

Đại học Quốc gia Thành phố Hồ Chí Minh
Trường Đại học Bách Khoa
Khoa Khoa học và Kỹ thuật Máy tính



LẬP TRÌNH NÂNG CAO (CO2039)

Bài Tập Lớn

XÂY DỰNG APP TRA CỨU SAO KÊ

GVHD: Lê Đình Thuận
Lớp-Nhóm: L01-503
SVTH: Võ Trần Phi Phong
MSSV: 2420006



Mục lục

1	Nội Dung Đề Tài	2
2	Phương Pháp Tiếp Cận	2
2.1	Ngôn Ngữ	2
2.2	Thư Viện	2
3	Cấu trúc dữ liệu	3
4	Thuật toán	4
5	Thiết lập môi trường	5
6	Demo ứng dụng	6
7	Các công việc cần làm trong tương lai	12

1 Nội Dung Đề Tài

Xây dựng một web/app tra cứu dữ liệu sao kê của Mặt trận Tổ quốc Việt Nam. Ứng dụng cần có giao diện (web/app/CLI) để tra cứu thông tin sao kê theo: số tiền, tên người gửi, nội dung.

2 Phương Pháp Tiếp Cận

2.1 Ngôn Ngữ

Python rất phù hợp để tìm kiếm dữ liệu trong file CSV. Số lượng bản ghi là một lượng dữ liệu tương đối đáng kể, nhưng các thư viện của Python có thể xử lý một cách hiệu quả.

Một số thư viện Python hỗ trợ đọc và tìm kiếm dữ liệu từ file CSV có thể kể đến như:

1. CSV: là một thư viện built-in của Python, làm việc với dữ liệu CSV thô mà không cần tải toàn bộ tập dữ liệu vào bộ nhớ bằng cách đọc dữ liệu theo từng dòng. Phương pháp này sử dụng ít bộ nhớ hơn nhưng có thể chậm hơn nếu tìm kiếm nhiều điều kiện hoặc thực hiện các truy vấn phức tạp.
2. Pandas: là một thư viện tuyệt vời để làm việc với các tập dữ liệu lớn. Nó có thể đọc nhanh các tệp CSV vào DataFrame, cho phép thao tác và tìm kiếm dữ liệu nhanh chóng. Pandas có thể xử lý dữ liệu với hàng triệu hàng và hỗ trợ các giải pháp tối ưu hóa như sử dụng `dtype` để có thể cải thiện hiệu suất tìm kiếm.
3. SQLite: để có chức năng tìm kiếm nâng cao hơn và tối ưu hóa, tải dữ liệu CSV vào cơ sở dữ liệu SQLite. SQLite nhẹ và cho phép truy vấn SQL nhanh.

2.2 Thư Viện

Trong bối cảnh của đề tài, việc tra cứu dữ liệu cần phải được thực hiện một cách nhanh chóng, hiệu quả. Thư viện Pandas là một lựa chọn tối ưu.

- Load dữ liệu: file CSV được load vào một biến toàn cục `data_df` sử dụng `pandas.read_csv`, mã hóa UTF-8 và xử lý Byte-order Mark (BOM) (`utf-8-sig`)

OS module cung cấp một cách thức để tương tác với hệ điều hành, chẳng hạn như đọc đường dẫn file, thư mục và các chức năng khác ở cấp độ hệ điều hành.

- `os.path.dirname` và `os.path.abspath`: tìm đường dẫn tuyệt đối của thư mục hiện tại và phục vụ tệp `index.html` một cách chính xác.

Flask là một web framework nhẹ được sử dụng để xây dựng các ứng dụng web. Trong đề tài này, Flask được sử dụng để tạo các API endpoint, xử lý các HTTP request, xử lý dữ liệu từ file CSV và trả kết quả cho người dùng. Flask đóng vai trò như một “backbone” của máy chủ web, quản lý các yêu cầu từ người dùng và trả về các phản hồi phù hợp.

