

Collection&Preprocessing.py

ดึงข้อมูล + ทำความสะอาด + แปลงให้อยู่ในรูปที่พร้อมสำหรับการทำ Embedding

```
===== RESTART: D:\Agnoa\llm\Collection&Preprocessing.py =====
กำลังรันสคริปต์
pw 25 คุย
[1/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%80%E0%B8%A2%E0%B8%B7%E0%B8%88%E0%B8%AD%E0%B8%A2%E0%B8%99%E0%B8%A1%E0%B8%A2%E0%B8%B1%E0%B8%A7%E0%B8%83%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%AA%E0%B8%9A/14
[2/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%9C%E0%B8%B7%E0%B8%88%E0%B8%99%E0%B8%95%E0%B8%B2%E0%B8%A1%E0%B8%A3%E0%B8%88%E0%B8%B2%E0%B8%87%E0%B8%81%E0%B8%B2%E0%B8%A2/6
[3/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%80%E0%B8%A2%E0%B8%B5%E0%B8%88%E0%B8%A2%E0%B8%87%E0%B8%95%E0%B8%88%E0%B8%AD%E0%B8%82%E0%B8%A3%E0%B8%84%E0%B8%8B%E0%B8%A1%E0%B8%80%E0%B8%A3%E0%B8%89%E0%B8%B2/267
5
[4/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%99%E0%B8%99%E0%B8%B3%E0%B8%83%E0%B8%99%E0%B8%AD%E0%B8%B9%E0%B8%84%E0%B8%A1%E0%B8%88%E0%B8%80%E0%B8%97%E0%B8%88%E0%B8%B2%E0%B8%81%E0%B8%B1%E0%B8%99/62
[5/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%81%E0%B8%A3%E0%B8%B0%E0%B8%80%E0%B8%9E%E0%B8%B2%E0%B8%B0%E0%B8%99%E0%B8%B1%E0%B8%AA%E0%B8%AA%E0%B8%B2%E0%B8%A7%E0%B8%B0%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%AA%E0%B8%9A/267
6
[6/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search
[7/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%80%E0%B8%A2%E0%B8%B5%E0%B8%80%E0%B8%A2%E0%B8%87%E0%B8%95%E0%B8%88%E0%B8%AD%E0%B8%82%E0%B8%A3%E0%B8%84%E0%B8%8B%E0%B8%A1%E0%B8%80%E0%B8%A3%E0%B8%89%E0%B8%B2/267
5
[8/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%81%E0%B8%A3%E0%B8%B0%E0%B8%80%E0%B8%9E%E0%B8%B2%E0%B8%B0%E0%B8%99%E0%B8%B1%E0%B8%AA%E0%B8%AA%E0%B8%B2%E0%B8%A7%E0%B8%B0%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%AA%E0%B8%9A/267
4
[9/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%95%E0%B8%B4%E0%B8%94%E0%B8%80%E0%B8%A2%E0%B8%87%E0%B8%99%E0%B8%AD%E0%B8%A1%E0%B8%B1%E0%B8%80%E0%B8%99%E0%B8%95%E0%B8%99%E0%B8%9C%E0%B8%A4%E0%B8%A7%E0%B8%A2%E0%B8%99%E0%B8%B0%E0%B8%87/2644
[10/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%8B%E0%B8%A7%E0%B8%94%E0%B8%97%E0%B8%99%E0%B8%AD%E0%B8%87%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B8%88%E0%B8%B3%E0%B8%80%E0%B8%94%E0%B8%B7%E0%B8%AD%E0%B8%99/1916
[11/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/%E0%B8%81%E0%B8%9C%E0%B8%A5%E0%B8%A3%E0%B8%B4%E0%B8%A1%E0%B8%AD%E0%B8%88%E0%B8%AD%E0%B8%99/1915
[12/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/11et7Tabemr.com
[13/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%8B%E0%B8%A7%E0%B8%94%E0%B8%97%E0%B8%99%E0%B8%AD%E0%B8%87
[14/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%A7%E0%B8%B4%E0%B8%99%E0%B8%81%E0%B8%81%E0%B8%B1%E0%B8%87%E0%B8%A7%E0%B8%A5
[15/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%B0%E0%B8%B3%E0%B8%84%E0%B8%97%E0%B8%A2%E0%B8%AD
[16/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%80%E0%B8%A2%E0%B8%A3%E0%B8%89%E0%B8%B2
[17/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%84%E0%B8%AD
[18/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%84%E0%B8%82%E0%B8%89
[19/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=11et7Tabemr.com
[20/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%84%E0%B8%82%E0%B8%99%E0%B8%A2%E0%B8%B1%E0%B8%94
[21/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%A0%E0%B8%B2%E0%B8%A7%E0%B8%B0%E0%B8%81%E0%B8%82%E0%B8%A3%E0%B8%9B%E0%B8%A3%E0%B8%B1%E0%B8%9A%E0%B8%95%E0%B8%B1%E0%B8%A7%E0%B8%9C%E0%B8%B4%E0%B8%94%E0%B8%B0%E0%B8%B0%E0%B8%87/2644
[22/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%82%E0%B8%A2%E0%B8%84%E0%B8%88%E0%B8%AD%E0%B8%88%E0%B8%AD%E0%B8%A2%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%A2%E0%B8%9C
[23/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%84%E0%B8%AD%E0%B8%A2%E0%B8%AD%E0%B8%A2%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%A2%E0%B8%9C%E0%B8%9A
[24/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%A5%E0%B8%B3%E0%B8%84%E0%B8%AA%E0%B8%99%E0%B8%AD%E0%B8%B1%E0%B8%B0%E0%B8%A2%E0%B8%9C%E0%B8%A7%E0%B8%A2%E0%B8%99%E0%B8%B1%E0%B8%87%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%AA%E0%B8%9A
[25/25] ดึงข้อมูลจาก: https://www.agnoshealth.com/forums/search?query=%E0%B8%9C%E0%B8%B7%E0%B8%88%E0%B8%99%E0%B8%9C%E0%B8%84%E0%B8%A7%E0%B8%A2%E0%B8%99%E0%B8%B1%E0%B8%87%E0%B8%AD%E0%B8%B1%E0%B8%81%E0%B8%80%E0%B8%AA%E0%B8%9A%E0%B8%B2%E0%B8%82%E0%B8%A3%E0%B8%89%E0%B8%81%E0%B8%9E%E0%B8%89
ไปติดที่เขียนจริง
>>> |
```

