Nexus-Graph (Technical Information)

The system has 5 core classes:

1. File Splitter.
2. Features Miner.
3. Materials & Suppliers Miner.
4. Instructor.
5. Organizer.

Let’s walk through it one by one!

1. File Splitter:

This class used to split the file into pieces based on some keywords ('TITLE','ABSTRACT','INTRODUCTION','METHODS','RESULTS','DISCUSSION','Abstract', 'Background', 'Introduction', 'Literature Review', 'Methods', 'Results', 'Discussion', 'Conclusion', 'Keywords') it could find, when it notices some of keys it split the text into two parts before the keyword and after the keyword.

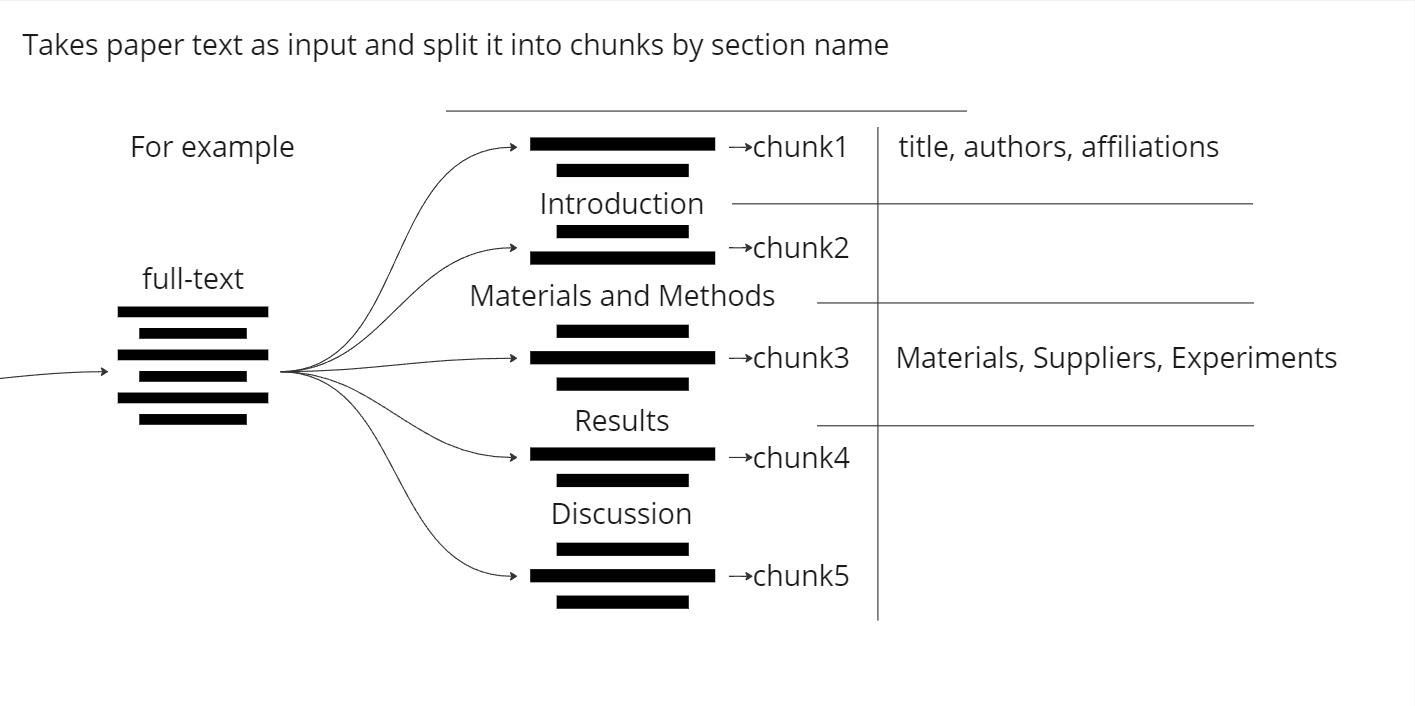


Figure1: File Splitter Diagram to describe how it’s works.