- **Flask**: core class, được sử dụng để khởi tạo ứng dụng
- **request**: xử lý dữ liệu HTTP request (ví dụ: tham số truy vấn)

- `jsonify`: chuyển đổi các đối tượng Python (như dictionary) thành kiểu dữ liệu JSON
- `send_from_directory`: phục vụ các tệp tĩnh (static), trong trường hợp này là `index.html`

3 Cấu trúc dữ liệu

Một file CSV (Comma Separated Values) là tệp văn bản chứa dữ liệu theo định dạng có cấu trúc, trong đó mỗi giá trị được phân cách bằng dấu phẩy. Mỗi dòng biểu diễn một bản ghi và mỗi giá trị trong dòng tương ứng với một trường trong bản ghi.

Ví dụ data từ file CSV: `date_time,trans_no,credit,debit,detail` là các header.

```
date_time ,trans_no ,credit ,debit ,detail
03/09/2024_5216.65140 ,9 ,200000 ,0 ,120167.030924.100642.NGUYEN
    HOAI NAM ung ho
03/09/2024_5017.43849 ,13 ,9000 ,0 ,888828.030924.121121.NGUYEN THI
    KIEU OANH chuyen tien
04/09/2024_5240.80637 ,55 ,500000 ,0 ,MBVCB.6944619812.chung tay
    gop suc .CT tu 0351000801710 TRAN DANH VU toi 0011001932418
    MAT TRAN TO QUOC VN - BAN CUU TRO TW
```

- Pandas cung cấp một phương pháp đơn giản để đọc và thao tác trên file CSV: `pandas.read_csv`.
- Người dùng gửi tham số tìm kiếm trong `index.html`. Flask xử lý các tham số này, áp dụng logic lọc bằng Pandas và trả về kết quả dưới dạng JSON.
- JavaScript sẽ lấy phản hồi JSON và xây dựng bảng HTML một cách động để hiển thị kết quả đã lọc.

Dữ liệu được trả về:

```
{
  "date_time": "03/09/2024_5216.65140",
  "trans_no": "9",
  "credit": "200000",
  "debit": "0",
  "detail": "120167.030924.100642.NGUYEN HOAI NAM ung ho"
},
{
  "date_time": "03/09/2024_5017.43849",
  "trans_no": "13",
  "credit": "9000",
  "debit": "0",
  "detail": "888828.030924.121121.NGUYEN THI KIEU OANH chuyen tien"
},
{
  "date_time": "04/09/2024_5240.80637",
  "trans_no": "55",
  "credit": "500000",
```

```
"debit": "0",  
"detail": "MBVCB.6944619812.chung tay gop suc .CT tu  
0351000801710 TRAN DANH VU toi 0011001932418 MAT TRAN TO  
QUOC VN - BAN CUU TRO TW"  
}
```

4 Thuật toán

Thuật toán xử lý việc lọc DataFrame (`data_df`) bằng Pandas. Thuật toán này áp dụng động các điều kiện lọc trên dữ liệu và trả về kết quả đã lọc dưới dạng danh sách dictionary.

