task5_030

August 30, 2024

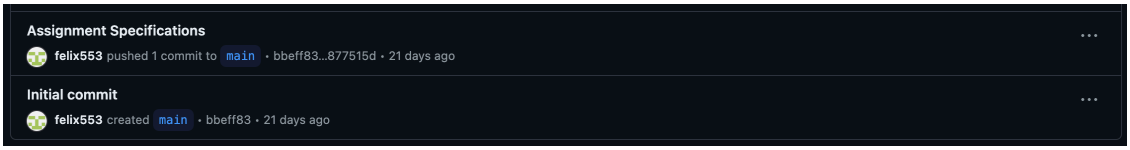
1 TASK 5 - Development History

1.0.1 Date: 09/08/2023

Today, we went through the assignment specification, and started exploring task 1 data.

contribution: Group member 1: - understood the data structure by exploring data via browsing the sample output and input

Group member 2: - understood the data structure by exploring data via browsing the sample output and input



proof

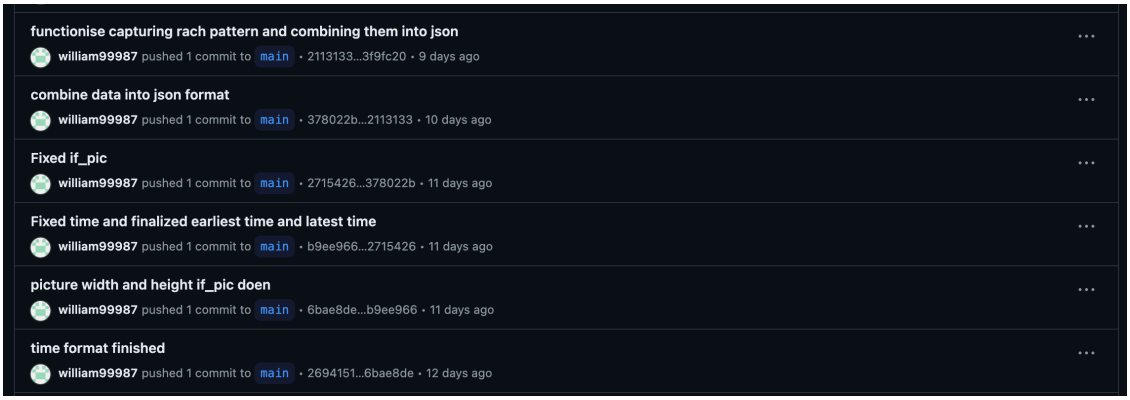
1.0.2 Date 16/08/2023

Today Michael xie implemented basic regex to extract the reviews on review text, ratings, time etc. Group member 1: - Set the structure to the whole extraction process

