

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆ TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

Tt (13)

**Extract information on the website by**

**Regular Expression**



**Giáo viên hướng dẫn: Huỳnh Thái Học và Đỗ Hà Phương**

**Thành viên nhóm :**

**1.Phạm Trọng Thi – 15055961 (Nhóm trưởng)**

**2.Quan Văn Sang – 15042721**

**3.Nguyễn Ngọc Vỹ –15056811**

**4.Trần Nhật Lâm – 15036101**

**PHẦN I: TỔNG QUAN**

**1) Mục tiêu** **của đồ án:** là tạo ra một phần mềm có thể lấy dữ liệu xuống máy tính từ website, những dữ liệu có thể lấy được bằng biểu thức chính quy như là email, số điện thoại, hình ảnh, ...

**2) Ứng dụng thực tế của phần mềm:** Phần mềm sẽ nhanh chóng lấy được dữ liệu mong muốn mà không cần đọc hết những bình luận của khách hàng, giúp cho những người chủ quản lý trang web bán hàng có thể lấy thông tin khách hàng như là :

**2.1)** Số điện thoại

**2.2)** Email

**2.3)** Hình ảnh

**3) Một vài ví dụ ứng dụng của phần mềm:** Phần mềm rất nhiều tiện ích trong thực tế, nhóm sẽ sơ lược một vài ứng dụng như sau:

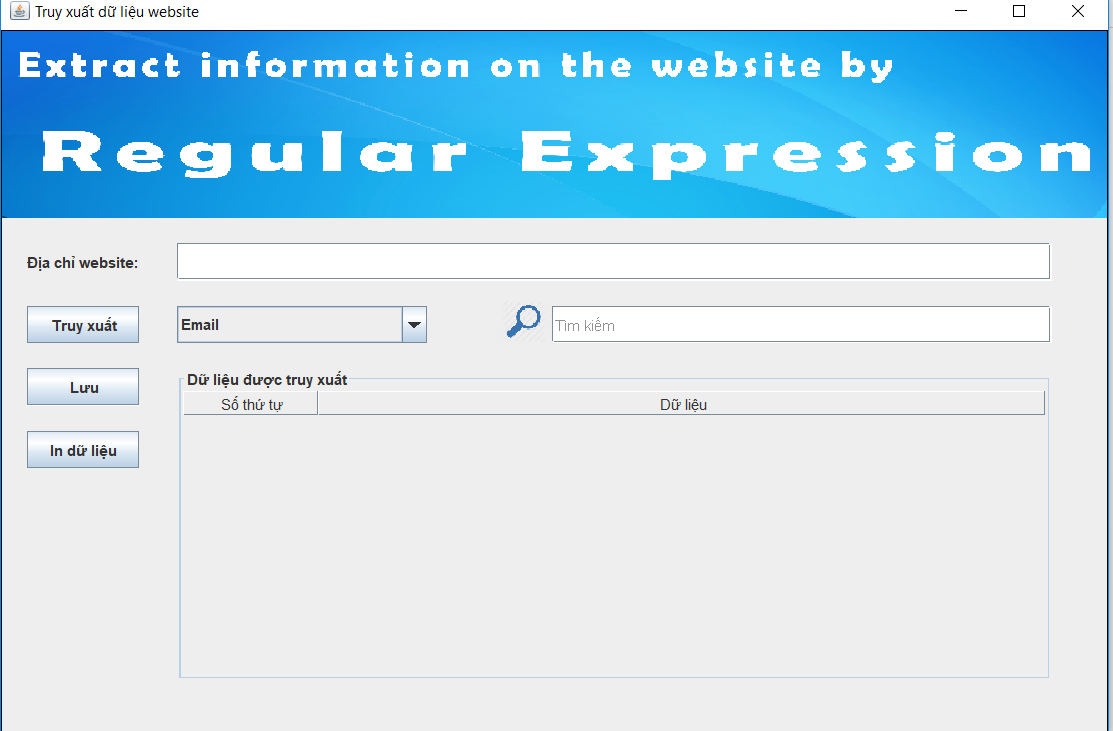
**3.1)** Một người thầy hay giảng viện,hoặc sinh viên muốn chia sẻ sách, tài liệu, hoặc code, thì họ chỉ cần thông báo, nếu có ai cần thì để lại Email, và họ chỉ cần dùng phần mềm này truy xuất tất cả Email của những ai đã bình luận ,và họ chỉ việc lấy Email mà không cần đọc bình luận của ai cả.