1. Xác thực đầu vào: hàm này kiểm tra xem có phải `data_df` là `None`: Nếu `data_df` là `None`, điều đó có nghĩa là tập dữ liệu chưa được tải và hàm sẽ ngay lập tức trả về một danh sách rỗng (`[]`).
2. Tạo một bản sao của DataFrame: một bản sao của `data_df` được tạo và lưu vào `filtered_df`. Điều này đảm bảo rằng tập dữ liệu gốc không bị thay đổi khi các quá trình lọc dữ liệu được thực hiện.
3. Lặp lại các điều kiện lọc:
 - `credit`: phạm vi số hoặc một số khớp chính xác
 - `date_time`: chuỗi chứa kết quả khớp - kiểm tra xem `value` có phải là chuỗi con của `date_time` hay không bằng cách sử dụng `str.contains`. `na=False` đảm bảo rằng các hàng có NaN trong cột sẽ bị loại
 - `trans_no`: chuỗi chứa kết quả khớp - đảm bảo `value` là kiểu dữ liệu số
 - Các cột khác: so sánh chuỗi không phân biệt chữ hoa, chữ thường bằng cách sử dụng `str.containswith` và `case=False`. Ví dụ: "Python" khớp với "python" hoặc "PYTHON"

```
# Filter data using pandas  
def filter_data(conditions):  
    global data_df  
  
    if data_df is None:  
        return []  
  
    filtered_df = data_df.copy()  
  
    for column, value in conditions.items():  
        if column == "credit":  
            if isinstance(value, tuple): # Range query  
                min_value, max_value = value  
                filtered_df = filtered_df[(filtered_df["credit"] >=  
                    min_value) & (filtered_df["credit"] <= max_value)]  
            else: # Single value  
                filtered_df = filtered_df[filtered_df["credit"] ==  
                    value]
```

```
elif column == "date_time":
    filtered_df = filtered_df[filtered_df["date_time"].str.
        contains(value, na=False)]
elif column == "trans_no":
    try:
        value = int(value) # Ensure the input is numeric
        filtered_df = filtered_df[filtered_df["trans_no"] ==
            value]
    except ValueError:
        # If value is not a valid number, skip filtering this
        column
        continue
    else:
        filtered_df = filtered_df[filtered_df[column].str.
            contains(value, na=False, case=False)]

return filtered_df.to_dict(orient="records")
```

Độ phức tạp của thuật toán:

- Đọc dữ liệu: $O(n)$ - đọc n dòng trong file CSV vào bộ nhớ.
- Lọc dữ liệu:
 - Lọc dữ liệu số: $O(m)$, trong đó m là số dòng
 - Tìm kiếm chuỗi: $O(m \times l)$, trong đó l là độ dài trung bình của chuỗi

5 Thiết lập môi trường

Docker là một nền tảng cho phép các nhà phát triển dễ dàng xây dựng, triển khai và chạy các ứng dụng bằng cách sử dụng các container. Các container có tính chất nhẹ, di động và tự cung cấp bao gồm mọi thứ cần thiết để chạy một phần mềm: mã nguồn, thời gian chạy, thư viện và các phụ thuộc (dependencies). Docker cung cấp một cách thức để đóng gói và phân phối các ứng dụng trong các môi trường biệt lập này (container), đảm bảo rằng chúng chạy nhất quán trên các máy hoặc hệ thống khác nhau.

Trong thư mục chứa dự án, tiến hành mở terminal:

1. Xây dựng Docker image: điều hướng đến thư mục chứa Dockerfile và docker-compose.yml, sau đó chạy: `docker-compose build`
2. Chạy ứng dụng: sau khi image được xây dựng, hãy khởi động ứng dụng: `docker-compose up`
3. Lệnh này sẽ khởi động ứng dụng Flask trên cổng 5000. Mở trình duyệt web hoặc sử dụng `curl` để truy cập ứng dụng tại `http://localhost:5000`.
4. Để kết thúc: nhấn tổ hợp phím CTRL+C trong terminal hoặc sử dụng: `docker-compose down`

6 Demo ứng dụng

Ứng dụng có các chức năng cơ bản như:

- Xác thực input, đảm bảo các giá trị nhập vào phù hợp cho việc tìm kiếm
- Tìm kiếm thông tin theo ngày, tên người gửi, số tiền, số hóa đơn
- Sắp xếp giá trị tăng dần/giảm dần theo cột
- Tính tổng số tiền dự trên kết quả tìm kiếm

Search Data

Date and Time	Transaction Number
<input type="text" value="Enter date"/>	<input type="text" value="Enter transaction number"/>
Credit Amount	Debit Amount
<input type="text" value="Enter credit amount"/>	<input type="text" value="Enter debit amount"/>
Transaction Detail	
<input type="text" value="Enter transaction detail"/>	
<input type="button" value="Search"/>	

Hình 1: Giao diện chính của ứng dụng

localhost:5000 says
Please fill in at least one field before searching.

