🔧 SOP: Automating Lead Generation with HTTP Requests in n8n

Overview

This SOP outlines how to automate the process of scraping emails and contact details from various sources—including Google Maps, LinkedIn, and company websites—using HTTP request nodes in n8n. It also demonstrates how to clean the extracted data, use AI agents to parse web content, and send API requests to platforms like Apify and Apollo. By following this guide, you will be able to build a powerful, semi-automated lead generation system that outputs structured data for use in CRMs or email campaigns.

Step-by-Step Guide

1. **Set Up a Basic Workflow in n8n**
    a. Create a new workflow in n8n.
    b. Add a `Trigger` node to initiate your flow.
    c. Add an `HTTP Request` node to make web requests.
2. **Perform a Basic GET Request**
    a. Configure the `HTTP Request` node:
        - Method: `GET`
        - URL: Include full URL with `http://` or `https://`
    b. Execute to fetch raw HTML content from the target website.
3. **Clean HTML Data**
    a. Add a `Code` node to parse the HTML.
    b. Use JavaScript to strip unwanted tags and isolate text.

c. Alternatively, use a `Markdown` node to convert HTML to plain text.

4. **Extract Contact Details Using AI**
   a. Add an `AI Agent` node.
   b. Feed the cleaned HTML to the agent.
   c. Use a system prompt like:

> You are a helpful assistant. Extract contact information and a short business description from the provided HTML.

5. **Send POST Requests to External APIs (e.g., Together.ai)**
   a. Use the API's playground to generate a CURL request.
   b. Convert the CURL into n8n's HTTP Request format.
   c. Configure headers with authentication:
      - Use `Generic Credential` with `Header Auth` to securely store API keys.
   d. Send body as raw JSON to match API documentation.
6. **Automate Google Maps Scraping**
   a. Create a JSON array of search queries (e.g., accountants in London).
   b. Loop through each query using n8n's `Execute Workflow` node.
   c. Hit Google Maps search pages and extract business URLs.
   d. Perform a GET request to each website URL.
7. **Extract Emails from Business Websites**
   a. Clean HTML as previously described.
   b. Use AI to extract emails, phone numbers, and business details.
   c. Format the output using an `Output Parser` node.
8. **Filter and Deduplicate Results**
   a. Aggregate all email results.
   b. Use filter nodes to remove duplicates and irrelevant emails.
   c. Export the cleaned data to Google Sheets, Airtable, or your CRM.
9. **Use Apify to Scrape LinkedIn and Apollo**

a. Create an Apify account and obtain your API key.

b. Install Apify community node in n8n.

c. Use `Run Actor` to:

- Scrape LinkedIn profiles.
- Extract emails and contact data from Apollo search pages.

d. Use `Get Dataset Items` to retrieve results using dataset ID.

10. **Integrate With Email Campaigns**

a. Filter collected data to retain leads with valid emails and phone numbers.

b. Use AI agents to generate personalized email icebreakers.

c. Send cold emails using n8n's email integration tools.

Warnings and Notes

- ❗ Always prepend URLs with `http://` or `https://` to avoid request errors.
- ⚠️ Avoid hardcoding API keys in headers; use `Generic Credential` for secure management.
- 🧠 Be mindful of rate limits when scraping services like Google Maps or LinkedIn. Add `Wait` nodes to avoid throttling.
- 📄 Not all websites expose emails directly—some use protection or hide data via JavaScript.
- 💸 Some features (e.g., Apollo and Apify scraping) require paid plans for full functionality.

Visual Diagram

```
flowchart TD
    A[Start Workflow] --> B[HTTP GET Request to Website]
    B --> C[Clean HTML using Code/Markdown]
    C --> D[Extract Data with AI Agent]
    D --> E[Store Results in Google Sheets]
    E --> F[Filter & Deduplicate Emails]
    F --> G[Send to CRM or Email Campaign]
```

```
    A2[Run Apify Actor for LinkedIn/Apollo] --> H[Get Dataset Items]
    H --> F
    subgraph Google Maps Scraper
        A3[Loop Search Queries] --> B3[Scrape URLs from Maps]
        B3 --> C3[Fetch Website HTML]
        C3 --> D
    end
```
