



Lesson Title / Topic

Understanding the structure and purpose of regular expressions, using Python/Jupyter Notebooks/ ML basics.

Content Area(s)

Computer Science (Machine Learning)

Grade Level(s)

10-12

MN Science Standards

Interactions Among Science, Technology, Engineering, Mathematics, and Society

9.1.3.4.6 - Analyze the strengths and limitations of physical, conceptual, mathematical and computer models used by scientists and engineers.

Student Objectives

Upon completion of this lesson, students will be able to:

- Describe the structure and composition of regular expressions.
- Identify the applicability of regular expressions for filtering data.

Materials


- A computer with Anaconda Navigator downloaded + Jupyter Notebooks installed (prior to or during the lesson). *Example of download source/instructions:*
<https://sparkbyexamples.com/python/install-anaconda-jupyter-notebook/>)
- Any publicly available dataset available at sites such as those listed below. Select data based on course curriculum or have students select based on their interests. Download this data as a .csv file, the easiest format to upload into Jupyter Notebooks.
 - data.gov
 - data.census.gov/cedsci/
 - datasetsearch.research.google.com/

- Background information or introduction to regular expressions to be used by the teacher or given to students (*Example introduction:* https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Regular_Expressions)

Time Required

45-60 minutes

Procedures

1. Ensure each student (or pair of students working together) has a fully charged computer. If not done before class, have students install Anaconda Navigator and run Jupyter Notebooks.
2. Walk students through what regular expressions are and their purpose.
3. Have students identify and download a dataset (.csv file) with publicly available text to use.
4. Open Jupyter Notebooks and previously established regular expression code.
(*Example code:*  Lesson Plan MJAS - Sample Script.pdf)
 - a. Based on experience level/age, students can either try to substitute different phrases into the regular expression code and search in their own dataset.

OR

- b. Students can build their own rudimentary regular expressions while still having the sample code to access when/if they get stuck.
5. Have students brainstorm creative examples of their applicability to solving real-life social issues. Examples will vary depending on the dataset used. *Example:* filtering social media posts to answer questions and address social issues.
6. Have students share their findings. Discuss ways coding/programming can help address social issues.

Additional Suggestion

Editing and building regular expressions for the first time can be a confusing and difficult process. Let students know that it is okay to feel overwhelmed or frustrated and to ask for help when they need it. Give students lots of positive feedback and encouragement when it gets tough!

Credit

Stavya Arora, Maple Grove High Senior High School (Maple Grove, MN), Class of 2023, developed this lesson. Stavya's science fair project, "*Overcoming Opioids: Analyzing, characterizing, and quantifying the sentiment of Reddit posts' tone and language prior to a relapse to allow for early intervention*" went to the National Junior Science & Humanities Symposium in 2022. Stavya is a 2022-23 Minnesota Junior Academy of Science officer.