Date and Time	Transaction Number
<input type="text" value="Enter date"/>	<input type="text" value="Enter transaction number"/>
Credit Amount	Debit Amount
<input type="text" value="Enter credit amount"/>	<input type="text" value="Enter debit amount"/>
Transaction Detail	
<input type="text" value="Enter transaction detail"/>	
<input type="button" value="Search"/>	

Hình 2: Người dùng phải nhập ít nhất một input để tìm kiếm



Search Data

Date and Time

09/09

Date must be in DD/MM/YYYY format.

Credit Amount

-50000

Credit must be a positive number.

Transaction Number

Enter transaction number

Debit Amount

Enter debit amount

Transaction Detail

Enter transaction detail

Search

Hình 3: Xác thực thông tin nhập vào



Search Data

Date and Time

09/09/2024

Transaction Number

Enter transaction number

Credit Amount

Enter credit amount

Debit Amount

Enter debit amount

Transaction Detail

Enter transaction detail

Search

6279 record(s) found.

Calculate Total Credit

Total Credit: 7,868,744,495

Credit Amount↑	Date and Time↑	Debit Amount↑	Transaction Detail↑	Transaction Number↑
55555	09/09/2024_5161.90230	0	PARTNER.DIRECT_DEBITS_VCB.MSE.6637 7041397.20240908.66377041397-0988549441_ Mong moi dieu tot dep a	687
200000	09/09/2024_5389.79977	0	020097042209082241272024U3UA760455.7997 7.224128.TRAN THI HANH TRINH ho tro ba con khác phục bao YAGI	688
2000000	09/09/2024_5242.92630	0	MBVCB.6979460863.PHAM THI LIEN chuyen tien.CT tu 0351000813059 PHAM THI LIEN toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	689
200000	09/09/2024_5241.92651	0	MBVCB.6979453135.Ung ho khác phục hau qua bao so 3 Yagi.CT tu 1024149794 KHUONG NGOC BICH toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	690
100000	09/09/2024_5017.66459	0	304088.080924.224230.HA THI HAI chuyen tien ung ho khác phục bao lu	691
100000	09/09/2024_5244.92702	0	MBVCB.6979454212.PHAM XUAN TRUNG ung ho khác phục bao so 3.CT tu 1019885185 PHAM XUAN TRUNG toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	692
2000000	09/09/2024_5219.67051	0	622852.080924.224307.HA THU HIEN UNG HO KHAC PHUC HAU QUA BAO SO 3 FT24253830067707	693
500000	09/09/2024_5189.81963	0	020097042209082243282024JU0W958288.8196 3.224329.DO THI Y PHUONG xin duoc gui de cung moi nguoi khác phục sau bao yagi	694
100000	09/09/2024_5017.68139	0	623580.080924.224331.Cuu tro bao so 3 FT24253997709200	695

Hình 4: Tìm kiếm tất cả giao dịch trong ngày 09/09/2024 và tính tổng số tiền



Search Data

Date and Time

09/09/2024

Transaction Number

Enter transaction number

Credit Amount

Enter credit amount

Debit Amount

Enter debit amount

Transaction Detail

Enter transaction detail

Search

6279 record(s) found.

Calculate Total Credit

Credit Amount↓	Date and Time↓	Debit Amount↓	Transaction Detail↓	Transaction Number↓
300000000	09/09/2024_5390.69958	0	020097042209091712522024VHQZ187912.6995 8.171254.Cong Dong Team Chau Phi ung ho khac phuc bao so 3 Yagi	2544
200000000	09/09/2024_5078.34288	0	MBVCB.6985552087.MINH MINH NGUYEN UNG HO KHAC PHUC HAU QUA BAO SO 3 YAGI VA LU QUET SAT LO TAI SAPA LAO CAI.CT tu 1048453609 NGUYEN TUYET MINH toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	2646
100000000	09/09/2024_5214.54756	0	823932.090924.001511.LE THI VIET ANH chuyen tien den MAT TRAN TO QUOC VN - BAN CUU TRO TW - 0011001932418	832
100000000	09/09/2024_5136.69938	0	IBBIZ6019569938.3BROTHERS MEDIA UNG HO KHAC PHUC HAU QUA BAO SO 3 YAGI	1143
100000000	09/09/2024_5219.99824	0	411689.090924.120601.Ancungbatuyet dong gop khac phuc hau qua bao lu	1531
100000000	09/09/2024_5078.81580	0	MBVCB.6984087181.PHAM THU HIEN UNG HO KHAC PHUC HAU QUA BAO SO 3 YAGI.CT tu 1047198390 PHAM THU HIEN toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	2129
100000000	09/09/2024_5214.63663	0	152818.090924.164305.HUELLEYROSE - CUU TRO KHAC PHUC BAO SO 3-090924-	2423

