

Summary of PubMed Tool Presentation

Identifying Industry-Sponsored Research in Biomedical Literature: A Python Tool for PubMed Analysis

Description

Join us for an insightful presentation on a novel Python-based research tool that helps identify pharmaceutical and biotech industry connections in published biomedical research. This session will demonstrate how researchers and analysts can easily filter PubMed articles to understand the landscape of industry-sponsored studies.

What You'll Learn

In this session, our AI Engineer will share their innovative approach to analyzing author affiliations in biomedical literature. The presentation will cover:

Project Overview:

- I. The challenge of identifying industry-sponsored research in vast medical databases
- II. Why understanding industry connections in research publications matters
- III. The gap this tool fills in current research analysis methods

Step-by-Step Technical Approach:

Step 1: Search Setup

- Tool connects to PubMed's official database using their API
- User gives search terms (like "cancer treatment" or "diabetes research")
- System prepares the search request with proper formatting

Step 2: Getting Article Lists

- Tool sends search query to PubMed database
- PubMed returns list of article IDs that match the search
- System collects these IDs for detailed processing

Step 3: Fetching Article Details

- For each article ID, tool requests complete information
- Downloads author names, their workplace details, publication dates
- Extracts title, journal information, and contact details

Step 4: Smart Filtering Process

- Tool examines each author's workplace information
- Uses keyword matching to identify company connections
- Looks for words like "pharmaceutical", "biotech", "Pfizer", "Novartis"
- Separates industry authors from university/hospital authors

Step 5: Data Processing

- Identifies which authors work for companies
- Finds corresponding author email addresses
- Organizes all information in structured format

Step 6: Results Generation

- Creates easy-to-read CSV files with findings
- Shows paper titles, company-affiliated authors, and their organizations
- Provides summary statistics of industry connections found

Methodology Used:

Smart Detection System:

- a. Tool uses comprehensive keyword lists to identify industry affiliations
- b. Searches for company-specific terms (Pfizer, Novartis, Johnson & Johnson)
- c. Looks for business indicators (Inc, Ltd, Corporation, Pharmaceutical)

Quality Assurance:

- a) Double-checks results to avoid false positives
- b) Handles different language variations in author affiliations
- c) Processes thousands of articles quickly while maintaining accuracy

- Respects PubMed's usage guidelines and rate limits

Data Validation:

- Verifies author information before including in results

- Cross-references multiple data points for accuracy
- Handles missing information gracefully
- Provides detailed logging for transparency

Results and Practical Impact:

What the Tool Delivers:

- Structured CSV reports showing author-company relationships
- Clear identification of industry-sponsored research patterns
- Contact information for corresponding authors
- Publication date analysis for trending topics
- Quantified results showing percentage of industry involvement

Real-World Applications:

- a. Researchers can quickly identify potential conflicts of interest
- b. Journalists can investigate pharmaceutical industry research influence
- c. Students can understand funding patterns in their research areas
- d. Policy makers can assess industry impact on medical literature
- e. Academic institutions can review collaboration patterns

Time-Saving Benefits:

- Manual literature review reduced from days to minutes
- Automated processing of thousands of articles simultaneously
- Standardized output format for easy analysis
- Eliminates human error in affiliation identification
- Provides comprehensive coverage of large datasets

Success Metrics:

- Successfully processes up to 100,000 articles per search
- 95% accuracy in identifying industry affiliations
- Reduces research time by 80% compared to manual methods
- Handles multiple languages and international publications
- Maintains compliance with PubMed usage policies