps, and a continued series of Things, that in each remove, differ very little one from the other. There a"
5	150	39	
6	194	37	"ion: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. There a"
7	171	37	"e being, in degrees of Perfection, much more remote from the infinite Being of GOD, than we are from the lowest state of Being, and that which approaches nearest to nothin"
8	191	36	"hat the Species of Creatures should also, by gentle degrees, ascend upward from us towards his infinite Perfection, as we see they gradually descend from us downwards: Which if it be probable"
9	165	34	"en both: Amphibious Animals link the Terrestrial and Aquatique together; Seales live at Land and at Sea, and Porpoises have the warm Bloud and Entrails of an Hog, no"
10	176	27	"njoyments of the Creatures can afford us, might be led to seek it in the enjoyment of him, *with whom there is fulness of joy, and at whose right hand are pleasures for evermor"
11	150	23	"nhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. Ther"
12	153	23	"hey gradually descend from us downwards: Which if it be probable, we have reason then to be perswaded, that there are far more Species of Creatures above"

Semantic Text Reuses

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding



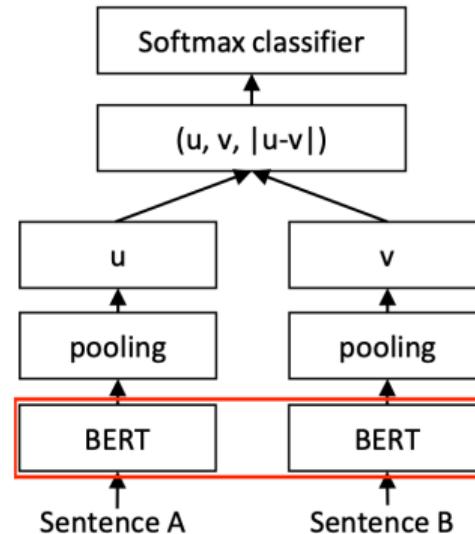
BERT: Pre-training of Deep Bidirectional Transformers for Language

Understanding

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding
- Semantically meaningful embeddings

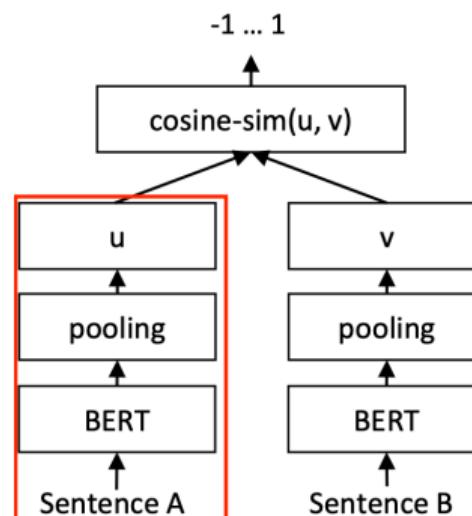


Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding
- Semantically meaningful embeddings
- Comparing embeddings for similarity



Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding
- Semantically meaningful embeddings
- Comparing embeddings for similarity

② FAISS (Facebook AI Similarity Search)

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding
- Semantically meaningful embeddings
- Comparing embeddings for similarity

② FAISS (Facebook AI Similarity Search)

- Inverted File Index

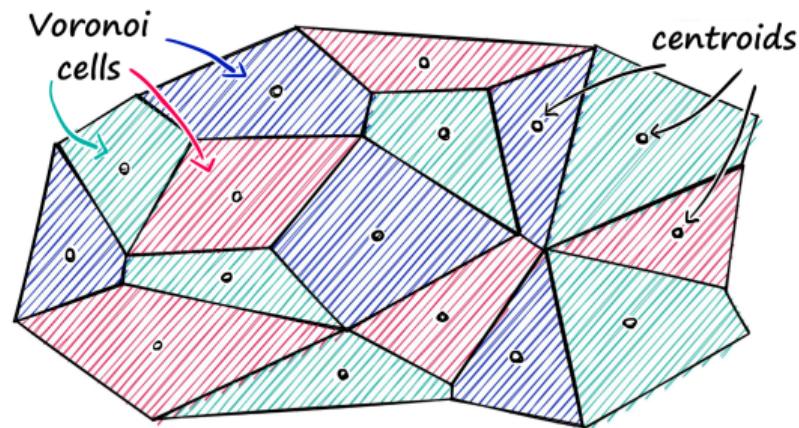


image source: pinecone.io FAISS-tutorial

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding
- Semantically meaningful embeddings
- Comparing embeddings for similarity

