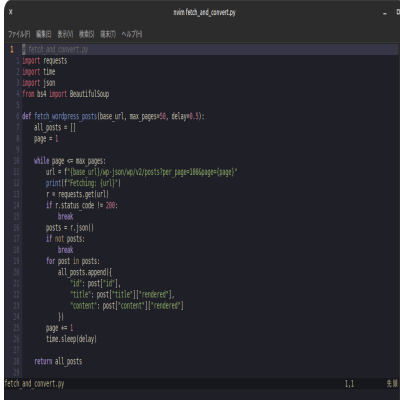




ã??ã?¼ã?¿ã?»ã??ã??ã??ã½?ã??ã?³ã?¼ã??ã•®ä¿®æ£ç??

Description

å?•å??ã-ã?ã•®ã??ã?ã°ã•®ë"?ã?ã??ã?ã-1ã-!ã-?ã¼?ã-?ã??ã??ã?ã°ã?ã-ã-?ã¼?ã-?ã-
?è"?ã?ã-?ã??ã??ã?¼ã?¿ã?»ã??ã??ã?ã½?æ?•ã-?ã??ã??ã?ã°ã?ã?ã?ã-ã-?
ã??ã?ã??¼?ã-ã-ã•®ã??ã?ã°ã?ã?ã?ã-ã-«ã-¾ã-ã-?ã-ã-ã-ã-»¥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« ă??æ?, ă•?ă•|ă••ă• ă
•?ă•?i¼?ă??{title}ă?•",
            "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?°•ã?ã?ã?®LLMã•§ã??ã?;ã?¤ã?³ã?•ã?¥ã?¼ã??ã?³ã?°ã??ã?•ã?|ã?¿ã??ã?•ã?æ?•ã?£ã?|ã?•ã?
¾ã?•ã??ã?•ã?®ã?•ã?•ã?¾ã?•è?ã°?ã?«ã?•ã?¾ã?•ã?ã??

[illegible][illegible][illegible]

???

???

???

???

???

???

Category

- ???
- ???

Tags

- AI
- LLM
- ???
- ???
- ???
- ???

Date Created

2025-06-27

Author

kazuo-tsubaki