

Task 1: Researcher Profile Mining and Analysis

Objective:

Each student will analyze a set of researchers by mining their Google Scholar profiles. The goal is to build a comprehensive profile for each researcher based on their recent publications, and perform insightful analysis of their research themes and diversity.

Instructions:

You have been assigned a list of researchers (names provided in your individual Excel sheet). For each researcher, complete the following:

Step 1: Fetch Researcher's Google Scholar Profile

- Use Google Scholar search (<https://scholar.google.com>) with the researcher's full name.
- Alternatively, explore tools like:
 - Publish or Perish (<https://harzing.com/resources/publish-or-perish>)
 - SerpAPI (for programmatic access — limited free usage)
 - Scraper tools (act wisely as some may be rate-limited or blocked)

Step 2: Extract Publications

- From the researcher's Google Scholar profile, fetch the most recent 20 publications.
- For each publication, record the:
 - Title
 - Abstract (may be available on Google Scholar or linked platforms)

Create one Excel sheet per researcher. Inside each sheet, include:

S.No	Researcher Name	Title of the Paper	Abstract
------	-----------------	--------------------	----------

Step 3: Build Author Profile

3.1 Identify the Core Area of Work

- Combine all abstracts of that author's papers into one text.
- Use NLP techniques like TF-IDF, NER, Topic Modeling, BERT-based keyword extraction.
- Optional: Generate a Word Cloud.

Example Output:

Researcher	Top Research Themes
Nishith Kotak	Natural Language Processing (NLP), Artificial Intelligence (AI), Internet of Things (IoT), Educational Technology

3.2 Analyze Research Diversity

- Compute semantic similarity between all 20 abstracts using sentence embeddings.
- Calculate average pairwise similarity score.
- High score: focused area; Low score: diverse topics.

Example Output:

Researcher	Average Similarity	Diversity Score
Nishith Kotak	0.72	Medium Diversity

Submission Format:

- One Excel file with:
 - Individual sheets for each researcher profiles.
 - Summary sheet titled 'Author_Profiles'.
 - Summary sheet titled 'Author_diversity'.
- Zip of the Code for entire process