② FAISS (Facebook AI Similarity Search)

- Inverted File Index

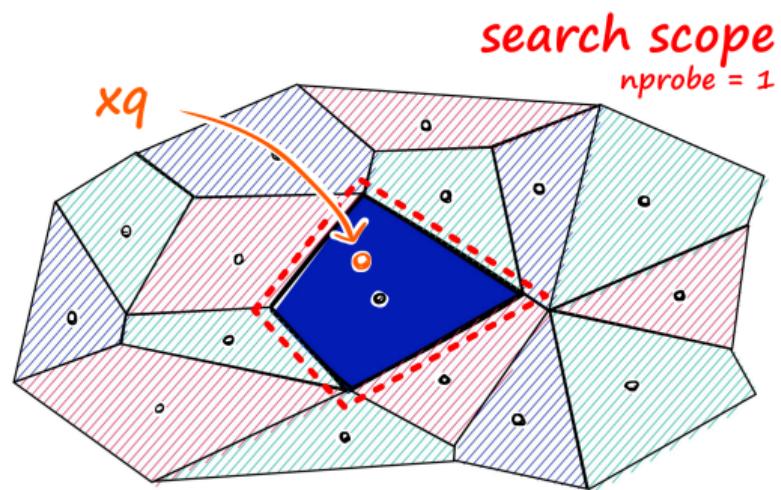


image source: pinecone.io FAISS-tutorial

Semantic Pipeline

① Sentence Transformers

- Sentence to vector embedding
- Semantically meaningful embeddings
- Comparing embeddings for similarity

② FAISS (Facebook AI Similarity Search)

- Inverted File Index
- Product Quantization

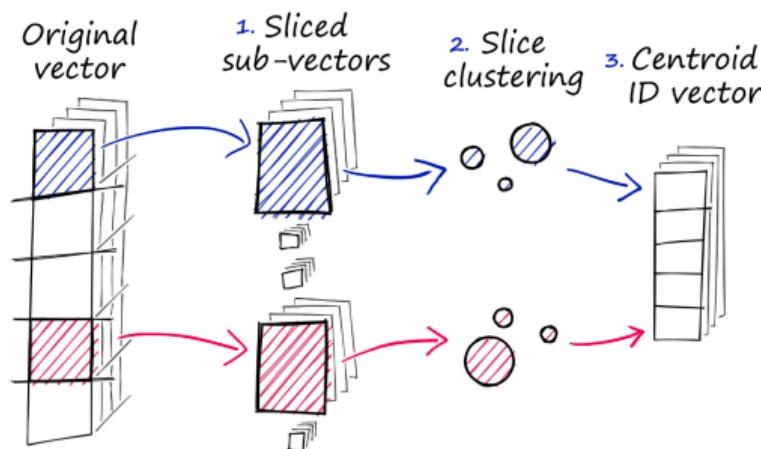


image source: pinecone.io FAISS-tutorial

Semantic Pipeline

① Sentence Transformers

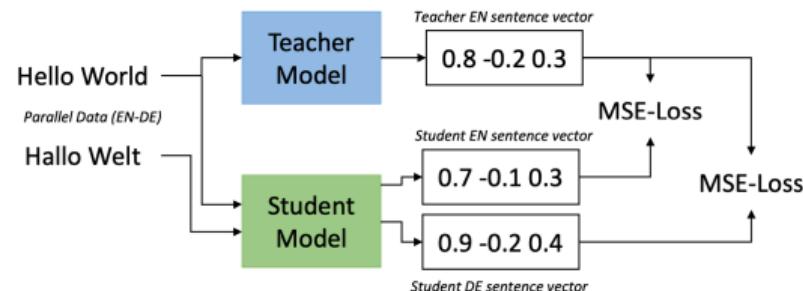
- Sentence to vector embedding
- Semantically meaningful embeddings
- Comparing embeddings for similarity

