

This is a supporting document for reference

To achieve the extraction and creation of the discussed JSONL pairs, I performed the following steps:

1. Fetching and Storing Data using API:

- You utilized the OpenAlex API (<https://api.openalex.org/>) to fetch relevant data.
- Endpoint examples include works, topics, and their corresponding details.

2. Creation of JSONL Pairs:

- You created JSONL pairs by combining different elements in specific formats.
 - **name_domain_field.jsonl**: Combining names with domain and field information.
 - **name_domain_subfield.jsonl**: Combining names with domain and subfield information.
 - **name_field_subfield.jsonl**: Combining names with field and subfield information.
 - **title_abstract.jsonl**: Combining titles with abstracts.
 - **title_key.jsonl**: Combining titles with keywords.
 - **title_ngrams_work.jsonl**: Combining titles with ngrams from related works.
 - **title_referenced_work.jsonl**: Combining titles with referenced works.
 - **title_related_work.jsonl**: Combining titles with related works.
 - **topic_name_key.jsonl**: Combining topics with names and keywords.

3. Formats for Creating Training / Triplets Pairs:

- {"texts": ["Anchor 1", "Positive 1"]}
- {"texts": ["Anchor 1", "Positive 1", "Negative 1"]}
- Anchors were topic display names, positives were domains, and negatives were fields.

[Expanded Diagram how Voyage API was used to find the closest/Nearest Neighbors](#)

Or see the diagram below.