1. Features Miner:

The Features Miner takes a block of text and it uses OpenAI API and starts to mine and search for title, authors, affiliations, and tags, then it form it in a JSON file..

If the miner didn’t got any title or tags it generates ones based on the context in the text.

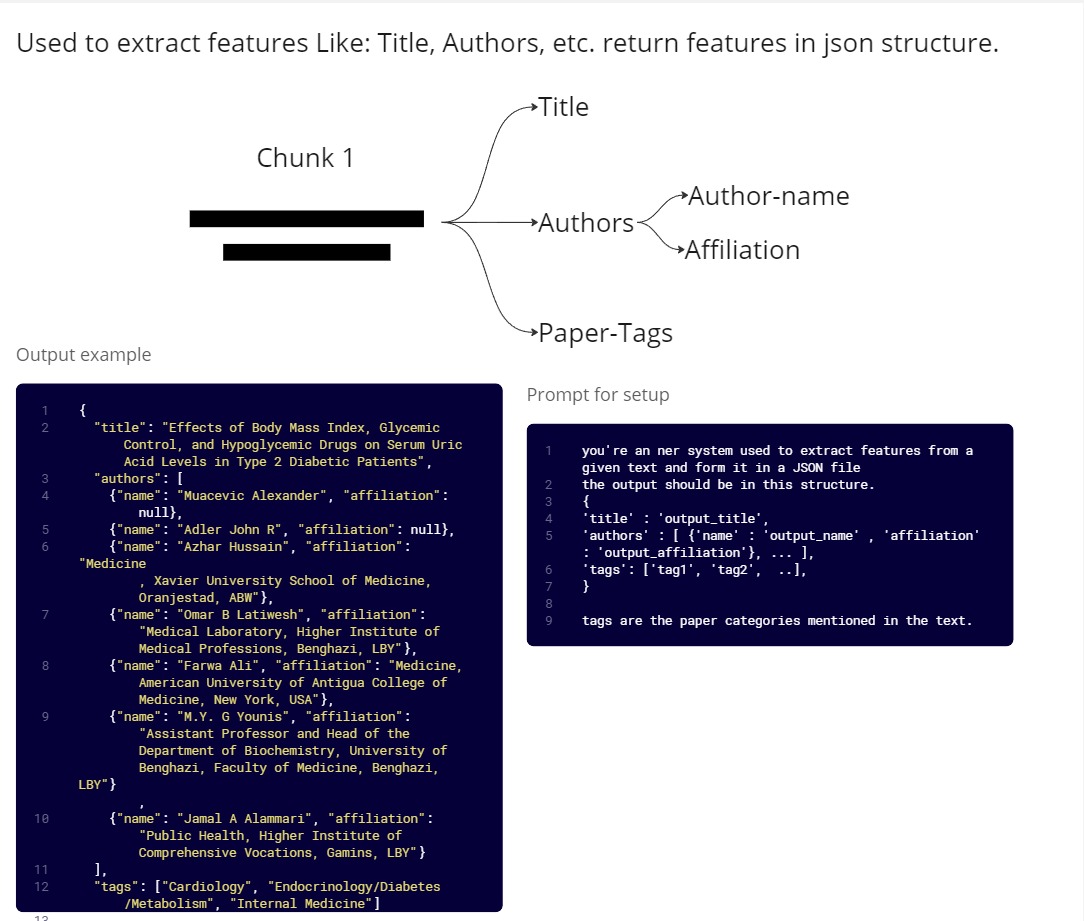


Figure2: Features Miner Diagram.

1. Materials & Suppliers Miner:

The Materials & Suppliers Miner gets the text and uses OpenAI API to mine and search for the materials and suppliers mentioned in text, then it form it in a JSON file.



Figure3: Materials & Suppliers Miner Diagram.

