CIS 41B Final Project Proposal

Team Member: Xuechi Zhao, Yueqi Wang

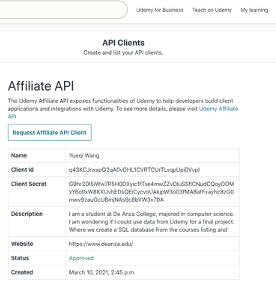
# **1.** **Project goal:**

Use Udemy developer API to query Development courses’ information, create databases and GUI for users to browse and save courses of interest.

# **2.** **Source of data:**

Udemy Affiliate API

<https://www.udemy.com/developers/affiliate/>



# 

# **3.** **Dataset information**

### 3.1 Scope of Dataset:

### - Default parameters

a. category = “Development” (“Development” including major computer science topics)

b. language = "en" (Filter courses with English only)

c. ordering = "newest" (Get the newest 1000 courses)

d. rating = "4.5" (Limit the courses with only rating higher than 4.5 out of 5)

### - Selected course subcategories:

|  |  |
| --- | --- |
| **Course Category** | **Count** |
| Web Development | 1174 |
| Programming Languages | 663 |
| Mobile Development | 360 |
| Game Development | 326 |
| Data Science | 282 |
| Software Engineering | 201 |
| Development Tools | 138 |
| Database Design & Development | 137 |
| Software Testing | 88 |
| No-Code Development | 51 |
| **Total** | **3420** |

### - Data scope of the project

Given the huge number of courses offered on Udemy, we decided to get the **1000 NEWEST highly rated** (rating>4.5) courses in the Development category and store them in a **SQL database**.

Multithreading (ThreadPoolExecutor) will be used to query a list of 1000 coursesID via “GET /api-2.0/courses/” and subsequently to fetch data for the 1000 courses, one course at a time.

### 3.2 API and data fields

- API root endpoint: https://www.udemy.com/api-2.0/

- Sample JSON file from API

- Sample 1: Return JSON file via “GET /api-2.0/courses/” see Appendix

- Sample 2: Return JSON file via Fields and Field Lists see Appendix

*Sample 2 Example url:* [*https://www.udemy.com/api-2.0/courses/1332278/?fields[course]=id,content\_info\_short,rating,num\_reviews,created*](https://www.udemy.com/api-2.0/courses/1332278/?fields%5bcourse%5d=id,content_info_short,rating,num_reviews,created)

# 4. GUI design

## 4.1 User interaction tree structure

Start: Udemy Development Courses Query System (URL)

|

|-- Choice 1: Courses Summary

| |-- Show subplots : paid/free, levels, duration, features

|

|-- Choice 2: look at courses composition

| |-- Choice 2.1: Categories: by subcategories under Development

| | |-- Show numbers of courses offered in each as pie charts (query by search)

| |

| |-- Choice 2.2: Language and Levels: by instruction languages and levels

| |-- Show side by side comparisons of courses delivered

| in different languages and various difficulty levels

|

|-- Choice 3: fetch courses by criteria

|-- Display search options:

| Entry : input search word

| OptionMenu :

| - Dropdown list of price (paid/free)

| - Dropdown list of instructional levels

| (all/beginner/intermediate/expert)

| Button : confirm choices

|

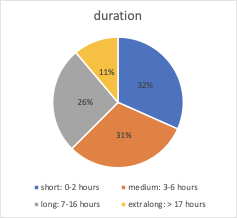
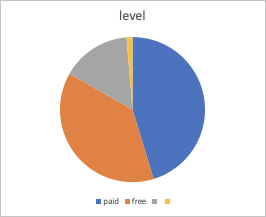
|-- Display treeview of query results with column sorts and checkboxes

