Langchain Setup

Thursday, 27 March 2025   11:00 PM

- Pip3 install langchain
- Pip3 install langchain-openai
- Pip3 install langchain-community
- Pip3 install langchainhub (dynamic prompt)
- Pip3 install black (to reformat python code)
  - Terminal : black .
- Pip3 install python-dotenv (to load env variables into our class)
- Pip3 install langchain-ollama
- Pip3 install pprint (pretty print the json response for an API)
- Pip3 install pipenv
- Pip3 install langchain-pinecone (vector store)
- Pip3 install langchain_text_splitters (Long text are splitted into chunk)
- Pip3 install pypdf (injest pdf into python)
- Pip3 install faiss-cpu (Local vector store)
- Pip3 install streamlit (Front end for GenAI based applications)
- Pip3 install firecrawl-py (scrape a website content)
- Pip3 install pillow (Display images in streamlit using pillow)
- Pip3 install -U langchain-huggingface
- Pip3 install langchain-experimental (lang chain's not production ready code's like virtual agents)
- Pip3 install qrcode (to generate qr code for code interpreter)
- pip3 install mysql-connector-python (if you need to connect to mysql)

Installed items are in : /Users/sugnanasekaran/Library/Python/3.12/bin
Export the path using : export PATH="$HOME/Library/Python/3.12/bin:$PATH"

If only running by terminal, install in main python instead of venv by running this in terminal:
- pip install --user mysql-connector-python
- python3 -m pip install --target=/Users/sugnanasekaran/Library/Python/3.12/lib/python/site-packages mysql-connector-python

Create New venv and install requirements file:
- python -m venv crewai_fresh
- source crewai_fresh/bin/activate
- Give pwd
- Streamlit run location/.py
- Pip3 install -r /Users/sugnanasekaran/wspace_python/ticketing/requirements.txt
  - To see if a dependency is installed inside venv e.g streamlit
    - pip list | grep streamlit
    - If not found then install -> pip install streamlit

Pip3 show …… will show the version

What is Firecrawl?
Firecrawl is a web crawling and scraping tool built for GenAI applications. It's designed to easily fetch web content, especially for LLMs (like ChatGPT) to read and reason over real-time information.

When building GenAI apps like chatbots, assistants, summarizers, or question-answering tools, you often want to:
- Fetch real-world content from websites
- Summarize or extract key info
- Feed that data to an LLM (for smart answers)

Feature	Description
🌐 Crawl Websites	Fetches pages recursively from a domain
Smart Scraping	Extracts clean text (no ads/scripts/noise)

- Fetch real-world content from websites
- Summarize or extract key info
- Feed that data to an LLM (for smart answers)

Key Features:

Feature	Description
🌐 Crawl Websites	Fetches pages recursively from a domain
📄 Smart Scraping	Extracts clean text (no ads/scripts/noise)
🌊 Content Cleaning	Gives structured and clean output
🪁 API-based	Easily integrates into any GenAI pipeline
🤖 LLM-ready	Returns output you can directly send to an LLM

Example: Using Firecrawl in a GenAI App

Let's say you're building a chatbot that answers user questions about company websites.

Step-by-step:

1. User says: *"What is on Tesla's career page?"*
2. You call **Firecrawl API**:
   GET https://api.firecrawl.dev/crawl?url=https://www.tesla.com/careers
3. Firecrawl fetches, extracts and returns clean HTML/text.
4. You pass the result to your LLM (like OpenAI GPT or Claude).
5. LLM reads and summarizes:
   "Tesla is hiring in engineering, AI, manufacturing..."