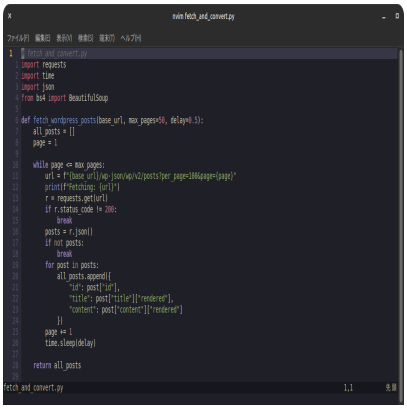

ã??ã?¼ã?¿ã?»ã??ã??ã??ä½?ã??ã?³ã?¼ã??ã•®ä¿®æ£ç??

Description

å?•å??ã•¯ã?ã•®ã??ã?ã?°ã•®è¨?äº?ã??ã•?ã•¹ã•¦å•?å¾?ã•?ã??ã??ã?ã?°ã?©ã? ã•¨å•?å¾?ã•?ã•?è¨?äº?ã•?ã??ã??ã?¼ã?¿ã?»ã??ã??ã??ä½?æ?•ã•?ã??ã??ã?ã?°ã?©ã? ã•?å?¥ã•§ã•?ã•?ã•?ã?••ã??ã??ã??ï¼?ã•¤ã•®ã??ã?ã?°ã?©ã? ã•«ã•¾ã•¨ã?•ã•?ã•®ã•?ä»¥ä¸?ã•§ã•?ã??

```
import requests
import time
import json
from bs4 import BeautifulSoup

def fetch_wordpress_posts(base_url, max_pages=50, delay=0.5):
    all_posts = []
    page = 1

    while page <= max_pages:
        url = f"{base_url}/wp-json/wp/v2/posts?per_page=100&page={page}"
        print(f"Fetching: {url}")
        r = requests.get(url)
        if r.status_code != 200:
            break
        posts = r.json()
        if not posts:
            break
        for post in posts:
            all_posts.append({
                "id": post["id"],
                "title": post["title"]["rendered"],
                "content": post["content"]["rendered"]
            })
        page += 1
        time.sleep(delay)

    return all_posts

def convert_to_qlora_format(posts):
```

```
    rec = 1
    dataset = []
    for post in posts:
        title = post["title"]
        content_html = post["content"]
        content_text = BeautifulSoup(content_html, "html.parser").get_text()
        if content_text.strip():
            dataset.append({
                "instruction": f"æ¬¡ã•®ã??ã?¼ã??ã•§æ??ç« ã??æ?¸ã•?ã•¦ã••ã• ã
•?ã•?ï¼?ã??{title}ã?•",
                "output": content_text.strip()
            })
        rec += 1
    print(f"{rec} records converted")
    return dataset

if __name__ == "__main__":
    wordpress_url = "https://zikuu.space/monologue"
    posts = fetch_wordpress_posts(wordpress_url)
    dataset = convert_to_qlora_format(posts)

    with open("qlora_dataset.json", "w", encoding="utf-8") as f:
        json.dump(dataset, f, ensure_ascii=False, indent=2)
    print("Dataset saved")
```


è¿?ã? ã?•å°•ã•?ã?•ã•®LLMã•§ã??ã?¡ã?¤ã?³ã?•ã?¥ã?¼ã??ã?³ã?°ã??ã•?ã•¦ã•¿ã?•ã??ã••ã•®ã•¨ã••ã•¯ã¾ã•?è¨?äº?ã•«ã•?ã•¾ã•?ã•ã??

