GIFKA Text Instructions

Hey Anique! My Startup is ls looking for a data specialist, for a long-term position, to parse, clean, and enrich millions of text fields. I have an Excel database with over 2 million fields. Before i can upload it to a MySQL server, I need to clean, parse, and enrich it. Doing this text manipulation with Excel or Google Sheets is risky; Information can be mistakenly deleted and the spreadsheet often crashes due to the large volume. Can you handle such a project? If so, What tools or technologies would you suggest?

For the sake of understanding, these are some examples of enrichment and parsing i need to do: 1. i have a list of brand names that are written differently although they are the same. e.g. Walgreens / Walgreen's / Walgreen's Pharmacy ... and they all should be standardized into one name 2. I have a list of websites where they begin with https: / http: / www ... and i need to standardize all of the formats 3. i have of text blocks out of which i need to find extract certain words 4. I need to link between different tables by defining a foreign key 5. i need to add new records and have all the pre-defined parsing and enrichment scripts be applied also on NEW records today i am enriching the data by using google Appscript. for example: 1. extracting the logo from a website of a certain brand 2. extracting the names of the products that are published in the website of a certain brand 3. matching between a brand and its list of stores and addresses 4. finding a Facebook or Instagram page of a certain brand provided that i have its website address  
  
Hey Anique! I would like to provide you with some additional information for you to tell me if you can provide such a service. Gifka, is a mobile app that helps users find where to redeem their gift cards. The user will upload his gift card to his digital wallet and the app will present him with a directory of brands and stores where it can be redeemed. Additionally, the app will offer a search box to find brands by search terms. The core value of the app is its well-indexed database that ties a Gift Card to a Brand to a Store to an Item. Building and managing such a database requires several steps like: Scraping, Parsing, Matching, Cleaning, Enriching, Indexing, and more. Obviously, the information in the database should be managed meticulously with extreme cautious, to ensure accuracy. Your responsibility it to build, manag and improve, a tool that gets raw data and automatically runs the text manipulation scripts to provide a clean, neat, and enriched database. In the link below you will find a sample structure of the database you will need to manage. The database contains 4 tables. 3 of the tables require text manipulation scripts. In each table begin by reading the blue-color instructions. Let me know of any questions, if such a long-term project is within your scope of expertise, and whether you are interested in joining our team. :)

[**https://docs.google.com/spreadsheets/d/12qN78ch-wDnI3pQmSnp7Mbs9beOrLWu0Xuh2Q40pX98/edit#gid=960768181**](https://docs.google.com/spreadsheets/d/12qN78ch-wDnI3pQmSnp7Mbs9beOrLWu0Xuh2Q40pX98/edit#gid=960768181)  
  
Based on the provided scope, how would you tackle such a project? What steps would you take? What technologies would you use? Coding languages? Libraries? In which of the tasks do you expect challenges? Will you need some paid tools, or will we rely on free, open-source ones? In which domains will you rely on existing libraries? In which you will write the code? Will you need to integrate with third-party solutions? How about getting AI capabilities? How about getting access to a repository of names of retail items?  
  
  
[**https://www.loom.com/share/7ec9b3d148864fb9878c4267c7cf1091?sid=a05006f4-4db4-46e3-aa23-b31e1b16064e**](https://www.loom.com/share/7ec9b3d148864fb9878c4267c7cf1091?sid=a05006f4-4db4-46e3-aa23-b31e1b16064e) [**https://www.loom.com/share/7ec9b3d148864fb9878c4267c7cf1091?sid=4714528e-6f67-4f8e-862a-e8403633d93f**](https://www.loom.com/share/7ec9b3d148864fb9878c4267c7cf1091?sid=4714528e-6f67-4f8e-862a-e8403633d93f)  
  
[**https://www.loom.com/share/7ec9b3d148864fb9878c4267c7cf1091?sid=511c72ae-84ab-451a-b1e6-5232fb4b20cd**](https://www.loom.com/share/7ec9b3d148864fb9878c4267c7cf1091?sid=511c72ae-84ab-451a-b1e6-5232fb4b20cd)  
[**https://www.loom.com/share/bdbd05fc55af4793a116ac233c698d07?sid=88564d01-c223-4825-8896-331a456c6002**](https://www.loom.com/share/bdbd05fc55af4793a116ac233c698d07?sid=88564d01-c223-4825-8896-331a456c6002)

you know what, i will first send you an exam for you to analyze and then we will have a video meeting

Hey Anique, in order to better assess your skills I would like to request you to complete an assignment to see which tasks of my project you can complete. i don't necessarily expect you to complete the entire task. Kindly complete whatever you can or provide any alternative proof to present your skills and convince us that you can cover many aspects of the project. in addition to this assignment, we have other Python scraping and scripting needs, if you have such skills please inform me, as I will see it as a major advantage for hiring you. feel free to ask me any questions to successfully complete the assignment.

Hey Anique! Any updates? Do you need my assistance? Your success is important for me

Assignment File Attached in the git repo.