####Proof

1.0.3 Date 18/08/2023 - 21/08/2023

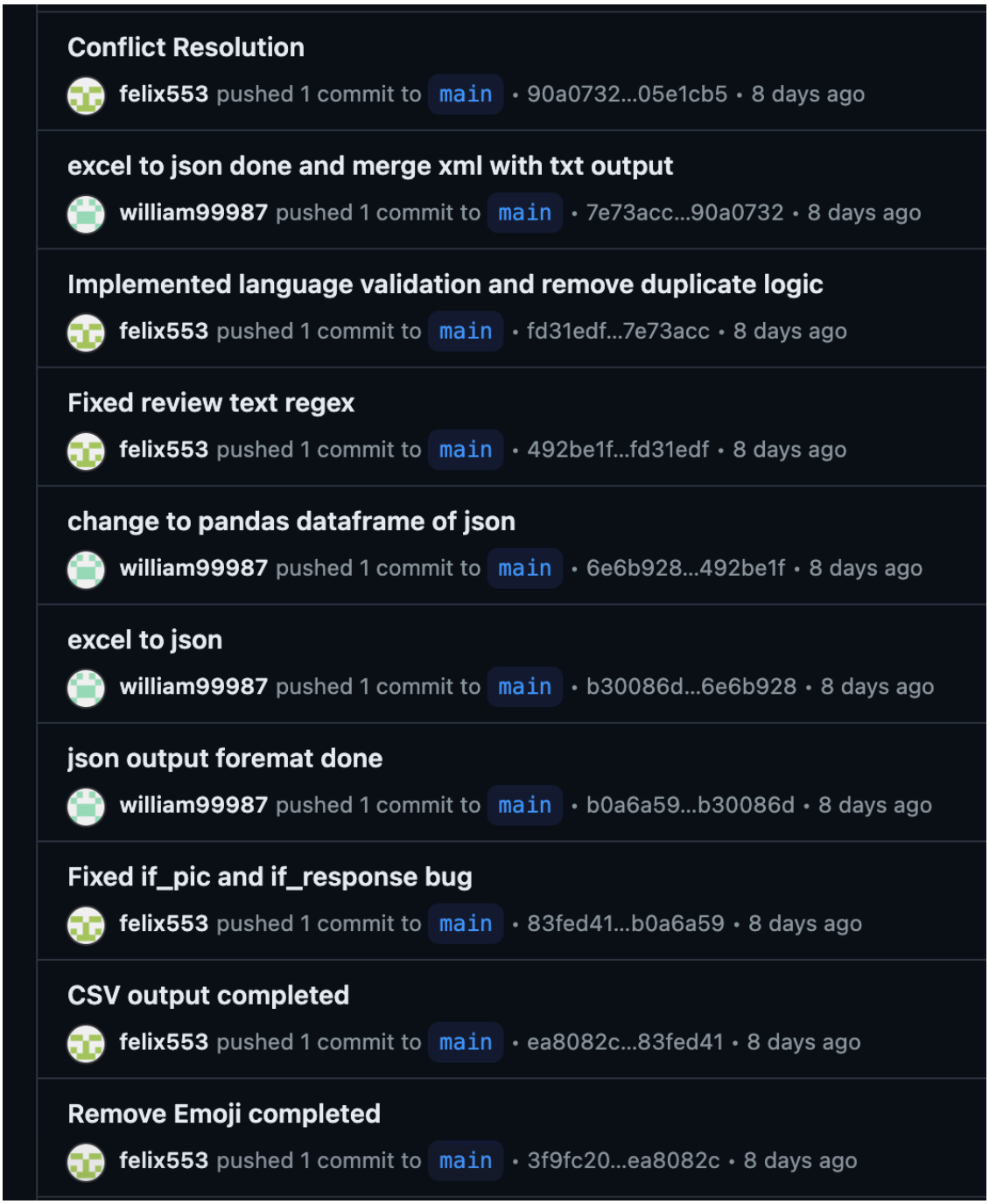
These days, Weiliang Huang further complete the exttraction on the xml file on if_pic, time etc, and functionized the extracting process. ####Contribution Group member 1: - Set extraction on if_pic, picture dimension and if_response etc.

####Proof 

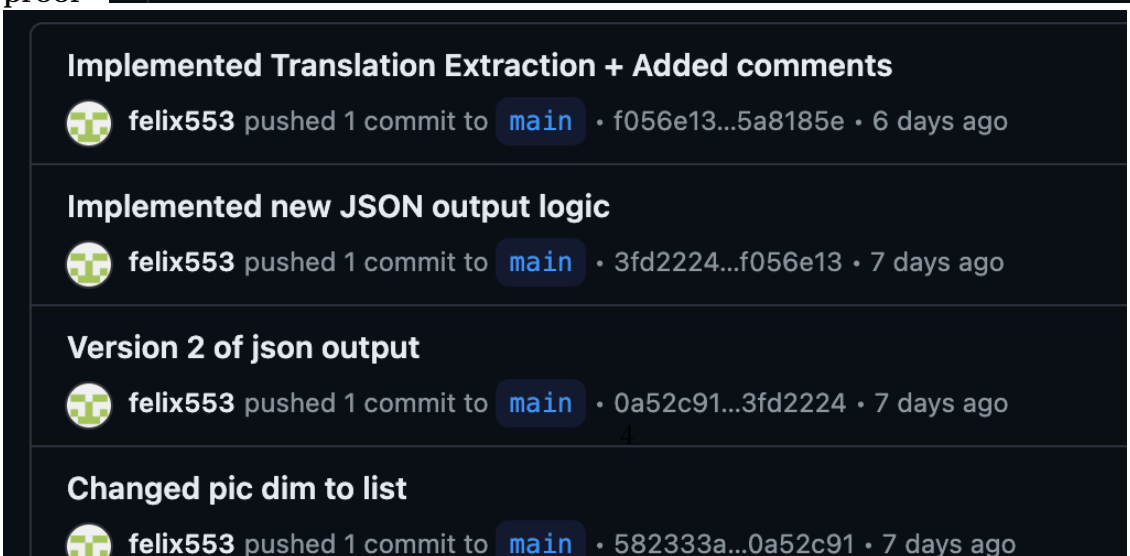
1.0.4 Date: 22/08/2023 - 24/08/2023

These two days, we finalized the task1 output

contribution: Group member 1: - write the code to produce the dataframe output of the excel file groupby the dataframe and merge it with the xml output. Group member 2: - Fixing the output format of the combinning json file, Format the task_1 notebook to make it look good, align the output with the sample ouput.