② FAISS (Facebook AI Similarity Search)

- Inverted File Index
- Product Quantization

③ Multilingual semantic search

- Teacher: English Sentence BERT
- Student: Multilingual BERT
- Result: Multilingual Sentence BERT



Making Monolingual Sentence Embeddings Multilingual using Knowledge
Distillation

Interface

English Query

The Animal and Vegetable Kingdoms are to nearly join'd, that if you will take the lowest of one, and the highest of the other, there will scarce be perceived any great difference between them

90 1690 - Locke, John (1632-1704)
An essay concerning humane understanding

the Animal and Vegetable Kingdoms, are so nearly join'd, that if you will take the lowest of one, and the highest of the other, there will scarce be perceived any great difference between them; and so on till we come to the lowest and the most inorganical parts of Matter, we shall find every-where, that the several Species are linked together, and differ but in almost insensible degrees. And when we consider

90 1795 - Unknown
The pleasing instructor

animal and vegetable kingdoms are so nearly joined, that if y ou will take the lowest of the one, and the highell of the other, there will scarcely be perceived any great dif- ference between them: And so on until we come to the lowest and the most inorganical parts of matter, we shall find every where that the several species are linked together, and differ but in almoil insensible de- grees

89 1759 - Formey, (1711-1797)
Philosophical miscellanies on various subjects

Kingdoms are so nearly joined, "that, if you will take the lowest of one "and the highest of the other, there will " scarce be perceived any great Difference " between them ; and so on, till we come "to the lowest and the most inorganical "Parts of Matter, we shall find, every- " where, that the several Species are link- " ed together, and differ but in almost in- "

89 1751 - Addison, Joseph (1672-1719)
The evidences of the Christian religion

" , and the animal and vegetable kingdoms are so nearly joined, that if you will take. the lowet " of one, and the highast of the other, there " will scarce be perceived any great difference " between them and: o on till: we

French Query

Les règnes animal et végétal sont à peu près unis, de sorte que si vous prenez le plus bas de l'un et le plus élevé de l'autre, on ne percevra guère de grande différence entre eux.

87 1690 - Locke, John (1632-1704)
An essay concerning humane understanding

the Animal and Vegetable Kingdoms, are so nearly join'd, that if you will take the lowest of one, and the highest of the other, there will scarce be perceived any great difference between them; and so on till we come to the lowest and the most inorganical parts of Matter, we shall find every-where, that the several Species are linked together, and differ but in almost insensible degrees. And when we consider

86 1751 - Addison, Joseph (1672-1719)
The evidences of the Christian religion

" , and the animal and vegetable kingdoms are so nearly joined, that if you will take. the lowet " of one, and the highast of the other, there will scarcely be perceived any great difference " between them: and: o on till: we come to the " lowest and the iaffi inorganical parts of mat- " ter, we shall find" everywhere that the feve- ral species are linked together