จะได้ไฟล์ agnos_forum_posts.json

โครงสร้างไฟล์

```
[
{
  "id": "post_001",
  "title": "ปวดหัวเรื้อรัง",
  "content": "มีอาการปวดหัวบริเวณขมับซ้ายมา 3 วัน",
  "url": "https://www.agnoshealth.com/forums/post/001",
  "date": "2023-10-01"
},
{
  "id": "post_002",
  "title": "แน่นหน้าอก",
  "content": "เวลาหายใจจะรู้สึกแน่นหน้าอกและหายใจสั้น",
  "url": "https://www.agnoshealth.com/forums/post/002",
  "date": "2023-10-02"
}
```

Embedding&Indexing.py

โหลดไฟล์ agnos_forum_posts.json จาก Phase ก่อนหน้า → แปลงเป็นเวกเตอร์ → สร้างดัชนีค้นหา FAISS → เก็บ Mapping

```
===== RESTART: D:\Agno\llm\Embedding&Indexing.py =====  
WARNING:tensorflow:From C:\Python39\lib\site-packages\tf_keras\src\losses.py:2976: The name tf.losses.sparse_softmax_cross_entropy is deprecated. Please use tf.compat.v1.losses.sparse_softmax_cross_entropy instead.  
กำลังโหลดโมเดล  
กำลังสร้าง  
Batches: 0% | 0/1 [00:00<?, ?it/s] Batches: 100% | 1/1 [00:00<00:00, 1.20it/s] Batches: 100% | 1/1 [00:00<00:00, 1.14it/s]  
เสร็จสิ้น  
>>> |
```

FAISS Index File เก็บเวกเตอร์ที่แปลงแล้ว เพื่อให้ค้นหาได้เร็ว จะได้ไฟล์ vector_index/faiss_index.idx

Metadata File เก็บ Mapping ระหว่างตำแหน่งเวกเตอร์กับข้อมูลเดิม (title, content, url) จะได้ไฟล์ vector_index/metadata.json

Search_QA.py

1. โหลดไฟล์จาก

vector_index/faiss_index.idx → สำหรับค้นหาเวกเตอร์

vector_index/metadata.json → เพื่อดึงข้อมูลระบุที่จริง

2. รับคำถามจากผู้ใช้

3. แปลงคำถามเป็นเวกเตอร์

4. ค้นหากระู้ที่ใกล้เคียงที่สุด (Top-k)

5. รวมเนื้อหาที่เจอเป็น context

6. ส่ง context + คำถาม เข้า LLM (เช่น GPT-4 หรือ LLaMA)

7. แสดงคำตอบ + แหล่งข้อมูลอ้างอิง (References)

