Group 3

Andrew Aziz - Ebram Thabet - Mohamed Alashkar - Omar Elwaliely

Documentation for web scraping the AUC Digital library
“https://digitalcollections.aucegypt.edu/digital/collection/p15795coll20/search”

Firstly, the user should run the Python script named "Webscrap_Selenium". This script will open the library website: “https://digitalcollections.aucegypt.edu/digital/collection/p15795coll20/search”. It is very important **not** to close this website while the script is running.

For each book on this website, a new folder will be created in the directory where the Python script is being executed. The folder name will match the book's name. All images **containing faces** from the book will be downloaded into this folder as “jpg” files. The image filenames will follow the format: “BookName_PageName”.

If two images have the same name, the first image will be saved as “BookName_PageName”, and the second will be saved as “BookName_PageName (1)”, with subsequent images following this numbering pattern.

If a folder is named “Unknown_Book_Name” or an image is named “Unknown_Book_Name_Unknown_Page_Name”, it means the code could not find the book name or page name. The user may need to rerun the code to resolve this issue.

Books will be processed and downloaded in the order they appear on the digital library website. If the user stops the script and reruns it, the script will start from the beginning, iterating through the books in order. However, if it finds a folder with the same name as a book, it will skip that book and continue to the next one. This process will repeat until it reaches a book for which no corresponding folder exists. Essentially, the script resumes the downloading process from the last book it was working on before being stopped.

This code relies on the **HTML** structure of the AUC Digital Library website. If there are any changes to the website's HTML, the code may stop working or crash.

This Google Drive folder contains the **expected output** that the user should receive after running the whole code (Use AUC email to be able to access it): “https://drive.google.com/drive/folders/1NAi29f37pdWhIwIJSzQ6YNC1V4s1YS56?usp=sharing”

For the **OCR** (Optical Character Recognition) functionality, we use the library “Pytesseract” to extract text from images in each book. The OCR process specifically targets names written in capital letters or those preceded by titles such as 'Mr.', 'Mrs.', or 'Dr.'.

To use the OCR code, the user needs to specify the path to the folder containing the images by setting a variable named “folder_path”. Once the path is set, the OCR code can be executed to perform the extraction and matching.

When the user runs the OCR code, it matches the extracted names with the correct face encodings in some images (you can find these images below) from the first book in the AUC digital library, titled 'Al Hawdaj of the Campus Caravan 1958'. This process correctly associates names with their corresponding face encodings and saves this data in a CSV file named 'faces_data.csv'.

Our GitHub repository also includes a CSV file named “Webscraped_Book.csv”, which contains the names and face encodings for the below images from the book 'Al Hawdaj of the Campus Caravan 1958'. Below is a list of images that the current OCR can process and match with the corresponding face encodings:

ESTREED NASSAR ... Palestinian ... came to AUC in 1954 from the English Mission School ... thesis: "A Reassessment of Arabi Pasha" ... hobbies: drawing, painting, sculpture music and sewing ... says: "AUC has affected me greatly and I owe her more than I could ever repay" ... feels vague about her future plans ... hopes to further her studies.

Humanities-History.

EVA EL ZIK ... a graduate of the American College for Girls ... joined AUC in 1955 ... Egyptian ... helped in the Psychological Clinic at the University ... thesis: "Studies in Conversion Hysteria" ... hobbies: music, reading, touring ... plans to work for a short time then continue her studies for the M.A. degree in Psychology ... says: "I feel sorry to leave AUC in which I spent the best years of my life."

Education.

FARIDA TARAZI ... Palestinian ... a graduate of the American College for Girls ... came to AUC in 1956 ... thesis: "Finishing Cotton Fabrics for Clothing" ... hobbies: reading and trips ... was a member of the Arab Cultural Club and the Science Club ... feels sorry to leave her friends but happy to meet new ones.

Home Economics.

FAROUK ABDUL RAHIM ... Jordanian ... came to AUC in 1954 from the Preparatory Section of AUB ... was a member in the Science, Arab Cultural and Armenian Clubs ... held the office of the presidency of his class in his Junior year and was elected the President of the Student Council in his senior year ... hobbies: swimming, basketball, drawing ... thesis: "Refining of Pertoleum With Refence to the New Refinary to be Constructed in Jordan" ... feels grateful to AUC for the happy days and memories ... says: "I am very proud to have been considered as a member of AUC community".

