British Pathé Video Crawler - Installation Guide

This guide will help you set up and run the British Pathé video crawler on your Mac. Follow these steps carefully, and don't hesitate to ask for help if you get stuck!

Prerequisites (One-time Setup)

1. **Install Command Line Tools**
   - Open Terminal (press Cmd + Space, type "Terminal", press Enter)
   - Copy and paste this command:
     `xcode-select --install`
   - Click "Install" when prompted
   - Wait for installation to complete (may take 5-15 minutes)
2. **Install Homebrew**
   - In Terminal, copy and paste this command:
     `/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"`
   - Follow any prompts that appear
   - When it's done, restart Terminal
3. **Install Python**
   - In Terminal, type:
     `brew install python`
   - Wait for installation to complete

Setting Up Your Project

1. **Create a Project Folder**
   - Open Finder
   - Click on "Documents" in the sidebar
   - Click "File" → "New Folder"
   - Name it "pathe-crawler"
2. **Save the Script**
   - Create a new file called `crawler.py`
   - Copy the entire Python script into this file
   - Save it in your "pathe-crawler" folder
3. **Install Required Packages**
   - Open Terminal
   - Type `cd` then drag your "pathe-crawler" folder into Terminal
   - Press Enter
   - Copy and paste this command:
     `pip3 install beautifulsoup4 requests`

Running the Script

1. **Prepare Your URLs**
   - Open `crawler.py` in TextEdit
   - Find the `urls = [` section near the bottom
   - Replace the example URL with your British Pathé lightbox URL(s)
   - Save the file
2. **Run the Script**
   - In Terminal (make sure you're in the pathe-crawler folder)
   - Type:
     `python3 crawler.py`
   - The script will start running and show progress messages
3. **Find Your Results**
   - When the script finishes, look for `pathe_data.csv` in your pathe-crawler folder
   - Open it with Numbers or Excel to view the collected data

Troubleshooting

If you see any error messages:

1. Make sure you're connected to the internet
2. Verify your British Pathé URLs are correct
3. Try running the script again
4. If problems persist, check the error message in Terminal and save it to share with technical support

Important Notes

- The script will create a CSV file called `pathe_data.csv` in your project folder
- Each time you run the script, it will add new videos to the existing CSV file
- The script will automatically skip videos it has already processed
- If you want to start fresh, simply delete the `pathe_data.csv` file

Need Help?

If you encounter any issues: 1. Take a screenshot of any error messages 2. Note what step you were trying to complete 3. Contact technical support with this information

Remember: Never share your login credentials or API keys with anyone!