**API Schema Extractor.txt**

Key Workflow Components

1. **Scheduling and Triggering:**
   * **Schedule Trigger:** The workflow is set to run on a specific schedule (Monday through Friday at 8 AM), ensuring that the latest research papers are periodically processed.
   * **Loop Over Items:** Processes multiple items (papers) in batches, ensuring efficient handling of a large number of documents.
2. **Web Request and Data Extraction:**
   * **Request Hugging Face Paper:** A HTTP Request node queries Hugging Face's papers, filtering based on a date parameter (e.g., fetching papers published in the last day).
   * **Extract Hugging Face Paper:** This node uses BeautifulSoup (via a Python code node) to parse the returned HTML content, extract relevant metadata (like title, URL, score, author, etc.), and compile a list of papers.
   * **Split Out:** The extracted results are then split into individual items for further processing.
3. **AI Analysis:**
   * **OpenAI Analysis Abstract:** Leveraging an OpenAI Chat Model, this node analyzes the paper abstract. The analysis extracts key details like core introduction, keywords, data/results, technical details, and academic classification.
   * **JSON Output:** The output is structured as JSON, ensuring consistent data formatting that can be stored in Notion.
4. **Notion Integration:**
   * **Store Abstract Notion:** This node takes the analyzed data and stores it in a Notion database. The database includes fields such as URL, title, abstract, publication date, classification, technical details, data/results, and keywords.

Data Flow Overview

1. **Initiation:** The schedule trigger starts the workflow at predetermined intervals.
2. **Web Scraping:** The workflow makes an HTTP request to fetch the latest research papers from Hugging Face. The Python code extracts key metadata from the HTML.
3. **AI Analysis:** The extracted paper abstracts are analyzed using an OpenAI model configured with a detailed system prompt that guides the extraction of key details. The output is structured as JSON.
4. **Data Storage:** The processed data is then uploaded into a Notion database, making it available for future reference, reporting, or further analysis.
5. **Batch Processing:** The use of split and loop nodes ensures that multiple documents can be processed in batches, enhancing scalability and performance.

Use Case and Customization

This workflow is ideal for research teams, academic institutions, or anyone interested in keeping a structured database of recent academic publications, especially from Hugging Face. By integrating with Notion, users can easily view, edit, and share the extracted information.

* **Customization Options:**
  + **Date Filters:** You can modify the date parameter to adjust the time window for the research papers you want to retrieve.
  + **System Prompts:** The OpenAI analysis prompt can be customized to extract different or additional information based on your research needs.
  + **Notion Database Structure:** Modify the Notion node to store additional fields or adjust existing ones based on the data you want to capture.