Hình 5: Sắp xếp theo thứ tự giảm dần số tiền gửi



Search Data

Date and Time

Enter date

Transaction Number

Enter transaction number

Credit Amount

50000

Debit Amount

Enter debit amount

Transaction Detail

NGUYEN THI THU TRANG

Search

10 record(s) found.

Calculate Total Credit

Total Credit: 500,000

Credit Amount↓	Date and Time↓	Debit Amount↓	Transaction Detail↓	Transaction Number↓
50000	10/09/2024_5217.22533	0	771005.100924.150724.NGUYEN THI THU TRANG Chuyen tien	76851
50000	10/09/2024_5220.71861	0	820652.100924.151906.NGUYEN THI THU TRANG chuyen FT24254680801845	78836
50000	10/09/2024_5241.22391	0	MBVCB.6993910987.NGUYEN THI THU TRANG UNG HO DONG BAO.CT tu 1016842354 NGUYEN THI THU TRANG toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	86526
50000	10/09/2024_5017.40473	0	327182.100924.172042.NGUYEN THI THU TRANG chuyen khoan	99538
50000	10/09/2024_5239.79987	0	MBVCB.6995675445.NGUYEN THI THU TRANG chuyen tien.CT tu 0731000928251 NGUYEN THI THU TRANG toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	108968
50000	10/09/2024_5390.69633	0	020097042209101952472024K4RM444172.6963 3.195248.NGUYEN THI THU TRANG chuyen tien	135900
50000	10/09/2024_5218.96892	0	332350.100924.201226.NGUYEN THI THU TRANG ung ho khac phuc bao FT24254764025466	142421
50000	10/09/2024_5240.30354	0	MBVCB.6997255888.NGUYEN THI THU TRANG chuyen tien.CT tu 1039488032 NGUYEN THI THU TRANG toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	143421
50000	10/09/2024_5189.77423	0	02009704220910202447202456TW517634.7742 3.202448.NGUYEN THI THU TRANG chuyen tien	146764
50000	10/09/2024_5216.14257	0	668703.100924.213024.NGUYEN THI THU TRANG ung ho dong bao sau bao yagi FT24255556078905	171893

Hình 6: Tìm kiếm theo tên người gửi NGUYEN THI THU TRANG và số tiền 50000



Search Data

Date and Time

Enter date

Transaction Number

Enter transaction number

Credit Amount

50000-100000

Debit Amount

Enter debit amount

Transaction Detail

NGUYEN THI THU TRANG

Search

51 record(s) found.