**3.2)** Một trang web bán bánh sinh nhật, khác hàng chỉ cần để lại số điện thoại và hình ảnh mẫu bánh mà khách hàng mong muốn, như vậy việc dùng phần mềm để lấy số điện thoại và hình ảnh là điều rất cần thiết, làm tăng năng suất làm việc và có độ chinh xác cao.

\**Hiện tại nhóm đã hoàn thành xong đồ án cũng đúng tiến độ đã đề ra.Tuy nhóm đã cố gắng nhưng cũng có những mặt hạng chế,mong thầy bỏ qua. Và thầy là người đã hướng dẫn, giúp chúng em hoàn thành đề tài đúng tiến độ.Xin chân thành cảm ơn thầy.*

**PHẦN II: CHI TIẾT VỀ PHẦN MỀM**

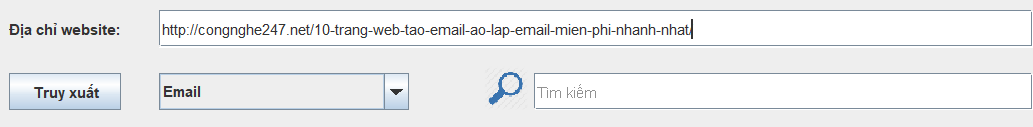
**1.Giao diện của phần mềm:**



Đoạn code của giao diện trên:



**2.Phần kết nối phần mềm với website.**

*a) Ta copy địa chỉ website cần lấy thông tin và bỏ vào textfield****.***

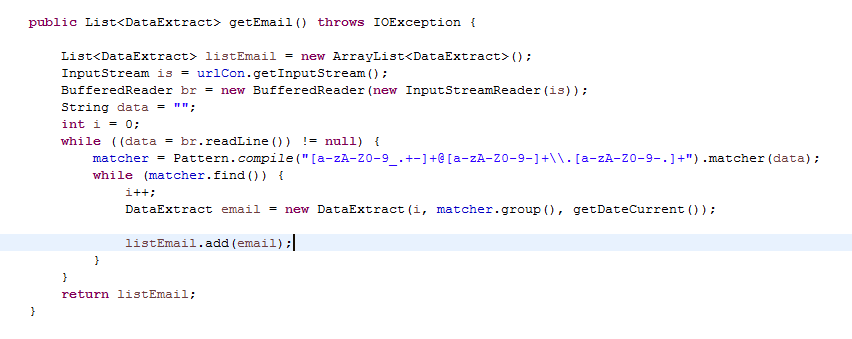
*b) Khi ta nhấn nút “Truy xuất” thì sẽ kết nối đươc website,mã code như sau:*



**3. Lấy dữ liệu**

**3.1.Quy định biểu thức chính quy**

*a)Khi ta chọn dữ liệu muốn truy xuất là Email thì ta sẽ truy xuất được Email xuống phần mềm. Biểu thức chính quy quy định Email như sau:*

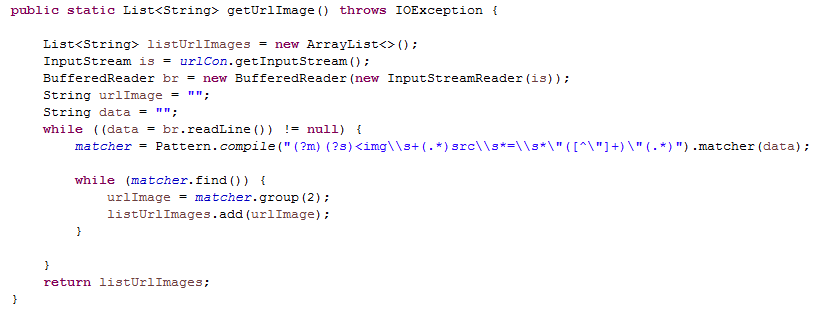


Hhhg

*b)Khi ta chọn dữ liệu muốn truy xuất là số điện thoại thì ta sẽ truy xuất được số điện thoại xuống phần mềm.Biểu thức chính quy quy định số điện thoại như sau:*



*c) Khi ta chọn dữ liệu muốn truy xuất là hình ảnh thì ta sẽ truy xuất được hình ảnh xuống phần mềm. Biểu thức chính quy quy định hình ảnh như sau:*

