

NVCK training 7 - NVCK raw data to text

Tuesday, December 28, 2021 2:54 PM

Ở bài này tôi sẽ giới thiệu project nvck_core

Project này chứa các file python để đọc file html và parse lấy thông tin

```
78 |
79 | # Get list all stories save to an array
80 | stories = os.listdir(input_base_url)
81 | # stories = ['con-nha-giau']
82 | logging.basicConfig(level=logging.INFO)
83 |
84 | count = 0
85 | for story in stories:
86 |     # for each story process_story and write to a file
87 |     logging.info(f'count: {count}')
88 |     try:
89 |         process_story(story, force=force, debug=debug)
90 |     except Exception as ex:
91 |         logging.error(f'process: {story}')
92 |         if debug:
93 |             traceback.print_exc()
94 |         pass
95 |     count += 1
```

Trong thư mục: G:\My Drive\PROJECT\nvck\truyenfull.vn\stories

Sẽ chứa các truyện, mỗi truyện là một thư mục

```
metajson
> co-vo-tre-con
> co-xuyen-ky-duyen
> coca-them-duong
> coi-long-co-don
> coi-sach-giai-tri
> con-dau-diem-gia
> con-duong-cam-hoa-xau-xa-cua-nu-chinh
> con-duong-dau-quy-nha-ho-tieu
> con-duong-sung-hau-sung-hau-chi-lo
> con-duong-sung-the
> con-duong-theo-duoi-nam-than
> con-gai-nha-nong
> con-khong-phai-vi-em-dang-yeu-sao
> con-loc-ky-giap
> con-mua-dinh-menh
> con-muon-lam-vo-ba
```

Trong một thư mục truyện sẽ có:

Các thư mục chương

Thông tin truyện

```

{} meta.json ●
co-vo-the-than-cua-tong-tai > {} meta.json > description
1  {
2    "version": "0.0.1",
3    "name": "Truyện",
4    "author_name": "Hạ Y Hoa",
5    "category_name": "Ngôn Tình, Ngôn Tình, Ngược, Sủng, 1",
6    "source": "wattpad.com/user/Drothyzz",
7    "status": "Full",
8    "description": "Số chương: chưa FullThể loại: Ngôn từ",
9    "truyenfull_id": "17944"
10 }

```

Danh sách chương

{} meta.json ● {} chapters.json ●

co-vo-the-than-cua-tong-tai > {} chapters.json > {} 0

```

1  [{
2    "id": "chuong-1",
3    "name": "Chương 1"
4  }, {
5    "id": "chuong-2",
6    "name": "Chương 2"
7  }, {
8    "id": "chuong-3",
9    "name": "Chương 3"
10 }, {
11   "id": "chuong-4",
12   "name": "Chương 4"
13 }, {
14   "id": "chuong-5",
15   "name": "Chương 5"
16 }, {
17   "id": "chuong-6",
18   "name": "Chương 6"
19 }, {
20   "id": "chuong-7",
21   "name": "Chương 7"
22 }, {
23   "id": "chuong-8",
24   "name": "Chương 8"
25 }, {

```

Trong một chương có: nội dung truyện và thông tin chương

▼ co-vo-the-than-cua-tong-tai

▼ chuong-1

≡ content.txt

≡ desktop.ini

{} meta.json

```
meta.json co-vo-the-than-cua-tong-tai • chapters.json • meta.json ...\chuong-1 •
co-vo-the-than-cua-tong-tai > chuong-1 > {} meta.json > ...
1  {}
2    "version": "0.0.1",
3    "name": "Cô Vợ Thế Thân Của Tổng Tài",
4    "prev": "",
5    "next": "chuong-2",
6    "truyenfull_chapter_id": "2366440"
7  }
```

```
meta.json co-vo-the-than-cua-tong-tai • chapters.json • meta.json ...\chuong-1 • content.txt X
co-vo-the-than-cua-tong-tai > chuong-1 > content.txt
1  Ngôi nhà cao to rộng lớn, được trang trí với màu nâu phối hợp với nude trong rất đẹp mắt, đi thẳng vào
2
3  "Ba à, con không muốn gả cho cái tên Tần Sở An gì đó đâu, không biết tên đó là người thế nào, lỡ tên
4
5  "Thanh Thanh con bình tĩnh đi nào, để ba nghĩ cách đã".
6
7  Cô gái ngồi lấp ló ngoài cửa bỗng dung bị một cánh tay nhéo lấy tay mà lôi vào phòng, tiếng la hét c
8
9  "Ông không cần phải nghĩ cách, tôi có cách rồi"
10
11 Người đàn ông đang nhìn ngoài cửa mặt cau mày nhí tay khoanh trước ngực, bỗng dung quay lại khi tiến
12
13 "Mẹ, mẹ có cách là cách gì, nói nhanh coi"
14
15 Người phụ nữ tay đang nhéo lấy tai của cô gái liền thả ra, xô cô gái ngã lẳng xuống sàn nhà, liền xo
16
17 "Từ từ, đừng vội, thì lấy con em con thầy thế chỗ con là được chứ gì"
18
19 "Mẹ đúng là thông minh đấy, yêu mẹ nhất"
20
21 Cô gái nghe thấy lời người phụ nữ vừa nói xong liền nói rồi nhảy đến ôm chặt người phụ nữ, cô gái đ
22
23 "Mẹ nói gì, mẹ bảo con thầy thế chỗ chị sao, không được, đó là người ta dạm hỏi chị mà"
24
25 Câu nói cô gái vừa xong thì người phụ nữ dùng dùng tức giận, nhanh bước đến gần cô gái nhéo lấy tai
26
27 "Mày còn cãi lại tao sao, tao bảo thì mày làm đi, có cần phải hỏi rề rà không hả?"
28
29 Cô gái liền bật khóc, tay cố kéo tay người phụ nữ ra khỏi tai cô nhưng mỗi lúc người phụ nữ càng nh
30
```