

# Automated Literature Review Curation Plan

## 1. Manual Google Scholar Search:

- Perform manual searches on Google Scholar using relevant keywords.
- Copy the contents of the first three pages.

## 2. Data Organization in Google Sheets:

- Paste the copied content into a structured Google Sheets format.
- Organize the data by query, contents of page one, two, and three.

## 3. Data Extraction:

- Extract key information from the search results:
  - Title of the research paper.
  - First author.
  - Number of citations.

## 4. DOI Retrieval:

- Use an API to obtain the DOI (Digital Object Identifier) of each research paper from Crossref.

## 5. Downloading PDFs:

- Pass the DOI to either SciDownload or PyGetPapers (Python services) to download the corresponding PDFs.
- Store these PDFs in a dedicated folder.

## 6. PDF Parsing:

- Utilize Nougat from Meta to convert the PDFs into LaTeX format.

## 7. Data Extraction with Language Models:

- Feed the LaTeX data to language models.
- Extract structured information:
  - Abstract of the paper.
  - Similarity score to the intended research goal.
  - Technologies mentioned.
  - Applications (e.g., IoT in agriculture).
  - Keywords.

## 8. Paper Evaluation and Ranking:

- Rank the papers based on:
  - Number of citations.
  - Similarity score.
  - Journal reputability (create a journal score based on reputability).
- Compute a total paper ranking score.

## 9. Final Review and Writing:

- Review the top-ranked papers.
- Write the comprehensive literature review based on the findings.

## Additional Notes:

- Make sure to comply with Google Scholar's terms of service during manual data collection.
- Ensure that the APIs and tools used for downloading and parsing PDFs are used legally and ethically.
- Consider the validity and reliability of the extracted data and the methods used for ranking the

papers.