ã??ã?©ã?¡ã?¼ã?¿ã?¼æ?°ã•?100å??1000å??ã??è¶ ã•?ã??ã??ã•?ã•ªå¤§ã••ã•ªè¨?èª?ã?¢ã??ã?«ã??ä¸?è?¬ã•®å??äººã??å°•ã•?ã•ªçµ?ç¹?ã•?é??ç?ºã•?ã??ã•®ã•¯ç¾æ??ç?¹ã•§ã•¯è²»ç?¨é¢ã•§é?£ã•?ã•?ã•§ã•?ã•?ã?•æ?¹æ³?è«?ã??å ·ä½?ç??ã•ªæ?¹æ³?ã??ç?¥ã•£ã•¦ã•?ã??ã•°ã?•ä½?è£?ã•®ã•?ã??çµ?ç¹?ã•«å ¥ã•£ã•?ã•¨ã••ã•«å®?è·µã•§ã••ã•¾ã•?ã??ã•ªã•®ã•§ã?•ã•?ã•®ã??ã•?ã•ªã•?ã••ã??å??äººã•?å??å¼·ã•?ã??æ?•ç¾©ã•¯ã?ã??ã•¾ã•?ã•?ã?•å°?æ¥?ã?•Alé?¢é?£ã•®ä¼•æ¥ã•«å°±è•·ã??å¸?æ??ã•?ã•¦ã•?ã??å¦ç??ã•?ã??ã••«ã•¨ã•£ã•¦ã??è?¯ã?è¨?ç·´ã•«ã•ªã??ã•®ã•§ã•¯ãªã•?ã•?ã•¨æ?•ã•£ã•¦ã•?ã•¾ã•?ã??ç•¾æ??ç?¹ã•§ã•¯ã?ã?¢ã??ã•¥ã••ã??ã?¡¾ã•?å¿?æ•´ã•§ã••ã??ã•®ã•¯ã?ã•®è¾ºã??ã•¾ã•§ã??

ã•?ã•?ã•?ã•?ã•¨ã??ã??ã•£ã•¦ã•?ã??ã•¨ã?•æ?¬å½?ã•®Alå¡¾é?·ã??ä½?ã•£ã•¦ã•¿ã•ªã••ã•ªã??ã•¾ã•?ã••ã??ã••ã•®ã•?ã?•ã•«ã•¯ã?ã??ã??ã??ã??ä½?ã??ã?ã••ã•ªã•£ã•¦ã••ã•¾ã•?ã??å•£ã??ç?®ã??è?ã??æ??ã??å??ã•?ã•?å?•ã•?ã•?ã•?ã??ã•°ä½?ã•?ã??è©±ã•?ã??é??å±?ã??ã?«ã?¼ã??ã?³ã•®ç´ æ••ã??å? å·¥ã•?ã•¦é?¨å?•ã??ä½?ã•£ã•?ã??ã?•ã?¨ã??ã?¸Alã??å??ã•?ã•?ã•?ã•®å°•ã•?ã•ªã?³ã?³ã??ã?¥ã?¼ã?¿ã?¼ã??å? è?µã•?ã?•å•?ç¨®ã?»ã?³ã?µã?¼ã??ã?¢ã?¼ã?¿ã?¼ã•?ã?¢ã?¯ã?•ã?¥ã?¨ã?¼ã?¿ã?¼ã??ä½¿ã•£ã•¦ã?•å¤?é?¨ã•?ã??ã•®å?•ã••ã•?ã•?ã??ç?¥è¦?ã•?å??ä½?ã??å®?è¡?ã•?ã??ã??å¦ã•¶ã•?ã•¨ã•?ã•?ã••ã•?ã??ã•?ã•£ã•¦é¢ç?½ã••ã•?ã•§ã•?ã??ã•?ã??ã••ã??ã•¨å•?æ??ã•«ã?•é??å±?å? å·¥ç?¨ã•®æ©?æ¢°ã??å¢?è¨ã•?ã•?ã?•ã?«ã?¼ã??ã?³ã??ç?¼ã••ç¯ã?æ¬²ã•?ã•?ã•¨ã?•é??é?•ã•ªã••è¨å??æ??è³?æ?•æ¬²ã•?æ¹§ã•?ã•¦ã••ã•¾ã•?ã??

ä½?ã•?ã??å°?ã•?ç?¥ã??ã•°ã??ã•£ã•¨ç?¥ã??ã•?ã••ã•ªã??ã??ä½?ã•?ã??ï¼?ã•¤ç?¥ã??ã•°ä»?ã•®ã•?ã•¨ã??ç?¥ã??ã•?ã••ã•ªã??ã??ç§•ã•¯ã?ã•®ã??ç?¥ã??ã•?ã•?ã?•ã•¨ã•?ã?•æ??æ? ã??è¡•å??ã??ç?¥æ?§ã•¨å?¼ã??ã•§ã•?ã??ã•®ã•§ã•?ã•?ã•?ã•?ã??ã•¯å¥½å¥?å¿?ã•¨ã•?ã•?è¨?è??ã•§è¡¨ç•¾ã•?ã•¦ã??è?¯ã?ã•§ã•

?ã??ã•?ã??