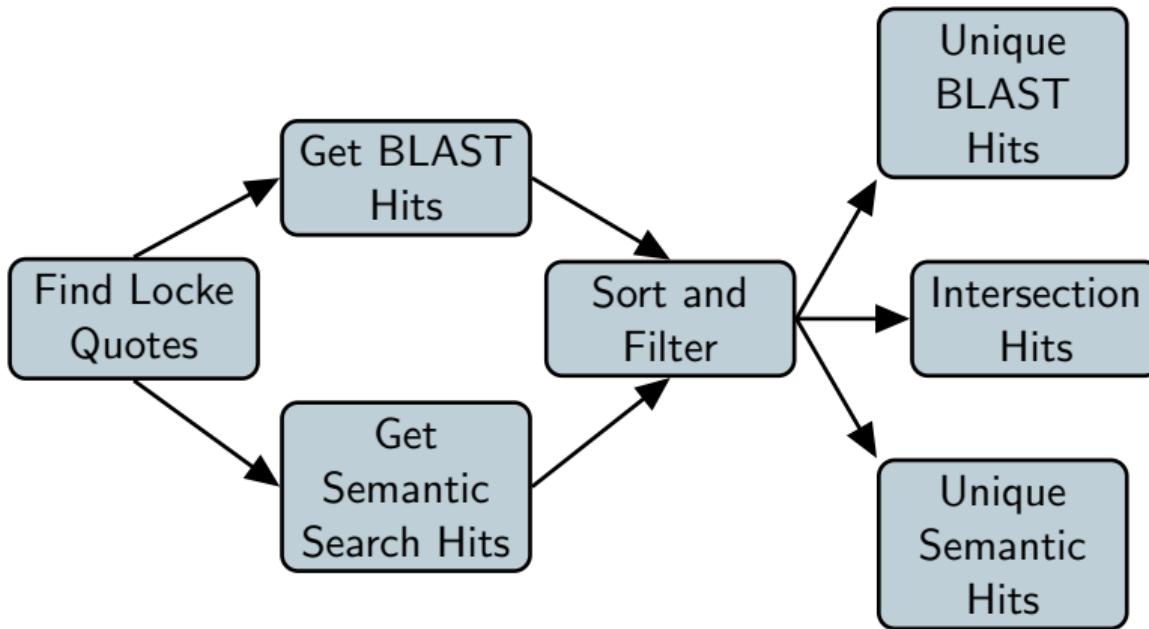
86 1795 - Unknown
The pleasing instructor

animal and vegetable kingdoms are so nearly joined, that if y ou will take the lowest of the one, and the highell of the other, there will scarcely be perceived any great dif- ference between them: And so on until we come to the lowest and the most inorganical parts of matter, we shall find every where that the several species are linked together, and differ but in almoil insensible de- grees

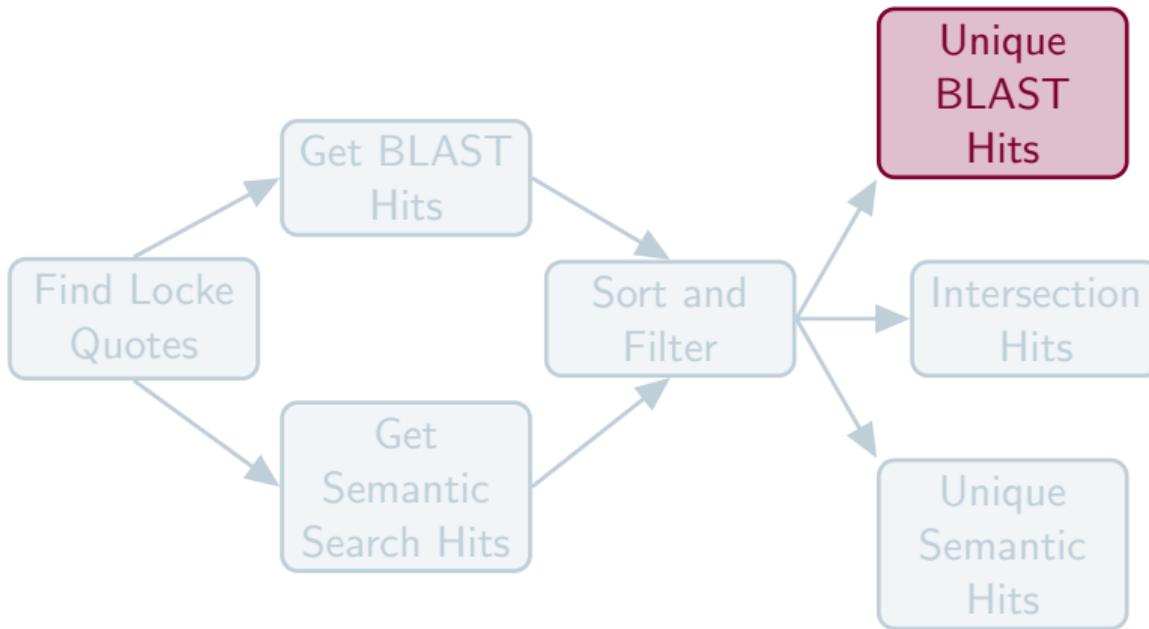
86 1783 - Monro, Alexander (1697-1767)
A treatise on comparative anatomy

connected, that if you take the highest of the one, and the lowest of the other, there will scarce be preserved any difference. For instance, what difference is there betwixt an oyster, one of the molt inorganised of the animal tribe, and the sensitive plant, the most exalted of the vegetable kingdom ? They both remain fixed to