Chemistry.

FAROUK QUDDOUMI ... Palestinian ... held the post of Treasurer of the Arab Cultural Club in 1954, President of the same club in 1956 ... took various main roles in the Arab Language Day programs, including the speaker of Souk Okaz and Kais ... was also a member of the Touring Club ... thesis: "The Causes and Effects of the 1930 Depression" ... hobbies: reading literature, playing football, swimming and boxing ... hopes to further his studies ... says: "I shall take part in the national movement arising nowadays."

Economics.

FATMA BYSHT ... Egyptian ... came to AUC in 1956 from the German School at Bab-el-Louk ... thesis: "Food Preservation in Egypt" ... hobbies: listening to soft music and reading novels ... happy to gradute, but worried about her future life.

Home Economics.

<>

FAUSTINO RORO ... says "Though I am happy to leave the university, yet I feel that I shall be missing something in my life and that is the friendly atmosphere which one finds in the campus among the students of various nationalities" ... joined AUC in 1954 coming from St. Anthony's College ... Sudanese ... was a member in the International Relations Club, the Touring Club, Square Dancing Club ... thesis: "Transfer of Power in the Sudan" ... hobbies: music, dancing and reading.

Political Science.

— SUMMER —

FUAD BAKHIT ... Saudi Arabian ... came to AUC in 1954 from Gerard Institute at Sidon, Lebanon ... thesis: "The Role of Government in Economic Life" ... was a member of the Arab Cultural Club, through which he conveyed information about the cultural and social life of Saudi Arabia ... hobbies: ping-pong, swimming, collecting stamps, reading short stories and novels ... feels sad for leaving his classmates and friends and happy to start serving his country ... plans to work in the Saudi Arabian Ministry of Finance for two years, then study for the M.A. degree in the U.S.

Economics.

GHADA EL RAFEY ... Lebanese ... came to AUC in 1955 as a junior from the American School for Girls in Tripoli, Lebanon ... thesis: "Foreign Trade Restrictions: Nature and Effect with Special Reference to Egypt" ... hobbies: collecting stamps, sewing, movies and politics ... was a member of the Music Club, Touring Club and the Arab Cultural Club ... plans to look for a job in the Lebanon and enlighten herself at the same time.

Economics.

<>

GHALIB HALASA ... Jordanian ... winner of Al Ahram Grant ... came to AUC in 1954 from the Bishop's School of Amman ... held many posts in the *Campus Caravan*, including Play Critic and Feature Editor... was a secretary of the Arab Cultural Club ... thesis: "The Colloquial Journalism of Abdulla Nadim ... hobbies: reading, smoking ... feels restless to enter practical life.

Journalism.

GILBERT DOSS ... Egyptian ... came to AUC in 1954 from St. George College ... was Secretary of both the Freshman and the Sophomore Class Committees and Vice-President in his Junior year ... thesis: "The Egyptian National Planning Commission: Its Role and Prospects" ... hobbies: journalism, touring on his scooter, and photography ... feels satisfied for the knowledge he has acquired ... hopes to further his studies in the States after completing his military service.

Economics-Business.

HAMDEYA HAMDY ... Egyptian ... was a member in the Maskers Club and a Social Manager of the Senior class ... came to AUC in 1956 from the American College for Girls ... thesis: "Realism in Youssef El Sebai Writings" ... feels that attendance in assembly should not be made obligatory ... hobbies: painting, tennis, reading and ballet.

Literature.

LAURICE SIDHOOM ... Egyptian ... joined AUC in 1954 from Menya School ... thesis: "Adjustment of the American Teachers to Selected Community Conditions" ... hobbies: singing lyrics and reading ... member of the Student Christian Association ... says: "...sociology has helped me in understanding my community."

Sociology.

LOIS GIRGIS ... Egyptian ... came to AUC in 1953 from the American College for Girls ... thesis: "The Project Method in Egyptian Government School" ... hobbies: reading, singing, and sewing ... feels that she is going to miss the advantage of being a student ... hopes that she will have the chance to complete her education abroad.

