

BÁO CÁO BÀI TẬP NHÓM THỰC HÀNH 1B

Nhóm 6

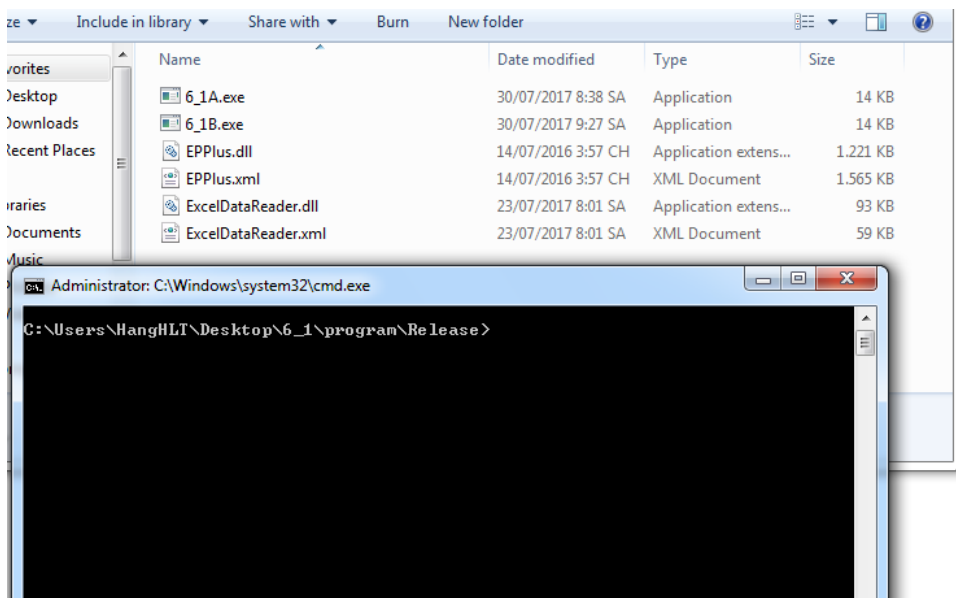
MSSV	Tên - công việc	Đóng góp
1642004	Trần Chí Bảo - Xóa mẫu rỗng	25%
1642008	Trần Lệ Diễm Châu - Xóa trùng lặp	25%
1642035	Nguyễn Thành Lai - Chuyển về diện tích km	25%
1642088	Bùi Thanh Vân - Import file text Export Excel tận dụng chương trình 1A	25%

Cách sử dụng:

Muốn sử dụng cần phải mở thư mục chứa phần mềm :

Name	Date modified	Type	Size
6_1A.exe	30/07/2017 8:38 SA	Application	14 KB
6_1B.exe	30/07/2017 9:27 SA	Application	14 KB
EPPlus.dll	14/07/2016 3:57 CH	Application extens...	1.221 KB
EPPlus.xml	14/07/2016 3:57 CH	XML Document	1.565 KB
ExcelDataReader.dll	23/07/2017 8:01 SA	Application extens...	93 KB
ExcelDataReader.xml	23/07/2017 8:01 SA	XML Document	59 KB

Mở commandline vào thư mục hiện tại chứa file thực thi exe:



Cú pháp sử dụng các chức năng như sau:

Trước tiên để thống nhất để dễ sử dụng các loại chỉ thị

[đường dẫn tới file input]: Là đường dẫn tới file input countries.txt

Vd: D:\countries.txt

-> Nên là được dẫn tuyệt đối như trên

-> Nếu file ngay cạnh thư mục thực hi thì chỉ cần tên file: vd: countries.txt

Phải đúng file countries.txt do giáo viên cung cấp.

[đường dẫn tới file output]: Là đường dẫn muốn xuất ra kèm tên của nó

Vd: D:\output.xlsx

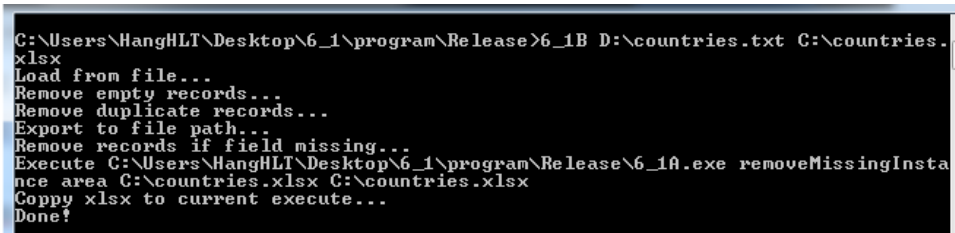
-> Nên là đường dẫn tuyệt đối như trên

-> Nếu file ngay cạnh thư mục thực thi thì chỉ cần tên file: vd: output.xlsx

Để thực hiện chức năng trên chỉ cần sử dụng câu lệnh sau:

6_1B [đường dẫn tới file input] [đường dẫn tới file output]

Vd:



```
C:\Users\HangHLT\Desktop\6_1\program\Release>6_1B D:\countries.txt C:\countries.xlsx
Load from file...
Remove empty records...
Remove duplicate records...
Export to file path...
Remove records if field missing...
Execute C:\Users\HangHLT\Desktop\6_1\program\Release\6_1a.exe removeMissingInstance area C:\countries.xlsx C:\countries.xlsx
Copy xlsx to current execute...
Done!
```

Vd:

6_1b D:\countries.txt C:\countries.xlsx

Để tiện kiểm tra file export phần mềm xuất ra file output ở cả [đường dẫn tới file output] và cả nơi chứa file thực thi

Và kết quả:

country	name	longName	foundingD	population	capital	largestCity	area
14	Abkhazia	Republic of	33044	242862	Sukhumi		5381,646
15	Abkhazia	Republic of	33110	242862	Sukhumi		5381,646
16	Abkhazia	Republic of	39686	242862	Sukhumi		5381,646
17	Abyei	Abyei Area	38361		Abyei (town)		6553,249
18	Abyei	Abyei Area	39691		Abyei (town)		6553,249
31	Adélie Land	Adélie Land					432000
36	Adjara	Autonomous Republic		393700	Batumi		1802,465
38	Aerica	Aerican Em	31905		Montreal		9000000
41	Afghanistan	Islamic Rep	7171	32564342	Kabul	Kabul	405276,3
42	Afghanistan	font-size:8	7171	32564342	Kabul	Kabul	405276,3
44			33392	1,05E+09		Nigeria	29865860
45			33392	1,05E+09		Lagos	29865860
63	Akrotiri and	Sovereign Base Areas		7700	Episkopi Cantonment		157,7157
64	Akrotiri and	of		7700	Episkopi Cantonment		157,7157
65	Akrotiri and	Akrotiri and Dhekelia		7700	Episkopi Cantonment		157,7157
68	Åland	Åland Island	7433	28666	Mariehamn		981,6998
73		Republic of	14342	2893005	Tirana	Tirana	17863,72

Lưu ý: Chương trình cần các dll cần thiết nên chỉ nó thể hoạt động khi bên cạnh nó có kèm các file cần thiết, và vì sử dụng chương trình 1A nên file 6_1A.exe cũng cần có trong cùng thư mục. Nếu không có sẽ bỏ qua bước tái sử dụng câu 1A

Đề xuất và phân tích hai giải pháp để có được dữ liệu đối với các thông tin bị thiếu (ví dụ: mẫu đầu tiên trong tập countries.txt thiếu thông tin diện tích, dân số...):

+ Ta có thể cho các dữ liệu thiếu thông tin như diện tích dân số = 0 và sử dụng binning chia giỏ theo chiều sâu -> sử dụng trung bình trong giỏ để gán giá trị

+ Tính trung bình toàn bộ field diện tích rồi gán những record thiếu giá trị bằng giá trị tính được

Đề xuất và phân tích hai giải pháp để kiểm chứng các thông tin trong tập dữ liệu là chính xác (ví dụ: thông tin về thành phố lớn nhất của các quốc gia có đúng chưa).

+ Tên nước phải tồn tại trong bản đồ thế giới, bởi vì một nước không thể nào lạc ra khỏi bản đồ thế giới vì thế ta có thể trích xuất dữ liệu danh sách tên các quốc gia trên một trang uy tín và sử dụng lọc nếu tên quốc gia này có thể nằm trong danh sách lấy được (không nhất thiết phải giống hoàn toàn chứa cũng được)

+ Thông tin về thủ đô của các quốc gia có đúng hay chưa, cũng vẫn như trên nhưng ta lấy danh sách tên quốc gia kèm thủ đô, sau đó kiểm tra records nào thỏa danh sách trên sẽ dữ lại.