Calculate Total Credit

Credit Amount	Date and Time	Debit Amount	Transaction Detail	Transaction Number
100000	10/09/2024_5214.61263	0	575660.090924.224958.NGUYEN THI THU TRANG xin quyên góp chia sẻ dòng hạnh dot bao Yagi và lu lút mong mọi người được bình an	6974
100000	10/09/2024_5212.87026	0	371515.090924.231015.Nguyen Thi Thu Trang ủng hộ khắc phục bao Yagi	7694
100000	10/09/2024_5242.53189	0	MBVCB.6988439781.Ủng hộ khắc phục hậu quả bao Yagi.CT từ 1021525599 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	10520
100000	10/09/2024_5240.54325	0	MBVCB.6988484506.YTT ủng hộ khắc phục hậu quả Bao số 3 Yagi.CT từ 9938829536 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	11517
100000	10/09/2024_5244.54349	0	MBVCB.6988484978.YNY ủng hộ khắc phục hậu quả Bao số 3 Yagi.CT từ 9938829536 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	11542
100000	10/09/2024_5240.55404	0	MBVCB.6988536976.Mong mọi người đều bình an .CT từ 1042208218 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	12505
100000	10/09/2024_5241.56177	0	MBVCB.6988567005.YTT ủng hộ khắc phục hậu quả Bao Yagi.CT từ 0181003437838 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	13124
100000	10/09/2024_5078.38818	0	MBVCB.6988626611.NGUYEN THI THU TRANG chuyển tiền ủng hộ đóng bao lu lút miền bắc.CT từ 1018518134 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	13818
75000	10/09/2024_5189.24343	0	020097042209100818092024LHV9692002.24343 .081810.NGUYEN THI THU TRANG chuyển tiền	17791
100000	10/09/2024_5189.38660	0	020097042209100827452024I52B819713.38660. 082737.NGUYEN THI THU TRANG chuyển tiền ủng hộ	18321
70000	10/09/2024_5239.34197	0	MBVCB.6991191760.NGUYEN THI THU TRANG chuyển tiền.CT từ 1036219816 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	37212
100000	10/09/2024_5218.8916	0	986413.100924.115609.NGUYEN THI THU TRANG Chuyển tiền	41406
100000	10/09/2024_5244.66035	0	MBVCB.6992180586.NGUYEN THI THU TRANG chuyển tiền ủng hộ đóng bao các tỉnh bị thiệt hại.CT từ 1032282623 NGUYEN THI THU TRANG tới 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	55478

Hình 7: Tìm kiếm theo tên người gửi NGUYEN THI THU TRANG và số tiền từ 50000-100000



Search Data

Date and Time

Transaction Number

Credit Amount

Debit Amount

Transaction Detail

Search

1 record(s) found.

Calculate Total Credit

Credit Amount↑	Date and Time↑	Debit Amount↑	Transaction Detail↑	Transaction Number↑
100000	10/09/2024_5243.68002	0	MBVCB.6992242464.chung tay dong gop cung dong bao to quoc chut nho.CT tu 0381000415593 NGUYEN THI THU TRANG toi 0011001932418 MAT TRAN TO QUOC VN - BAN CUU TRO TW	57000

Hình 8: Tìm kiếm theo số hóa đơn 57000

7 Các công việc cần làm trong tương lai

Với những hạn chế hiện tại của cửa đề tài, các công việc cần làm trong tương lai bao gồm:

- Tránh sao chép toàn bộ: thay vì sao chép toàn bộ DataFrame, áp dụng lọc tại chỗ (in-place) để giảm mức sử dụng bộ nhớ.
- So sánh chuỗi hiệu quả hơn: pre-indexing hoặc các thư viện như **pytables** cho các tập dữ liệu lớn có thao tác chuỗi thường xuyên.
- Phân đoạn: đối với các tập dữ liệu rất lớn, cần xử lý theo khối với Pandas hoặc các framework phân tán như **Dask** hoặc **PySpark**.

Những phương án tối ưu hóa này có thể giảm đáng kể thời gian chạy và chi phí bộ nhớ cho dữ liệu quy mô lớn.



Tài liệu tham khảo

- [1] W3Schools. (n.d.) *Pandas Tutorial*. Retrieved December 2024, from: <https://www.w3schools.com/python/pandas/>.
- [2] GeeksforGeeks. (n.d.) *Pandas Read CSV in Python*. Retrieved December 2024, from: https://www.geeksforgeeks.org/python-read-csv-using-pandas-read_csv/.
- [3] GeeksforGeeks. (n.d.) *Pandas DataFrame*. Retrieved December 2024, from: <https://www.geeksforgeeks.org/python-pandas-dataframe/>.