Education.

MAHMOUD EL TUKHI ... Egyptian ... came to AUC in 1955 from the Police Academy.

Education.

<>

MAHA ABD EL MONEM ... Egyptian ... came to AUC in 1956 from the American College for Girls ... thesis: "The United Nation's Emergency Force as a Measure for the Implementation of Peace".

Political Science.

<>

Mrs. MARIE GIRGIS ... Egyptian ... came to AUC in 1955 from the American Mission for Girls at Assiut ... thesis: "Birth Control and the Population Problem in Egypt" ... hobbies: playing piano, organ and accordion, singing, cycling and sports ... feels quite relieved from the studies which took much of her son's time and very happy to obtain her B.A. one month before having another child ... says: "I will probably complete my music education, then work after my children go to school - *in sha' Allah.*"

Education.

MARO BOGHIKIAN ... Egyptian ... came to AUC in 1953 from the American College for Girls ... thesis: "A Study on Non-Directive Play Therapy as a Guidance Technique" ... hoppies: music and sports ... feels sorry to leave AUC.

Psychology and Education.

<>

MADIHA NAGUIB ... joined AUC in 1956 after graduating from the American Mission for Girls at Assiut ... Egyptian ... thesis: "Conflict of Attitudes of Recent Pressly Memorial Institute and the Mothers Regarding the Role of the Girls in the Family" ... was a member of the Christian Club ... hobbies: playing piano, reading, ping pong, acting ... feels sorry to leave AUC.

Home Economics.

MAGDA TADROS ... Egyptian ... came to AUC in 1954 from the "Lycée Français" ... was an active member of Journalism, held offices from reporter to editor ... hobbies: reading philosophy books, music, painting movies ... thesis: "Misr Newsreel Before and After 1952 Revolution" ... feels the value of her achievement and realizes the wealthy and invaluable experience she gained from the University ... says: "I like my major which I have been practicing at *Al Ahram* since last November."

Journalism.

MAHMOUD BAHLOUL ... Bahreini ... came to AUC in 1955 from the Preparatory Section in AUB ... thesis: "Toward the Unity of Syria, Egypt and Jordan" ... a member of the Arab Cultural Club ... hobbies: reading political books related to the Arab states ... believes that AUC paves the way for the student to work independently and feel responsible.

Political Science.

MADIHA ABUL FUTUH ... Egyptian ... came to AUC in 1954 from the Lycée Francais, Cairo ... feels that her personality has changed at AUC to the better. ... "I am glad that I finished my studies, after which I will lead a new life," says Madiha ... hobbies: sports and reading ... thesis: "Rehabilitation of T.B. Patients."

Sociology.

MIREILLE MORDO ... Greek ... hobbies: reading and listening to music... came to AUC in 1954 from St. Claire College ... was a member in the Touring Club ... thesis: "The Non-Economic Aspects of Economic Development with Special Reference to Savings" ... feels relieved at last with a degree in hand but afraid to start life on her own ... says: "AUC has been useful in the sense that it was a world in miniature with its good and bad sides, its sad and joyful experiences"

Economics.

MOHAMMAD MEDHAT HUSSEIN ... Egyptian ... a graduate of the Lincoln School ... joined AUC in 1953 to major in Economics-Business Administration ... thesis: "Building up the Work Force in the Anglo-Egyptian Oil Field" ... was a Sports Manager in his Sophomore year ... hobbies: tennis, reading books on economics and business ... believes that the Student Council is ineffective.

Economics-Business.

NADIA EL SHAZLY ... Egyptian ... a graduate of the Mere de Dieu ... came to AUC in 1955 as a sophomore ... was a social manager of her class committee in the Junior year ... a member of the Arab Cultural Club, Music Club, and the Touring Club ... thesis: "Handicaps of the Evening Press in Egypt" ... hobbies: painting, collecting stamps, and listening to violin concerts ... says: "Although my education at AUC enables me to understand people and things better, I feel that I haven't yet come to the end of my studies and never will" ... feels restless because she has to start a new life and face new situations.

Journalism.