1. Instructor:

The Instructor gets the text and it uses OpenAI API to understand the context and extract the methodologies and generate instructions about how to perform the experiments and apply on methodologies have been found in the text, then it form it in a JSON file.

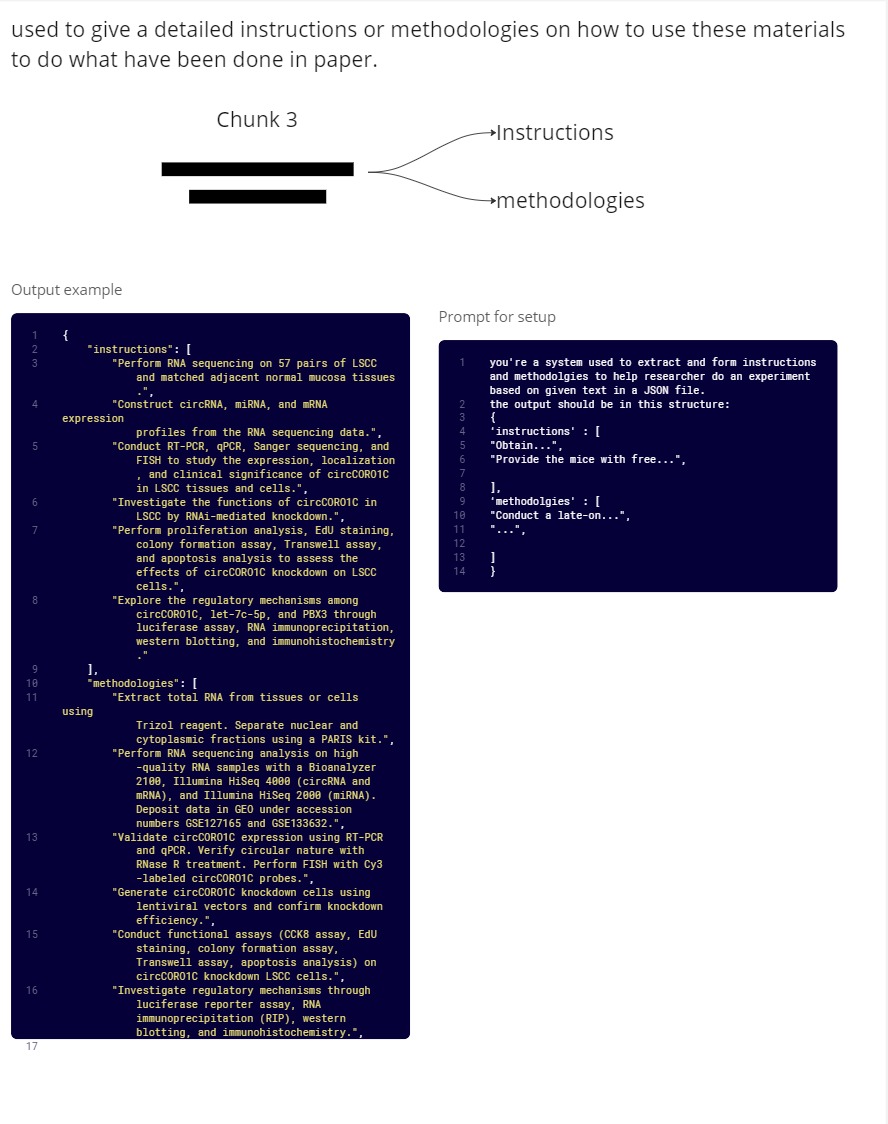


Figure4: Instructor Diagram

1. Organizer:

The organizer creates the tables of database, and takes Json files and merge it in one file and check if it’s stored or not, if not it’s storing the information one by one in database.

