Regex (cont.) & APIs and Data Sets

CSCI 77800 – Ethics and Computer Science

Session 5 – 9/26/2024

Edgar E. Troudt, Ph.D. <et3076@hunter.cuny.edu>

Event of Interest:
GenAI Faculty Development

CUNY Special Virtual Professional Development Series on Generative AI

AI Session 1 - Engaging with Generative AI: Which Tool Does What?

Date & Time: Sep 27, 2024 11:00 AM

Description

- “We will go back to some basics of generative AI and demonstrate several generative AI tools (free and paid versions) to compare. Participants will (re)learn what AI does, and understand functions, strengths, and weaknesses of the tools so their basic knowledge is updated.”

- Registration: https://us02web.zoom.us/meeting/register/tZMsceGtpzwjHNLdoEZWONuyaW7P1fyXm3nM#/registration
- Part II and Part III of the series on the flyer linked in Blackboard.

Interesting Document:
ACM CS 2023 Curriculum Statement

- https://csed.acm.org/final-report/
- What should a college's CS curriculum look like?
- May help you with design and alignment of a HS curriculum.

Finishing Regex

Session 5 - Part I

Vocabulary Review

- Set
- Dictionary
- List
- Tuple

- Regular Expression
- re vs. regex libraries

- grep (tool) – Linux/Ubuntu/MacOS
- findstr (Windows, somewhat analog to grep).
- **Optional reading:** https://mkyong.com/linux/grep-for-windows-findstr-example/

Regex works in Java too!

- Demo: **regex.java**, **regex2.java**

- **Optional reading, the Java Regex docs:** https://download.java.net/java/early_access/valhalla/docs/api/java.base/java/util/regex/Pattern.html

- **Optional reading, Java literal string blocks:** https://openjdk.org/jeps/378

Python works in Java too!

- Note on Jython: https://www.jython.org/.
  - Allows you to integrate Python commands into Java programs.
  - My Ubuntu setup:
    - **curl -O https://repo1.maven.org/maven2/org/python/jython-standalone/2.7.4/jython-standalone-2.7.4.jar**
    - **javac -classpath "./jython-standalone-2.7.4.jar" jython.java**
    - **java -classpath "./jython-standalone-2.7.4.jar:." jython**

Pair-wise Exercise: Names

- Modify the **regex2.java** input checker to determine if a name is valid.

- For example:
  - Edgar Troudt.
  - Edgar E. Troudt.
  - Edgar E. Troudt, Ph.D.

How did your name **RegEx** do?

- “…people with the characters é and ñ in their names are pretty common”
- “…RegEx will definitely mark those people’s names as invalid.”
- **Source:** https://medium.com/codex/your-regex-could-be-racist-and-you-dont-know-about-it-36b35341cae

- Thoughts about your code:
  - Does it take into account hyphenated names?
  - Does it take into account alphabets other than the one used by the English language?
  - What else might it be missing?

Localization

- “Localization is the process of translating an application's resources into localized versions for each culture that the application will support.”

- **Optional reading and quote source – Localization:** https://learn.microsoft.com/en-us/dotnet/core/extensions/localization
- **Optional reading – Localizability review:** https://learn.microsoft.com/en-us/dotnet/core/extensions/localizability-review

APIs and Data Sets

Session 5 - Part II

Libraries vs. APIs

- Library – collection of functions, run locally.
  - **example:** Jython.jar, gives us code that invokes Python within Java.
  - Might contain some interesting data, **Java examples**:
    - Math.E
    - Math.PI
    - Math library reference: https://docs.oracle.com/javase/8/docs/api/?java/lang/Math.html
    - Remember, it's traditional to capitalize constants.
  - Data is not usually extensive, nor dynamic.

Could we open a public database?

- Could do this via an open SQL server?
- Could we do this with a shared file/spreadsheet?
- How much control do you want over access and searching?

Could we open a public database?

- Could design a web-based API = application programmer interface.
- Web-based APIs allow “calls” to your software through web URLs.
- API developer hosts the algorithm and data and specifies how it can be manipulated.

- powerful algorithms whose code or models you don’t want to share (e.g. ChatGPT);
- algorithms that require high-end or specialized computing (e.g., GPU farm);
- specialized hardware (e.g., web cameras);
- data sets that get frequently updated, that you only want to share in a limited fashion, or after algorithmic processing;
- access control / subscription charges.

Vocabulary and Simple API Example

- API = application programmer interface.

- API key = a code that identifies either you or your program.
  - Allows either for access control, billing, or controlling unwieldy applications.
  - Forces you to register to use, rather than being wide open.

- JSON =
  - JavaScript Object Notation. A way to transmit data.

Simple Example, Food Photos

- First, load main URL:
  - https://foodish-api.com

- Second, load API URL:
  - https://foodish-api.com/api/images/burger/
  - Reloading will give several different images.
  - Passing data through the URL address.
  - Post vs. Get. This is “get”.

- Demo: **simpleFood.py**, **simpleFood2.py**.

NASA APOD – Astronomy Photo of the Day

- APOD website: https://apod.nasa.gov/apod/astropix.html
- APOD API: https://api.nasa.gov/

- **Optional reading, APOD documentation:** https://github.com/nasa/apod-api

- Demo: **api-nasaphoto.py**

NASA APOD – Astronomy Photo of the Day

- Setting up my virtual environment with appropriate libraries:
  - **python3 -m venv apidemo**
  - **source apidemo/bin/activate**
  - **pip install json5**
  - **pip install urllib3**
  - **pip install imutils**
  - **pip install opencv-python**

Images in Jupyter

- Demo: **778-s5-notebook1.ipynb**

Let’s examine some APIs

- **Pair-wise exercise:**
  - **ALL ROOMS –** Find 3 exciting data sets to share with the class.
  - **Breakout room #1:**
    - Explore public, mostly non-governmental APIs: https://github.com/public-apis/public-apis
    - Scroll down. There are a few highlighted in the first table.
    - Much further down is the “Index” with a large category list to explore.
  - **Breakout room #2:**
    - Explore https://data.gov/.
    - Choose “Data” from the top menu.
    - Feel free to explore more widely.
  - **Breakout room #3:**
    - Explore https://opendata.cityofnewyork.us/.
    - Choose “Data” on top menu. Then, “Popular Datasets”.
    - Feel free to explore more widely.
  - **ALL ROOMS – OVERTIME:**
    - If you have found three from your website, move on to another room’s website.

Regex (cont.) & APIs and Data Sets

CSCI 77800 – Ethics and Computer Science

Session 5 – 9/26/2024

Edgar E. Troudt, Ph.D. <et3076@hunter.cuny.edu>