Conflict Resolution
felix553 pushed 1 commit to main • 90a0732...05e1cb5 • 8 days ago
excel to json done and merge xml with txt output
william99987 pushed 1 commit to main • 7e73acc...90a0732 • 8 days ago
Implemented language validation and remove duplicate logic
felix553 pushed 1 commit to main • fd31edf...7e73acc • 8 days ago
Fixed review text regex
felix553 pushed 1 commit to main • 492be1f...fd31edf • 8 days ago
change to pandas dataframe of json
william99987 pushed 1 commit to main • 6e6b928...492be1f • 8 days ago
excel to json
william99987 pushed 1 commit to main • b30086d...6e6b928 • 8 days ago
json output foremat done
william99987 pushed 1 commit to main • b0a6a59...b30086d • 8 days ago
Fixed if_pic and if_response bug
felix553 pushed 1 commit to main • 83fed41...b0a6a59 • 8 days ago
CSV output completed
felix553 pushed 1 commit to main • ea8082c...83fed41 • 8 days ago
Remove Emoji completed
felix553 pushed 1 commit to main • 3f9fc20...ea8082c • 8 days ago

proof


Implemented Translation Extraction + Added comments
felix553 pushed 1 commit to main • f056e13...5a8185e • 6 days ago
Implemented new JSON output logic
felix553 pushed 1 commit to main • 3fd2224...f056e13 • 7 days ago
Version 2 of json output
felix553 pushed 1 commit to main • 0a52c91...3fd2224 • 7 days ago
Changed pic dim to list
felix553 pushed 1 commit to main • 582333a...0a52c91 • 7 days ago

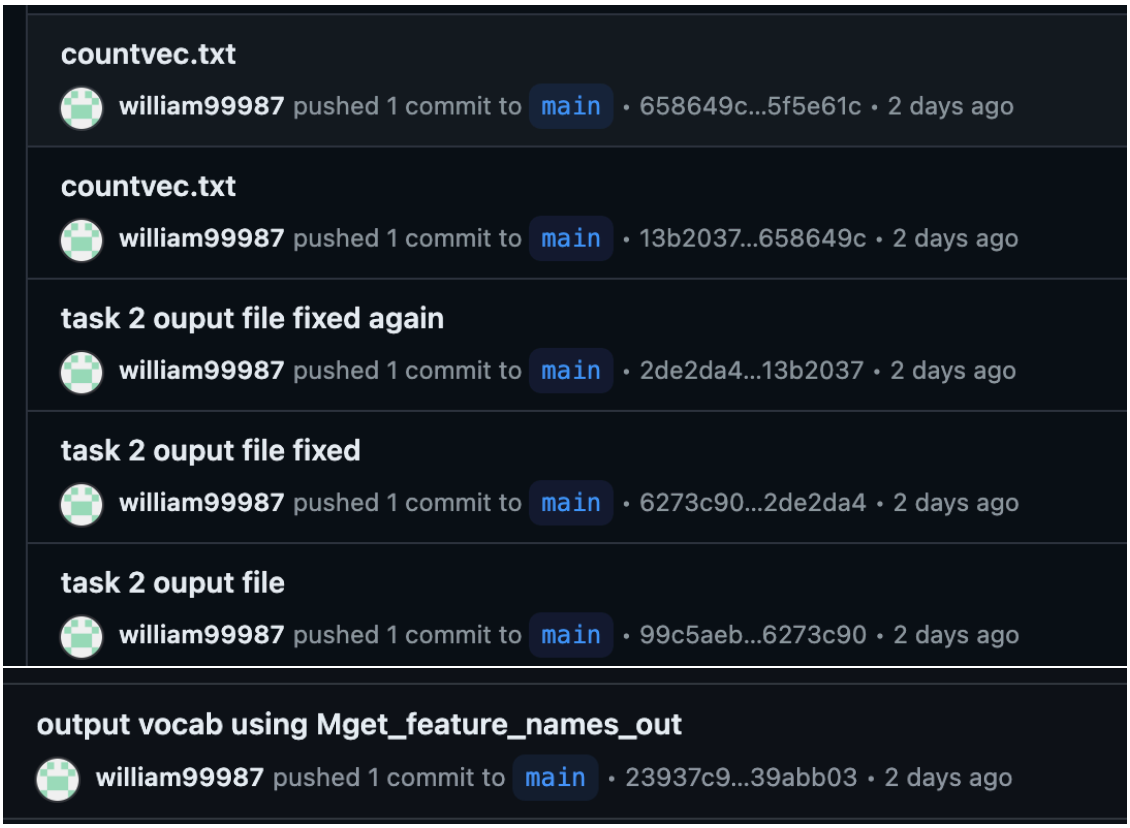
##Task 2

1.0.5 Date:26/08/2023 - 28/08/2023

These days, Weiliang Huang finished task2.

contribution: Group member 1: - Weiliang Huang finished generating the countvect and vocab txt.