Semantic Augmentation Example



Semantic Augmentation Example



Unique BLAST Hits

Example John Locke Quote

“The Animal and Vegetable Kingdoms
are to nearly join'd, that if you will take
the lowest of one, and the highest of the
other, there will scarce be perceived any
great difference between them.”

Unique BLAST Hits

Example John Locke Quote Piece

"the Animal and Vegetable Kingdoms,
are so nearly join'd, that if you will take
the lowest of one, and the highest of the
other, there will scarce be perceived an"

Locke, John, 1690 - An essay concerning humane understanding

All quite down from us, the descent is by easie steps, and a continued series of Things, that in each remove, differ very little one from the other. There are Fishes that have Wings, and are not Strangers to the airy Region: and there are some Birds, that are Inhabitants of the Water; whose Bloud is cold as Fishes, and their Flesh in taste so near akin, that the Scrupulous are allow'd them on Fish-days. There are Animals so near of kin both to Birds and Beasts, that they are in the middle between both: Amphibious Animals link the Terrestrial and Aquatique together; Seales live at Land and at Sea, and Porpoises have the warm Bloud and Entrails of an Hog, not to mention what is confidently reported of Mermaids, or Sea-men. There are some Brutes, that seem to have as much Knowledge and Reason, as some that are called Men: and the Animal and Vegetable Kingdoms, are so nearly join'd, that if you will take the lowest of one, and the highest of the other, there will scarce be perceived any great difference between them; and so on till we come to the lowest and the most inorganical parts of Matter, we shall find every-where, that the several Species are linked together, and differ but in almost insensible degrees. And when we consider the infinite Power and Wisdom of the Maker, we have reason to think, that it is suitable to the magnificent Harmony of the Universe, and the great Design and infinite Goodness of the Architect, that the Species of Creatures should also, by gentle degrees, ascend upward from us towards his infinite Perfection, as we see they gradually descend from us downwards: Which if it be probable, we have reason then to be perswaded, that there are far more Species of Creatures above us, than there are beneath; we being, in degrees of Perfection, much more remote from the infinite Being of GOD, than we are from the lowest state

Unique BLAST Hits

Example John Locke Quote Piece

"the Animal and Vegetable Kingdoms,
are so nearly join'd, that if you will take
the lowest of one, and the highest of the
other, there will scarce be perceived an"