NADIA ARAFA ... Egyptian ... joined AUC in 1956 coming from the American College for Girls ... thesis: "The Real Motives Behind Pope's Satire" ... hobbies: painting and reading ... was a member of the Maskers ... feels that the sports tournaments were not organized, and regards the food in the cafeteria as not delicious.

Literature.

<>

NAJWA IMAM ... Palestinian ... joined AUC in 1956 coming from the American College for Girls ... thesis: "E.M. Forster's Novels in the Light of His Theory of Novels" ... plans to get married after graduation ... feels sorry to leave her friends, but happy to start a new life ... hopes that the Maskers will continue presenting its wonderful plays.

Literature.

<>

NASSER EL URI ... Palestinian ... came to AUC in 1954 from Al Ikhaa Secondary School in Beirut ... thesis: The Nature and the Working of Cinemascope ... was a member in the Science and the Arab Cultural Club ... hobbies: ping pong, photography, reading political articles, music and travel ... hopes to study nuclear engineering in the States.

Physics Mathematics.

NAWAL AL RIFAI ... Egyptian ... came to AUC in 1956 from the American College for Girls ... thesis: "Bernard Shaw and the Theatre" ... hobbies: reading, swimming, listening to Oriental music... prefers work as a secretary in any embassy ... feels the need for increasing social activities in AUC.

Literature.

NORA EL KADY ... Egyptian ... came to AUC in 1956 from the German School at Bab-el-Louk ... thesis: "Taxation and Financing of Economic Development" ... hobbies: reading novels, playing piano and listening to classical music ... feels satisfied with the knowledge she has acquired ... plans to look for a job.

Economics.

PANDELIS HALAMANDARIS ... Greek ... came to AUC in 1954 from the Greek School of Suez ... thesis: "Vocational Training as an Element in Efficiency on the Job" ... hobbies: sports, camping, reading, touring ... says: "AUC has developed in me self confidence" ... plans to work for sometime then go to the States for further studies.

Education.

RAGHEB JAOUNI ... Palestinian ... joined AUC in 1954 coming from Mubarakia School at Kuwait ... thesis: "Comarative Study Between the Attitudes of Selected Lebanese and Selected Syrian Newspapers in Regard to the Eisenhower Doctrine" ... was a permanent member of the Arab Cultural Club, Make up Editor of the *Campus Caravan* and Copy Editor in his Senior year ... feels happy to face actual life ... plans to further his studies after two years of work in Kuwait.

Journalism.

RASHAD KAMAL ... came to AUC in 1954 from Victoria College ... Egyptian ... an active member in the Maskers Club ... thesis: "Dramatization of St. Joan Theme by Anouilh and Shaw: A Study in Development" ... hobbies: painting, riding, golf, swimming ... says: "It is after four years of study that I am prepared to face the turbulent precipices and difficulties that lie ahead. My major has helped me a great deal in facing problems with a clear and open mind so that the magnitude of these oncoming are much minimized."

English Literature.

RAYMOND TADROS ... Palestinian ... came to AUC in 1954 from St. George College in Heliopolis ... thesis: "A Case Study of the Delta Trading Company with Special Reference to Integration" ... hobbies: ping-pong, reading and collecting stamps ... says: "I am happy to graduate but sorry to leave my close friends.

Economics-Business.

REGINA PROEDROU ... Greek ... came to AUC in 1954 from the Greek Gymnasium in Greece ... Thesis: "Unmet Emotional Needs and Their Relations to Some Other Variables; an Experimental Study with Special Reference to the Fifth and Sixth Grades of the Greek Patriarchal School" ... hobbies: music, collecting stamps, sewing, swimming ... feels the need for more extracurricular activities in the Faculty of Education ... hopes to receive a scholarship from the University of Pennsylvania.

Psychology.

RUQAYYA ABDEL HALIM ... Egyptian ... came to AUC in 1956 from Giza Secondary School ... thesis: A Study of the Fellah's Diet in an Egyptian Village" ... feels anxious but happy about starting a new life.

Home Economics.

<>

SABA ALARJA ... Jordanian ... was Vice-President of the Science Club in 1956, 57, 58; a Treasurer of the Student Council in 1957-58; and a member of the Arab Cultural Club ... thesis: "Enamel Industry in Egypt" ... hobbies: ping-pong, chess, and hunting ... plans to study mechanical engineering in the States ... feels that AUC has taught him self-independence.

