

README

1.图书馆项目1： 基于AI技术的南科大讲座海报著录与分析



➤ 背景

- ◆ 图书馆将讲座海报（PNG/JPG，中\英文混合）人工著录后收集在图书馆特色数字资源平台上。

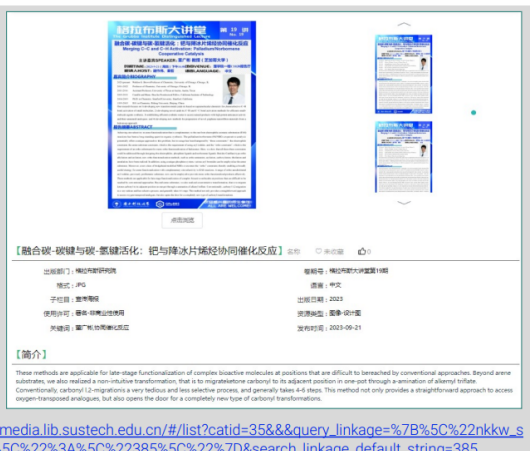
➤ 难点

- ◆ 工作量大，效率低
- ◆ 图书馆员缺乏相关专业知识，编写摘要难度大
- ◆ 人工著录易出错

➤ 现状

- ◆ 之前其他课程学生已尝试 ChatPGT/Bing 自动生成摘要。

图书馆导师：杜娟，duj3@sustech.edu.cn
王伟，wangw@sustech.edu.cn



29

Description

A project to implement automatic crawling and analyzing of lecture posters at Southern University of Science and Technology.

Contribut

| SID | Name |
|----------|------|
| 12112729 | 杨烜 |
| 12110813 | 刘圣鼎 |
| 12110817 | 张展玮 |
| 12111427 | 勾业备 |

Repository Structure

```
.  
├─ convert.py  
├─ get_info.py  
├─ get.py  
├─ update.py  
├─ visualize.py  
├─ GUI.py  
├─ README.md  
└─ stop_words_CN.txts
```

Requirements

- Python 3.9
- numpy 1.24.0
- pandas 2.0.0
- matplotlib 3.7.0
- beautifulsoup4 4.12.0
- fake_useragent 1.4.0
- tqdm 4.66.0
- paddleocr 2.7.0
- wordcloud 1.9.0
- jieba 0.42.0
- ttkbootstrap 1.10.0

Usage

```
python GUI.py
```

学术讲座海报资源管理系统

| | | | |
|---------|-----------|--------------------------------------|----------------------|
| 元数据及海报： | 起始页码： | <input type="text"/> | |
| | 终止页码： | <input type="text"/> | |
| | 语言模型： | <input type="text" value="chatGPT"/> | |
| | 元数据输出路径： | <input type="text"/> | 选择路径 |
| | 海报图片输出路径： | <input type="text"/> | 选择路径 |
| | | | 导出 |

统计分析：

| | | | |
|------|--------------|---------------------------------------|---|
| | 起始日期： | <input type="text" value="2024/1/9"/> |  |
| | 终止日期： | <input type="text" value="2024/1/9"/> |  |
| 热点词： | 元数据输入路径： | <input type="text"/> | 选择路径 |
| | 热点词列表输出路径： | <input type="text"/> | 选择路径 |
| | 词云图输出路径： | <input type="text"/> | 选择路径 |
| | 添加停用词：（, 分隔） | <input type="text"/> | |
| | | | 导出 |

Poster crawling and analyzing

Crawl the poster and generate information:

- Choose the number of posters to crawl and analyze(one page has the latest 6 posters).
- Choose the chatgpt or Wenxinyi model we provided.
- Select the output path for the poster and analysis data.
- Click "导出".

Poster statistical analysis

Generate poster word cloud data:

- Select the entry path for the poster information form.
- Select word cloud and hot word output path.
- Add stop word (if not, leave it blank)
- Click "导出".

ENGINEERING 机械设计
思维 综合 DESIGN 设计
制造 开发 CIRCUITS THINKING
回收 成型 装置
技术 系统 ANALOG TECHNIQUES
MANUFACTURING 电路系统 RAPID PROTOTYPING 环保 基础
MECHANICAL