

پروژه اول

نوشتن خودکار انگیزه نامه

در این پروژه می‌خواهیم به کمک ورودی کاربر و اطلاعاتی که از دیتاست های دانشگاهها استخراج میکنیم، یک انگیزه نامه مخصوص رشته و دانشگاه مورد نظر کاربر بنویسیم.

در این برنامه ابتدا از کاربر دانشگاه مورد نظر و رشته ای که قصد دارد برای آن اقدام کند را میگیریم. سپس از او تعدادی کلید واژه دلخواه میگیریم. این کلیدواژه ها اطلاعات شخصی خود کاربر مانند معدل، رشته، دانشگاهی که در آن درس میخواند و ... میباشند. سپس با توجه به نام دانشگاه وارد شده و رشته وارد شده در دیتاست جست و جو میکنیم و چند کلیدواژه مرتبط به مجموع کلید واژه ها اضافه میکنیم.

در مرحله بعدی این کلید واژه ها را به www.gomoonbeam.com که یک دستیار نوشتاری هست میدهیم تا برای ما انگیزه نامه را تولید کند.

نمونه ورودی :

Standford University

Computer Science

Bachelor of Computer Engineering, Ferdowsi University of Mashhad, GPA of 4, TA of Data Mining Course, 3 years of work experience

نمونه خروجی :

انگیزه نامه ای که این کلمات در آن به صورت مستقیم یا غیر مستقیم استفاده شده باشند.

قسمت اول پروژه crawling :

crawl کردن وبسایت مورد نظر با استفاده از کتابخانه هایی مثل selenium و beautiful soap.

برای crawl سایت gomoonbeam نیاز هست تا ابتدا با جیمیل خود در آن ثبت نام کنید. سپس در قسمت wizard دسته ی student را که آدرس آن <https://app.gomoonbeam.com/headstart/template?collection=student>

هست را انتخاب میکنید و بعد از آن دسته essay را انتخاب میکنید و ادامه عملیات خزش را انجام میدهید.

- توجه شود که نمیتوانید از نرم افزار های آماده برای تولید کد crawler استفاده کنید. کد شما بایستی تمیز و خوانا باشد و اکثر exception ها را بتواند به درستی کنترل کند. برخی از exception ها هنگامی رخ میدهند که element های آن صفحه به طور کامل load نشده باشند، برای رفع این مشکل میتوانید از scroll صفحه و یا صبر کردن برای load کامل element ها استفاده کنید.

- با استفاده از وقفه های رندوم مثل sleep در پایتون، احتمال مشکوک شدن سایت مورد نظر را کاهش دهید. همچنین میتوانید هنگام نوشتن در داخل فیلد ها هم از وقفه استفاده کنید.
- برای اینکه با هربار اجرای برنامه نیاز به لاگین نداشته باشید دایرکتوری user data مرورگر خود را به عنوان تنظیمات به selenium بدهید. قطعه کد زیر را اگر با selenium کار میکنید میتوانید استفاده کنید.

```
options = webdriver.ChromeOptions()
options.add_argument("user-data-dir=<Profile path of your chrome browser>")
options.page_load_strategy = 'normal'
page =
webdriver.Chrome(executable_path=chromedriver_path,
chrome_options=options)
```

- برای پیدا کردن profile path میتوانید وارد لینک chrome://version در داخل کروم شوید.

قسمت دوم پروژه preprocessing:

- Text Preprocessing & keyword extraction

در این قسمت باید با توجه به آنچه در درس آموختید، پیش پردازش را روی دیتاست انجام دهید. مراحل زیر نمونه پیش پردازش هایی هست که باید انجام شوند. پیش پردازش های دیگر مانند stemming و یا lemmatization و غیره را نیز میتوانید در صورت نیاز انجام دهید.

