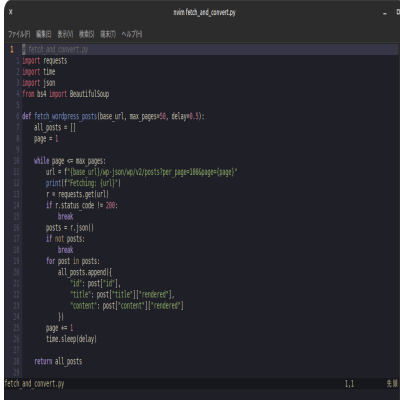




ã??ã?¼ã?¿ã?»ã??ã??ã??ã½?ã??ã³ã?¼ã??ã•®ä¿®æ£ç??

Description

å?•å??ã-ã?ã•®ã??ã?ã°ã•®ë"?ã?ã??ã?ã¹ã!ã?ã¼?ã?ã??ã??ã?ã°ã?ã•ã?ã¼?ã?ã•
?è"?ã?ã?ã?ã??ã?¼ã?¿ã?»ã??ã??ã½?æ?•ã?ã??ã??ã?ã?ã?ã?ã?ã?ã?ã?ã?ã?ã?ã?ã?
ã??ã?ã??¼?ã•ã•®ã??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•?i¼?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")

```

e¿?ã? ã?°•ã?ã?ã•®LLMã•Sã??ã?jã?¤ã?³ã?ã?¥ã?¼ã?ºã?³ã?°ã???ã?ã|ã•¿ã???ã?ã•æ?ã•£ã|ã?ã•
¾ã?ã???ã?ã•®ã•ã•ã•¾ã?è``äº?ã«ã•?ã•¾ã?ã•ã??

[illegible][illegible]

ä½?ä•ä??â°?ä•?ç?¥ä??â°ä??ä•£ä•ç?¥ä??ä•ä•ä•ä??ä??ä½?ä•ä??ä??¼?ä•□ç?¥ä??ä•°ä»?ä•®ä•?ä•
ä??ç?¥ä??ä•ä•ä•ä??ä??ç§ä•-ä?ä•®ä??ç?¥ä??ä•ä•ä•ä•ä•ä•ä•æ??æ? ä??èj•â??ä??ç?¥æ?§ä•
â°¼ä??ä•§ä•ä??ä•®ä•§ä•ä•ä??ä•?ä??ä•-â½¼â¥?â¿?ä•ä•ä•ä•è""è??ä•§èjç¾¼ä•ä!ä??è?-ä?ä•§ä•

???

???

???

???

???

???

Category

- ???
- ???

Tags

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

Date Created

2025-06-27

Author

kazuo-tsubaki