|-- Open course webpage on click

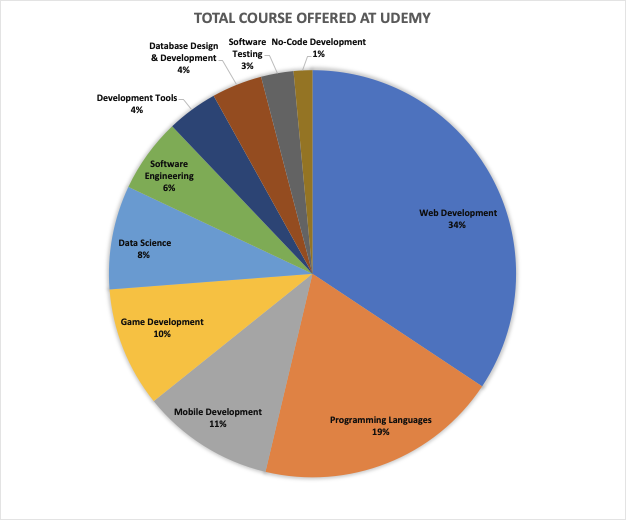
|-- Save checked courses to file

## 4.2 GUI sample outputs:

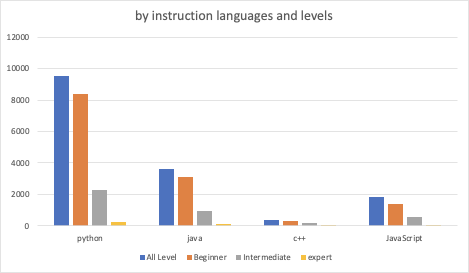
1. *Start page :*
2. *Choice 1: Courses Summary chart:*



1. *Choice 2.1: Categories chart: by subcategories under Development*



1. *Choice 2.2 Language & Levels sample chart:*



# 5. Member tasks

Backend:

Query for Choice 1 and Choice 2: Yueqi Wang

Query and build database : Xuechi Zhao

Frontend:

Plotting functions for Choice 1 & Choice 2: Yueqi Wang

Choice 3: Xuechi Zhao & Yueqi Wang

# 6. New features

1. ThreadPoolExecutor

<https://docs.python.org/3/library/concurrent.futures.html>

1. Treeview

<https://docs.python.org/3/library/tkinter.ttk.html>

1. OptionMenu

<https://docs.python.org/3/library/tkinter.tix.html>

# 

# 7. Appendix

## Sampe 1: Return JSON file via “GET /api-2.0/courses/”

{

"count": 754,

"next": "https://www.udemy.com/api-2.0/courses/?language=en&ordering=highest-rated&page=2&page\_size=12&ratings=4.0&subcategory=Data+Science",

"previous": null,

"results": [

{

"\_class": "course",

"id": 2922202,

"title": "Analysing Tweets using R",

"url": "/course/analysing-tweets-using-r/",

"is\_paid": true,

"price": "$94.99",

"price\_detail": {

"price\_string": "$94.99",

"currency\_symbol": "$",

"currency": "USD",

"amount": 94.99

},

"price\_serve\_tracking\_id": "4vuvtMcNRHK\_Cmhpw-Jhhw",

"visible\_instructors": [

{

"image\_100x100": "https://img-a.udemycdn.com/user/100x100/47581910\_9cab\_3.jpg?Gg07p8i4XPuxGvm\_nXpIkL4RIwbh9gbCnSROUvAQsfc1mnPvoOc1wWllQI1D8j0MXo9OqxpN0FCfSNXfqnAO3ZvDAxDJP6oYdrh6xBKfkOwFWINkNvIAiiHaQe4JFwo",

"display\_name": "Partha Majumdar",

"job\_title": "Just a programmer",

"initials": "PM",

"title": "Partha Majumdar",

"name": "Partha",

"image\_50x50": "https://img-a.udemycdn.com/user/50x50/47581910\_9cab\_3.jpg?IkkzBo-x494xUsST6-sC3cKZU22I3eu\_1ftkBtXsMbS5lSSwE1xzu4iQoLXmNWp-m-UINv\_WPv5aKAVF2oEZMNNEQPvA3ylOdH8MnTWe8\_d5DnNaBAfanaNPaDv7",