proof



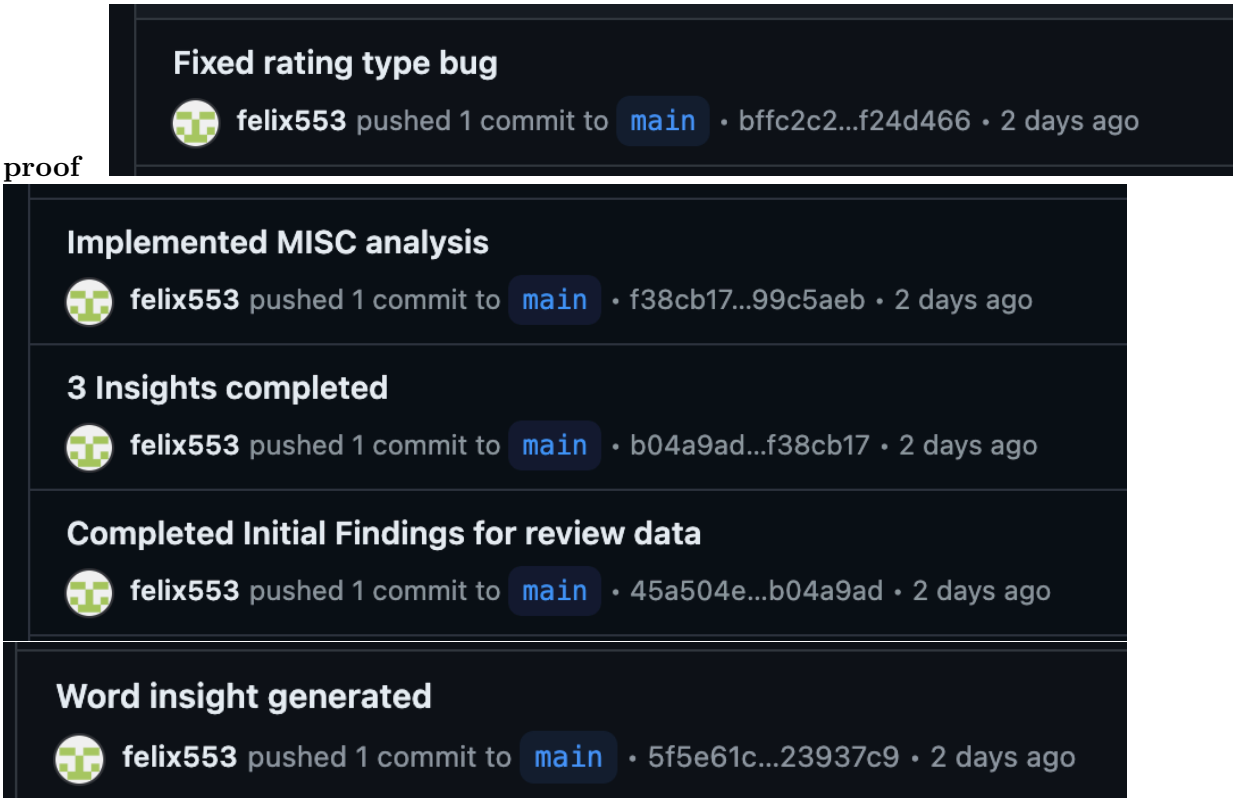
##Task 3

1.0.6 Date:26/08/2023 - 28/08/2023

These days, Michael Xie finished task3

contribution: Group member 1: - Michael xie used out put from task 1 and 2 to generate data exploration insights and charts.

proof



1.0.7 Date: 29/08/2023 - 30/08/2023

These two days, Michael Xie prepared the presentation slides and contents for the presentation. Weiliang Huang write dev history and write methodology to task1.

contribution: Group member 1: - prepared the presentation slides and contents for the presentation

Group member 2: - write dev history and write methodology for task1

task1 explaination
william99987 pushed 1 commit to main • 5a9d5f0...4c9fa5b • 1 hour ago
Completed Presentation Powerpoint
felix553 pushed 1 commit to main • 404db04...5a9d5f0 • 2 hours ago
Added some more comments Task3
felix553 pushed 1 commit to main • 893d044...404db04 • 6 hours ago
Completed final insight
felix553 pushed 1 commit to main • 5c4de3d...893d044 • 13 hours ago
Added comments for count vec
felix553 pushed 1 commit to main • cef338f...5c4de3d • 14 hours ago
Added comments for price and category
felix553 pushed 1 commit to main • ae088dd...cef338f • 17 hours ago
Added Comments for MISC and Location
felix553 pushed 1 commit to main • c06760f...ae088dd • 18 hours ago
Added comments for Dataset review stage
felix553 pushed 1 commit to main • 2639b9c...c06760f • 20 hours ago
mehodology
william99987 pushed 1 commit to main • 6313f7a...2639b9c • yesterday
Comment codes
william99987 pushed 1 commit to main • 39abb03...6313f7a • yesterday

proof