Report on Research Paper Retrieval from PubMed

1. Introduction

This report outlines the approach, methodology, and results of the Python program developed to fetch research papers from PubMed based on user-specified queries. The program aims to identify papers with authors affiliated with pharmaceutical or biotech companies and output the results in a structured CSV format.

2. Approach

The program automates the retrieval of research papers from PubMed using the Entrez API. The approach consists of the following key components:

1. **User Query Input:** The user provides a query term related to a research topic (e.g., "cancer research").
2. **PubMed API Search:** The program sends the query to the PubMed database using the NCBI Entrez API.
3. **Data Filtering:** The program extracts metadata such as paper ID, title, year of publication, authors, and affiliations.
4. **Affiliation Analysis:** The extracted affiliations are checked for references to pharmaceutical or biotech companies.
5. **Output Generation:** The final processed data is saved as a CSV file or displayed in debug mode.

3. Methodology

The methodology involves the following structured steps:

3.1 Data Retrieval

- The Entrez API is used to fetch research paper metadata from PubMed.
- The search query is processed and formatted to optimize retrieval.

3.2 Data Extraction

- The program extracts metadata including:
    - Paper ID (PMID)
    - Title
    - Year of publication
    - Author names
    - Author affiliations

3.3 Affiliation Analysis

- The affiliations of authors are analysed using keyword matching techniques.
- Keywords such as "pharmaceutical", "biotech", "biopharma", and company names are used for identification.

3.4 Output Generation

- The extracted and filtered data is saved in CSV format.
- Debug mode allows real-time viewing of processed results in the console.

4. Results

The program successfully retrieved research papers relevant to the specified query and identified papers associated with pharmaceutical or biotech companies. Sample results include:

A1 | PubMed ID

	A	B	C	D
1	PubMed ID		Title	Non-academic Authors
2	40066659	Inhibition of HDAC4 in granulosa cells improved co-cultured oocyte maturation in	2025	Zuo Z, Lin L, Li H, Wang T, Xia G, Wang C
3	40066446	Exercise boost after surgery improves survival in model of metastatic breast	2025	Stagaard R, Schauer T, Oria VO, Zornhagen KW, Albrechtsen R, Christensen JF
4	40066442	A comprehensive prognostic and immunological implications of Gremlin 1 in lung	2025	Xiao J
5	40066440	In-depth analysis of the safety of CAR-T cell therapy for solid tumors.	2025	Jin Y, Su T, Chen S
6	40066333	Cepharanthine hydrochloride: a novel ferroptosis-inducing agent for prostate	2025	Guan JS, Jia J, Zhou YQ, Zheng JJ, Lin QM, Fan JL, Wang Y
7	40066322	Biomimetic Self-Oxygenated Immunoliposome for Cancer-Targeted Photodynamic	2025	Tang Y, Tang T, Li Y, Wu J, Liu X, Xiang D, Hu X
8	40066226	Depletion of Acetyl-CoA Carboxylase 1 Facilitates Epithelial-Mesenchymal	2025	Lai J, Liu S, Chen Y, Chen J, Liang Z, Mei X, Feng Y, Liu J, He H, Zhong W
9	40066115	Melatonin Aids in Treating Mood and Sleep Problems Resulting from Hormonal	2024	Vaziri N, Shakourifar M, Sattari P, Moghaddas A
10	40066114	Malleatin A and B: New Premyrsine-Type Diterpenes from Euphorbia malleata with	2024	Zolfaghari B, Akbari F, Esmaeili S, Aghaei M, Mosaffa F, Ghorbanhosseini SS
11	40066113	Evaluation of Oral Nano-Silymarin Formulation Efficacy in the Prevention of	2024	Karbasforooshan H, Arasteh O, Ghavami V, Elyasi S

E1 | Company Affiliations

	E	F
1	Company Affiliations	Corresponding Author Email
2	State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, , State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, , State Key L	Not Available
3	Biotech Research and Innovation Center (BRIC), University of Copenhagen (UCPH), , Biotech Research and Innovation Center (BRIC), University of Copenhagen (UCPH), , Biote	Not Available
4	Institute of Disinfection and Infection Control, Heilongjiang Provincial Center	Not Available
5	Guangdong Provincial Engineering Research Center for Precision Medicine in , School of Clinical Medicine, Guangdong Pharmaceutical University, Guangzhou, , Guangdong P	Not Available
6	Guangdong Provincial Key Laboratory of Large Animal Models for Biomedicine, South , Guangdong Provincial Key Laboratory of Large Animal Models for Biomedicine, South , G	Not Available
7	Department of Pharmacy, The Second Xiangya Hospital, Central South University, , Hunan Provincial Engineering Research Center of Translational Medicine and , Institute of C	Not Available
8	Guangdong Provincial Key Laboratory of Urology The First Affiliated Hospital of , Guangdong Provincial Key Laboratory of Urology The First Affiliated Hospital of , Guangdong Pro	Not Available
9	Department of Clinical Pharmacy and Pharmacy Practice, School of Pharmacy and , Department of Clinical Pharmacy and Pharmacy Practice, School of Pharmacy and , Depa	Not Available
10	Department of Pharmaceutics, Novel Drug Delivery Systems Research Centre, School , Isfahan Pharmaceutical Sciences Research center, Isfahan University of Medical , Dep	Not Available
11	Department of Clinical Pharmacy, School of Pharmacy, Mashhad University of , Department of Clinical Pharmacy, School of Pharmacy, Mashhad University of , Department of	Not Available

5. Conclusion

The implemented Python program effectively automates research paper retrieval and affiliation analysis for pharmaceutical and biotech-related studies. It provides a structured and efficient approach to identifying relevant scientific publications, aiding researchers in literature reviews and industry-related research monitoring. Future enhancements could include natural language processing (NLP) for improved affiliation analysis and automated company name recognition.

Keywords: PubMed API, Research Paper Retrieval, Affiliation Analysis, Pharmaceutical Industry, Biotech Companies, Python Automation