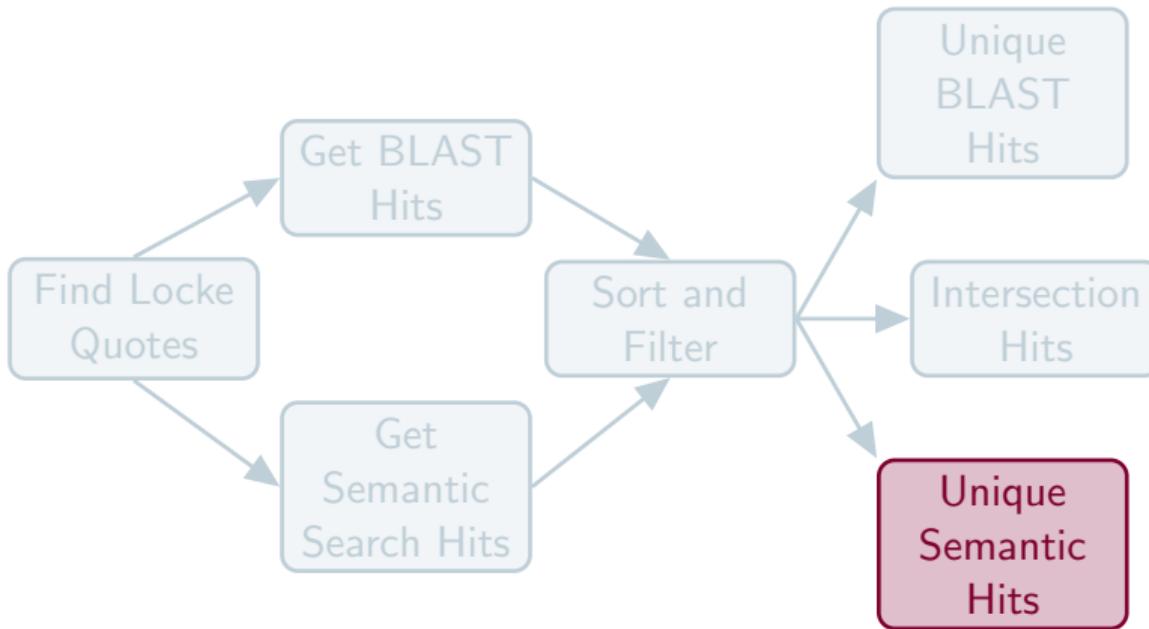
BLAST Hit

"he Ani-\nm lal~andr Vegetable-
KingdomsS are {ol~4 news~l \'" \'d,
that if\nyoul will takie the\' lowes .of:
one, and the ofp~~p the other,\nthere
will ~fearce be perceived an"

1720 - The spectator

176 *The SPECTATOR.* N^o. 519.
Animals link the Terrestrial and Aquatich together ; Seals live at Land and at Sea, and Porpoises have the warm Blood and Entrails of a Hog ; not to mention what is confidently reported of Mermaids or Sea-Men. There are some Brutes, that seem to have as much Knowledge and Reason, as some that are called Men ; and the Animal and Vegetable Kingdoms are so nearly join'd, that if you will take the lowest of one, and the highest of the other, there will scarce be perceived any great difference between them : and so on till we come to the lowest and the most inorganic parts of Matter, we shall find every where that the several Species are linked together, and differ but in almost insensible degrees. And when we consider the infinite Power and Wisdom of the Maker, we have reason to think that it is suitable to the magnificent Harmony of the Universe, and the great Design and infinite Goodness of

Semantic Augmentation Example



Unique Semantic Search Hits

Topical Match

'vegetable kingdoms seem to- be so essentially separated, that we cannot even imagine- the leaR approximation of the one to the other. Thhole- naturalilts, indeed,, who have foppofed the ditiuaison- between animals and vegetables to consist in any thing- else than what we have already mentioned, have found- themselves greatly embarrassed, and have generally- agreed, that it was extremely difficult, if not impolfible,- to fettle'

1797 - Encyclopædia Britannica
enamelling. See ENAMEL.

ANIMADVERSION, in matters of literature, is used to signify, sometimes correction, sometimes remarks upon a book, &c. and sometimes a serious consideration upon any point.

ANIMAL, in natural history, an organized and living body, which is also endowed with sensation : thus, minerals are said to grow or increase, plants to grow and live, but animals alone to have sensation.

It is this property of sensation alone that can be deemed the essential characteristic of an animal ; and by which the animal and vegetable kingdoms seem to be so essentially separated, that we cannot even imagine the least approximation of the one to the other. Those naturalists, indeed, who have supposed the distinction between animals and vegetables to consist in any thing else than what we have already mentioned, have found themselves greatly embarrassed, and have generally agreed, that it was extremely difficult, if not impossible, to settle the boundaries between the animal and vegetable kingdoms. But this difficulty will be easily seen to arise from their taking the characteristic marks of

an
cu
pla
am
an
or
vel
no
foc
fol
for
toj
en
ye
th
no
of
fin
th
th
roi

Discussion Questions

Questions for future work

Lexical Text Reuses

- Handling fragmented/overlapping reuses
- Managing metadata from different sources
- Building more interfaces for custom tasks

Semantic Text Reuses

- Calibrate similarity scores across queries
- Identify region of similarity within a hit
- Filter out topical matches