1. Text cleaning
2. Stop word removal
3. Tokenization

- ستونی به اسم extracted_keywords داشته باشید و برای هر رشته کلیدواژه هایی که استخراج کردید را در آن بگذارید. سپس word cloud این ستون را رسم کنید.

بعد از پیش پردازش، باید کدی نوشته شود که با گرفتن نام دانشگاه و رشته مورد نظر (خط اول و دوم در نمونه ورودی بالا)، تعدادی از کلمات مرتبط آن رشته را به کاربر برگرداند.

در اینجا کلمات مرتبط میتوانند از دسته های زیر باشند :

نام درس های آن رشته، شهر آن دانشگاه، پیش نیاز های آن درس یا رشته، رنک دانشگاه و

یا برای مثال میتوانید کلیدواژه "affordable" را اگر رشته مورد نظر شهریه کمتر از \$5000 داشت در نظر بگیرید.

لینک های مفید برای این قسمت :

<https://www.kaggle.com/code/sudalairajkumar/getting-started-with-text-preprocessing#Introduction>

<https://www.kaggle.com/general/69712>

توضیحات دیتاست :

This dataset contains 60425 master degree programs from around the world

Data columns (total 23 columns):

1. country_name: the name of the country where the university offering the program is located.
2. country_code: the ISO 3166-1 alpha-2 country code for the country where the university offering the program is located.
3. university_name: the name of the university offering the program.
4. university_rank: the rank or rating of the university offering the program, which could be based on various criteria such as research output, teaching quality, reputation, etc.
5. program_name: the name of the master's program being offered by the university.
6. program_type: the type of program, which could be MSc, MBA or etc.
7. deadline: the deadline for submitting an application to the program.
8. duration: the duration of the program in months or years.
9. Language: the language of instruction for the program.
10. tuition_1_currency: the currency of the tuition fees for the program.
11. tuition_1_money: the amount of the tuition fees for the program, in the currency specified in the previous column.
12. tuition_1_type: the type of tuition fees, which could international, national, or some other type.
13. tuition_2_currency: the currency of an additional tuition fee (if applicable) for the program.
14. tuition_2_money: the amount of the additional tuition fee (if applicable) for the program, in the currency specified in the previous column.
15. tuition_2_type: the type of the additional tuition fee (if applicable), which could be international, national or some other type.
16. tuition_price_specification: This column contains additional information about the tuition fees, such as whether the tuition fee is for one semester or a year.
17. start_date: the start date of the program.
18. ielts_score: the minimum required IELTS score for non-native speakers of English to apply to the program.
19. Structure: This column contains information about the structure of the program, such as the name of the courses in the program.
20. academic_req: academic requirement for applying to this program.
21. facts: some information about this program and university.
22. City: the city where the program will be held.
23. program_url: the link to the program website.

موفق باشید

SOP STRUCTURE SUGGESTION

INTRO

1 paragraph

Explain your career goals and how grad school helps them. In 1-2 sentences, summarize what you have done to prepare for grad school (Part I). Summarize what you plan to do in grad school (Part II). Summarize why the program you are applying to helps you achieve your goals (Part III). Expand on each below.

I. PAST EXPERIENCE

2+ paragraphs

Describe research projects and papers (authors, title, responsibility, outcome). Conference presentations. Awards. Technical skills. Specific courses and what you learned in them. Relevant job experience or internships. Contextualize all with respect to your goals and plans. Explain why you have done them. Be Concrete!

II. FUTURE PLANS

1-2 paragraphs

Describe future areas of interest. Specific research questions. Problems you'd like to tackle and solve. Classes you'd like to take. Skills you'd like to learn. Explain how they further your goals.

III. WHY US

1-2 paragraphs

How does the program help you achieve your goals? Department research that matches yours. Specific lab you like to join. Classes you'd like to take. Professors you'd like to work with. Make sure information here is accurate!

CONCLUSION

0-1 paragraph

Optional ending, possibly reiterating the points in the intro. Alternatively explain why the program should accept you (WHY ME)