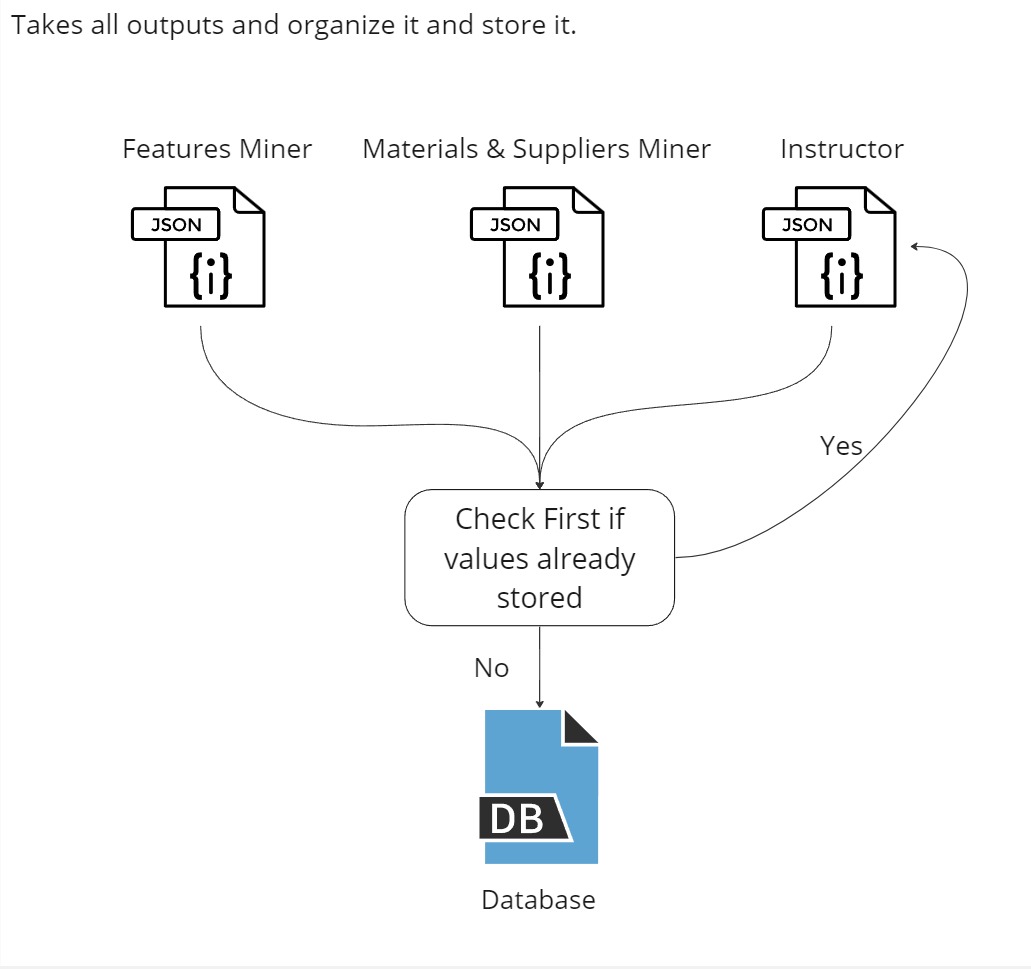


Figure5: Organizer Diagram

WorkFlow:

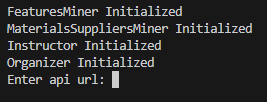
The workflow going like this:

1. Initialize all classes and Ask for the API link to retrieve the papers.

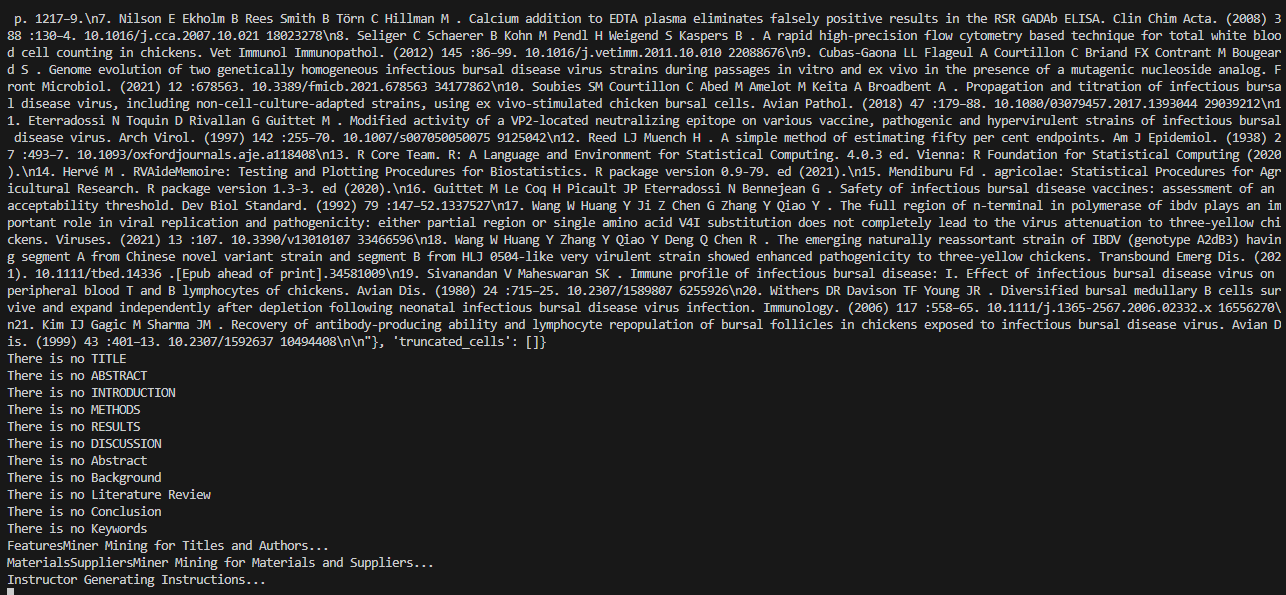
Note: File splitter don’t need to be initialized.

1. Retrieve the papers one by one.
2. Every paper goes through a pipeline:
   1. File Splitter: Split the file into chunks.
   2. Features Miner: takes the first chunk and extract authors, affiliation, title, tags.
   3. Materials & Suppliers Miner: get the methods section (third chunk) and extract Materials & Suppliers.
   4. Instructor: get the methods section (third chunk) and gives Instruction about how to perform mentioned experiments and methodologies.
   5. Organizer: get all Json files from other classes and merge it in one file and store it in the database.
3. Once data is stored we go back for step 2.

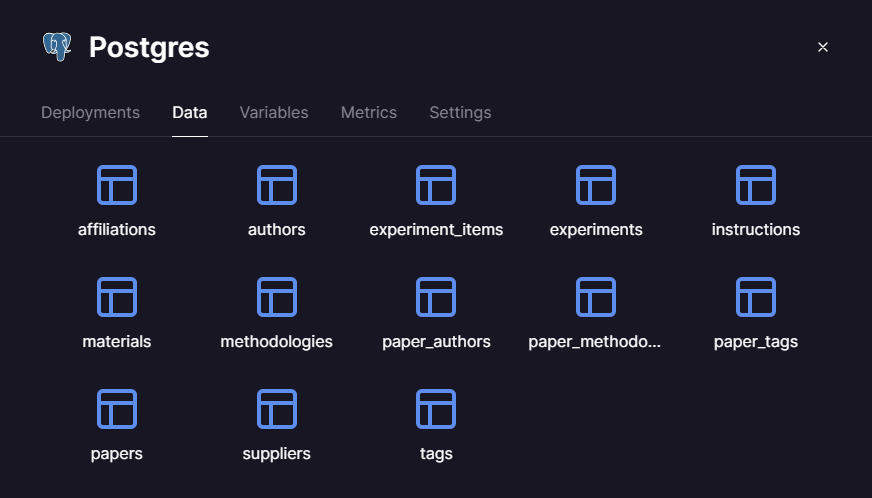
Screenshots:



Initializing classing and asking for api url.



Reading papers and mining for information.



Database Structure.