"url": "/user/partha-majumdar-6/",

"\_class": "user"

}

],

"image\_125\_H": "https://img-a.udemycdn.com/course/125\_H/2922202\_84ad\_6.jpg?930r8YCe7mLZ1UgVpIaXyR803mGXLD6iNnzO9mNow3WwIFcMsCARnbI2G9s4iGV0ZkVc6P4\_rqYFXYREe\_ZKGFovTxArAOGDfJHYPuPyohFg4zGg8FNaGUIgMliwPA",

"image\_240x135": "https://img-a.udemycdn.com/course/240x135/2922202\_84ad\_6.jpg?-GjP4Qxz5OlBcr2zJfb0PDJtRPCPVD\_9E8BMbd\_1MuRnk2qd-xciThuTz978rkQ1isaERZeOdAEMQsiUDfBheaL5Hv6geqCgdHoKzc\_IbhpTvR5rBDmqN2wuFilhplFR",

"is\_practice\_test\_course": false,

"image\_480x270": "https://img-a.udemycdn.com/course/480x270/2922202\_84ad\_6.jpg?eyvGOk0dB5URKr7gII8ldQnCOP3WSt-8sk3FjK6vsUwAZmU4aos-U3arnSF5B2FLQBcv4VN0Kj3E2V2OlGjv-IlkZ2CDwqHChWdciFcIu\_Jido98JRP7RfwL8XFRcsXf",

"published\_title": "analysing-tweets-using-r",

"tracking\_id": "zFJ7YzmbTiO\_oqGzZ9WjUg",

"predictive\_score": null,

"relevancy\_score": null,

"input\_features": null,

"lecture\_search\_result": null,

"curriculum\_lectures": [],

"order\_in\_results": null,

"curriculum\_items": [],

"headline": "In this Course, we go through the process of analysis of Twitter Data for Emotion Analysis.",

"instructor\_name": null

},

{

"\_class": "course",

"id": 2846378,

"title": "Master Deep Learning using Case Studies : Beginner-Advance",

"url": "/course/master-deep-learning-using-case-studies-beginner-advance/",

"is\_paid": true,

"price": "$94.99",

"price\_detail": {

"price\_string": "$94.99",

"currency\_symbol": "$",

"currency": "USD",

"amount": 94.99

},

"price\_serve\_tracking\_id": "dUE0ZLPrQnSb032zEtaIVw",

"visible\_instructors": [

{

"image\_100x100": "https://img-a.udemycdn.com/user/100x100/34310318\_731b\_4.jpg?Cf2-PlECfoaiZKswyADNmhd9jdlddoM1RJ4heMgDYUEDK2u63rsNMWskMhFwfH4CmgiH5j4BAoCz2TDQ2jWXmmW7zpc4YaGLNS423YW\_yF99aqTt5BhbBEFHv-wobns",

"display\_name": "Geekshub Pvt Ltd",

"job\_title": "BigData and Analytics",

"initials": "GP",

"title": "Geekshub Pvt Ltd",

"name": "Geekshub",

"image\_50x50": "https://img-a.udemycdn.com/user/50x50/34310318\_731b\_4.jpg?srSWi4MdYxJD8TI-xGDEqtTMF5rCQym1eF1N4Xm0uQY-4C7lJvT4AoiDJcjWmQrqDPKtRLh11g9YsmgloU7fDyhJabcAnZhwV3kbXLL-ajWa8ODAKWBabSyQWoVK",

"url": "/user/geekshub/",

"\_class": "user"

}

],

"image\_125\_H": "https://img-b.udemycdn.com/course/125\_H/2846378\_14ed\_2.jpg?secure=L1MipOkQKy69fI1LdReVfw%3D%3D%2C1615679639",