Chemistry.

SUHAIL SABANEGH ... believes AUC is the most suitable place for developing personality ... joined AUC in 1954 coming from Bishop's School at Amman ... was a member of the Touring Club and the Arab Cultural Club... hobbies: trips, reading, painting and gardening ... feels sorry for leaving AUC but happy for being relieved of exhausting academic work.

Economics-Business.

SALIMA SHAMI ... Jordanian ... came to AUC in 1955 from C.M.S. at Amman ... thesis: "Helping Problem Children Help Themselves Through Play Therapy" ... hobbies: swimming and shooting ... was a member of the Touring Club and in her Senior year the Social Manager of the Arab Cultural Club ... feels happy for obtaining the B.A. degree, which will pave the way for her contribution in raising the standard of the Jordanian woman.

Sociology.

SALWA MOHAREB ... Egyptian ... came to AUC in 1954 from St. Mary's English School ... helped in producing and directing many plays ... thesis: "Translation of Tawfik El Hakeem Play: Departure From Paradise with Study of the Authors Characteristics" ... feels the need for more cooperation among AUC students ... hobbies: writing short stories, painting and music.

Literature.

<>

SAMI KHAWAJA ... Jordanian ... came to AUC in 1953 from the Friends' School at Ramallah ... was Secretary of the Science Club in 1956-57, member of the Arab Cultural Club, Vice-President of the Student Council in his Junior year, and a Student Advisor in 1957-58 ... hobbies: reading, movies, ping-pong, chess, music ... thesis: "The Possibility of Manufacturing Phosphate Fertilizers in Jordan" ... believes that AUC provides each graduate with the means through which he can reach the end... says: "I believe that AUC provided me with a golden frame, but I have to put the right picture in that frame."

Chemistry.

SAMI MANNA ... came to AUC in 1955 as a freshman ... completed his secondary education in Acca, Palestine ... was a Vice-President of the Senior Class Committee and a member in the Student Council... thesis: "The Role of Public Investment Under a Mixed Free Economy" ... hobbies: tennis, swimming, and all kinds of music ... feels happy to graduate in seven semesters ... determined to further his education even if he had to quit his job in the Arab League.

Economics-Business.

SAMIA HANNA ... Egyptian ... came to AUC in 1954 from the English School at Cairo ... thesis: "Air-Conditioning for Human Comfort", ... was a member of the Science Club and the Touring Club ... hobbies: swimming, tennis, and reading ... feels relieved of hard work and regrets leaving AUC.

Physics-Mathematics.

SAMIHA BIBAWY ... Egyptian ... came to AUC ip 1956 from the American College for Girls ...

Home Economics.

<>

SIHAM MITRI ... Lebanese ... came to AUC in 1956 from Tripoli College ... thesis: "Utilization of Solar Energy with Special Reference to Egypt" ... was a member in the Arab Cultural Club and the Science Club ... hobbies: reading science articles, politics, listening to music and movies ... feels sad to leave AUC but happy to graduate.

Physics-Mathematics.

SAMIRA MOSAAD ... Egyptian ... came to AUC in 1956 from the American College for Girls ... hobbies: music, dramatic stories and ping pong... thesis: .A Comparative Study of Radio Programs in Normal and Emergency Situations ... feels satisfied for fulfilling her desire.

Sociology.

SAMIRA GHOBRIAL ... Egyptian ... came to AUC in 1954 from Sanieh Secondary School in Cairo ... was a Feature Editor in her Junior and Senior years ... a member in the Arab Cultural Association and the Library Club ... played the role of Laila in "Magnoon Laila" presented in Arab Language Day in 1955, 56 ... is glad to graduate but feels sorry for leaving her friends.

Journalism.

SAMIRA KIROLLOS ... Egyptian ... joined AUC in 1954 coming from St. Mary's English School ... was Social Manager of the Maskers Club in her Junior year and Vice President of the same club in her Senior year ... thesis: Masques in Shakespeare's Plays" ... hobbies: ballet, violin, and acting ... plans to work for an M.A. in English Literature.

Literature.