ã?¢ã??ã•¥ã••ã??å¡¾ã•®ä½¿å?½ã•¯ã?è?ªå??ã•®ä¸ã•«ã•?ã??ç?¥æ?§ã??å¥½å¥?å¿?ã??å¼?ã••å?ºã•?ã•?ã•¨ã• ã•¨æ?•ã•£ã•¦ã•?ã•¾ã•?ã??ç§•ã•®å½¹å?²ã•¯ã•ã•®æ©?æ¢°ã•¨ç?°å¢?ã??æ•?ã•?ã??ã•?ã•¨ã??

ã•?ã•ªã•?ã•¯è?ªå??ã®ç?¥æ?§ã??å¥½å¥?å¿?ã??å¼?ã••å?ºã•?æ?¥ã? ã??é?•ã•£ã•¦ã•?ã•¾ã•?ã•?ï¼?

ä»¥ä¸?ã?•ä½?è«?ã•§ã•?ã??

é•?å?»ã•«å¤§ç? ã??æ?£ã•£ã•?ã??ä°?æ? ã•«é•ã•£ã•?ã??ã•¨ã??æ»ã??ã•?ã??ã•?ã??ã•ªã•?ã•?ã•ªã?•ã•¨ã•?ã•?çµ?é¨?ã??ä½?å°¦ã•?ã•?ã•¾ã•?ã•?ã??èª°ã??ã•?ã•?ã•?ã??ã•¯æ»ã«ã•¾ã•?ã•?ã??ã?•æ»ã•¬ã•?æ»ã•ªã•ªã•?ã•?ã•§ã•¯ãªã•«ã?•ã•?ã•¤æ»ã•¬ã•?ã•©ã•?æ»ã•¬ã•?ã•?é?¢å¿?ä°?ã•§ã•?ã??ã•ã??å•è?²ã•¦ã•?çµ?ã?•ã??ã••ã??ã•?ã•®æ??æ??ã•«ã•ªã•£ã•?ã??ã?•è?ªå??ã•?ã•©ã•?æ»ã•¬ã•?ã??è??ã•?ã•¦ã??è?¯ã•?ã•?ã•ªã•¨æ?•ã•?ã•¾ã•?ã??

ã?¢ã??ã•¥ã••ã??å¡¾ã•¯æ??è¡?ã??ç?¥è?ã??å ±æ??ã?ã??äºã•?å¢?ã•?ã??ã•?ã•¨ã??æ??ã??ã•§é•?å?¶ã•?ã•¦ã•?ã??æ©?é?¢ã•§ã•?ã•?ã??ã?•ä¸?å•ç?¸ä¼•ã•®ã??ã•?ã•ªç§?å¯?ã?•ã•?ã•?ä¼•æ?¿æ?¹å¼•ã•§ã•¯ãªã••ã•?ã•¤ã?•ã•ªã??ã•¹ã•§ã•¯Alå¡¾é?·ã®ã??ã•?ã•ªã??ã?¸ã?§ã?¯ã??ã¯ã?¢ã??ã¥ã••ã??å¡¾ã•®æ??å•?æ?§ã•«å•?ã•£ã•¦ã•?ã??ã??ã•?ã•«æ?•ã•?ã•¾ã•?ã•?ã?•å??ä°ç??ã•«ã??è?ªå??ã•®æ»ã•«æ?¹ã•¨ã•?æ??é•©ã•ªã•®ã•§ã•¯ãªã•?ã•?ã•¨æ??ã•?ã•¾ã•?ã??

Category

1. æ?¥ã? ã•®æ´»å??
2. ç?¬ã??è¨?

Tags

1. AI
2. LLM
3. ã??ã?¡ã?¤ã?³ã?•ã?¥ã?¼ã??ã?³ã?°
4. ä¸?é??å??å¸?
5. å¤§è¦•æ¨¡è¨?èª?ã?¢ã??ã?«
6. å±±æ¢¨ç??

Date Created
2025å¹´6æ??27æ?¥
Author
kazuo-tsubaki