Library Segmentation Installation and Usage

Pre-Requisites

Python: v2.7.10 was used for development

Zbar-py (v1.0.4): https://github.com/zplab/zbar-py
Note: **pip** (https://pip.pypa.io/en/stable/installing/) can also be used to install zbar-py:

```
pip install zbar-py
```

OpenCV-Python (v3.3.0.10):
https://docs.opencv.org/3.3.0/df/d65/tutorial_table_of_content_introduction.html
Note: **pip** can also be used to install opencv-python:

```
pip install -I opencv-python==3.3.0.10
```

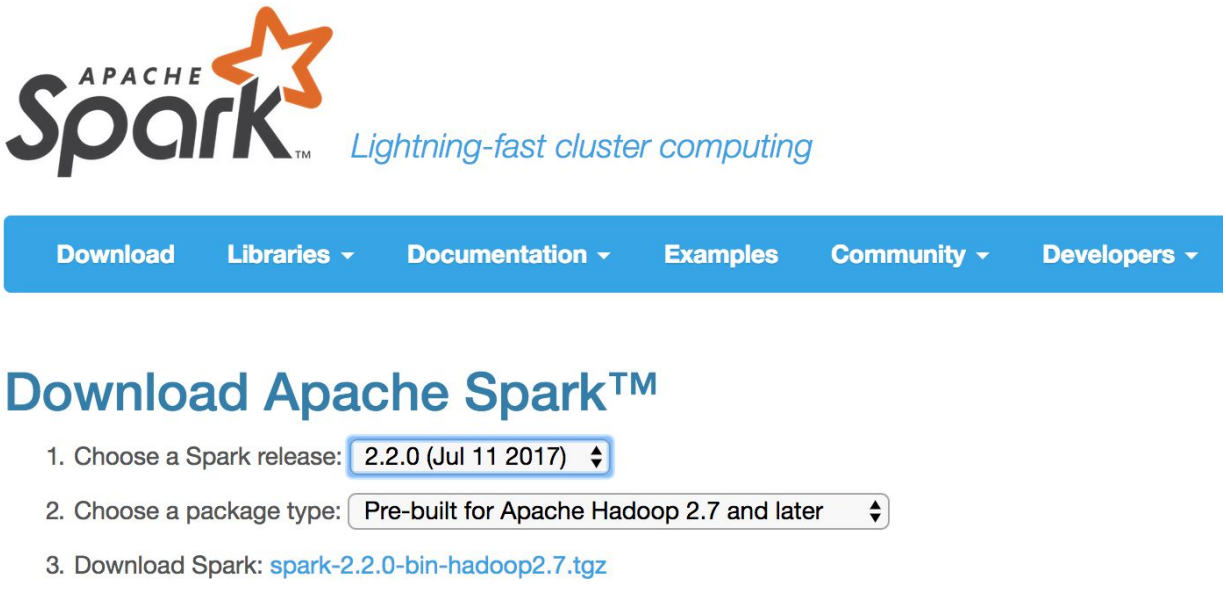
Spark (v2.2.0):

The current program only utilizes Spark standalone mode, but because it acts as a local cluster, it should work deployed to a proper clustered Spark system.

Spark Standalone Mode: http://spark.apache.org/docs/latest/spark-standalone.html

Spark standalone mode comes with the vanilla installation of Spark. Java is required and allows the Spark installation to just be downloaded: http://spark.apache.org/downloads.html

This program was written using version 2.2.0.



Note: The program doesn't utilize HDFS, so does not actually require the Hadoop installation

Program

As of 1/26/2018, the current version of the program is located on the OneDrive at Team Library/Programs/Spark

main.py and all_annotations.csv will be located in this folder and should be on disk

Running

Note: This program was being run on OSX/Unix-like OS; Windows OS will be different.
Let `$SPARK_ROOT` be the root folder of the spark download

1. Start the cluster manually:
   `$SPARK_ROOT/sbin/start-master.sh`
2. Determine the Spark URL:
   a. After starting the cluster, it will print out a log file path
   b. Print this file and find the line "`Starting Spark master at URL`" to find the Spark URL
3. Create a temporary dump folder. This is where the segments reside for processing
4. Submit the main.py job:
   `$SPARK_ROOT/bin/spark-submit --master $URL $MAIN_PATH $ANNOTATIONS_PATH $DUMP_PATH $IMAGES`
   a. $URL: The Spark URL
   b. $MAIN_PATH: The path to the main.py on disk from OneDrive
   c. $ANNOTATIONS_PATH: The pain to the all_annotations.csv on disk from OneDrive
   d. $DUMP_PATH: The path to the temporary dump folder
   e. $IMAGES: Path(s) to the Image(s) for processing
5. The current program prints segments found (>> 107 segments found) and writes to the current directory: segmentation.log and detection.log
   a. segmentation.log contains information about the segmentation with IoU
   b. detection.log contains information about the accuracy of the detection
6. To stop the cluster manually:
   `$SPARK_ROOT/sbin/stop-master.sh`