"image\_240x135": "https://img-b.udemycdn.com/course/240x135/2846378\_14ed\_2.jpg?secure=z9PscIgxB26q0Wu5Wv9c0w%3D%3D%2C1615679639",

"is\_practice\_test\_course": false,

"image\_480x270": "https://img-a.udemycdn.com/course/480x270/2846378\_14ed\_2.jpg?P-nKVVXJ4amFOo8L6OetwuFup9VJ7qJ9X2s-7rmv5ViBySoQ842KyLyNw7gNLftPf8GwerdX06ALJo5ks--x8zqVHcDAYsJPCEO-siqYhQAF7beW3-c7ReXtFLOdUFfu",

"published\_title": "master-deep-learning-using-case-studies-beginner-advance",

"tracking\_id": "DqxNOHMkSAipxS46F5f39Q",

"predictive\_score": null,

"relevancy\_score": null,

"input\_features": null,

"lecture\_search\_result": null,

"curriculum\_lectures": [],

"order\_in\_results": null,

"curriculum\_items": [],

"headline": "Master Deep Learning Algorithms Using Python From Beginner to Super Advance Level including Mathematical Insights.",

"instructor\_name": null

},

…… MORE CLASSES INFO HERE

"aggregations": [

{

"options": [

{

"count": 696,

"value": "price-paid",

"key": "price",

"title": "Paid"

},

{

"count": 58,

"value": "price-free",

"key": "price",

"title": "Free"

}

],

"id": "price",

"title": "Price"

},

{

"options": [

{

"count": 374,

"value": "all",

"key": "instructional\_level",

"title": "All Levels"

},

{

"count": 260,

"value": "beginner",

"key": "instructional\_level",

"title": "Beginner"

},

{

"count": 106,

"value": "intermediate",

"key": "instructional\_level",

"title": "Intermediate"

},

{

"count": 14,

"value": "expert",

"key": "instructional\_level",

"title": "Expert"

}

],

"id": "instructional\_level",

"title": "Level"

},

],

"search\_tracking\_id": "nNh0NP53TV-65\_hbzMI93g"

}

## Sample 2: Return JSON file via Fields and Field Lists

{

"id": 3486916,

"title": "Natural language processing with NLTK and python",

"url": "/course/natural-language-processing-with-nltk-and-python/",

"is\_paid": true,

"price": "$29.99",

"rating": 5.0,

"num\_subscribers": 8,

"num\_reviews": 1,

"primary\_subcategory": {

"id": 558,

"title": "Data Science",

},

"created": "2020-09-09T12:58:16Z",

"last\_update\_date": "2020-09-18",

"instructional\_level\_simple": "Beginner",

"content\_info\_short": "1.5 hours",

"requirements\_data": {

"items": [

"students should have some experience with Python",

"students should be able to download and install NLTK"

]

},

"what\_you\_will\_learn\_data": {

"items": [

"Natural language processing with NLTK and python",

"tokenization",

"named entity extraction",

"chunking and chinking",

"text summarization",

"parts of speech tagging",

"N-grams"

]

},

"who\_should\_attend\_data": {

"items": [

"people who are new to NLTK (Natural Language Tool Kit and natural language processing)",

"people who are new to natural language processing"

]

},

"course\_has\_labels": [

{

"\_class": "course\_has\_label",

"label": {

"title": "Natural Language Processing",

},

},

{

"\_class": "course\_has\_label",

"label": {

"title": "NLTK",

},

}

],

"description": "<p>This course is aimed at people who are new to natural language processing (NLP). The course has two main sections, the first section explains all of the main concepts such as tokenization, parts of speech tagging, named entity extraction and so on with examples and code, the second section looks at ideas like text summarization and sentiment analysis and how we can use the core concepts from part one to solve these problems.</p>",

"headline": "Learn and apply all of the fundamental concepts in natural language processing using NLTK",

}