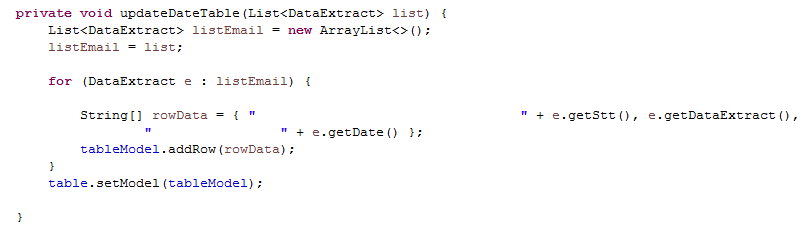


**3.2.Hiển thị dữ liệu lên table.**

*a)Hiển thị dữ liệu:*

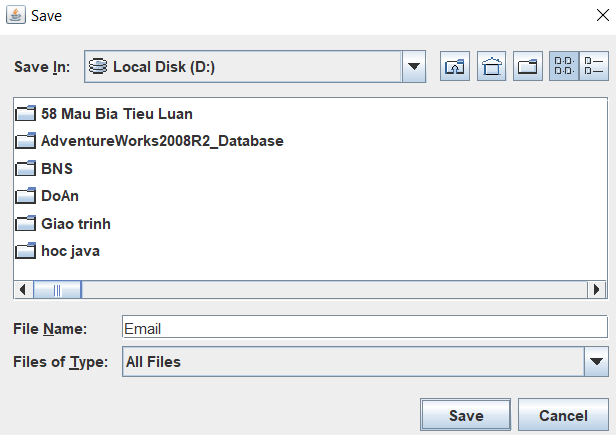
****

*b)Đoạn code như sau:*

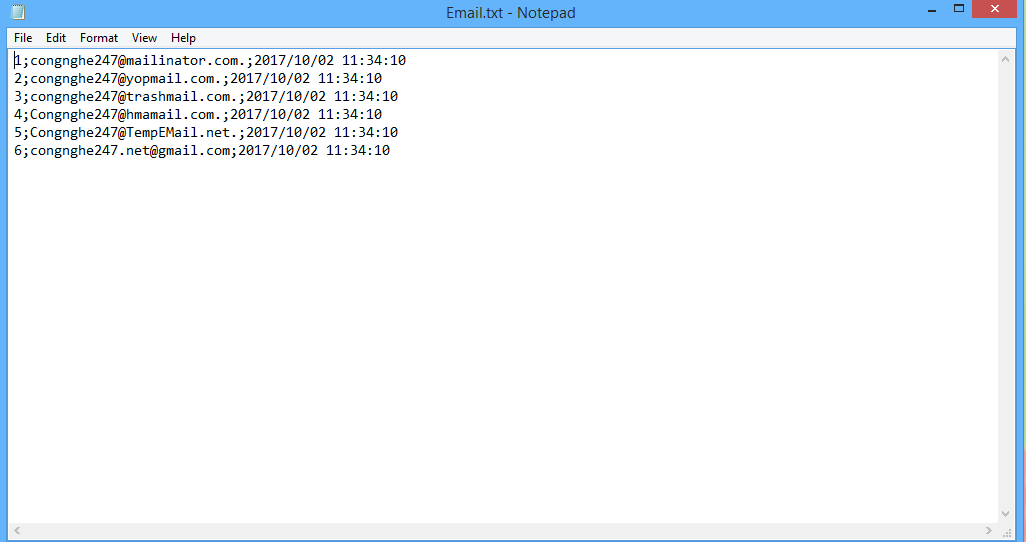


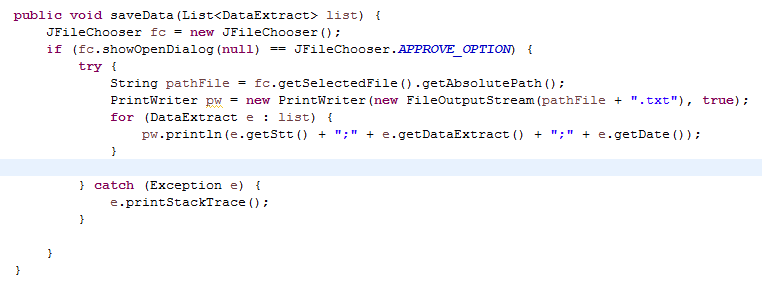
**3.3.Lưu dữ liệu xuống file text**

*a)Khi nhấn nhấn nút lưu thì sẽ mở hộp hội thoại để tìm thư mục mình muốn lưu file và cuối cùng là đặt tên cho file.*



*b)Đây là file text sau khi lưu file.*

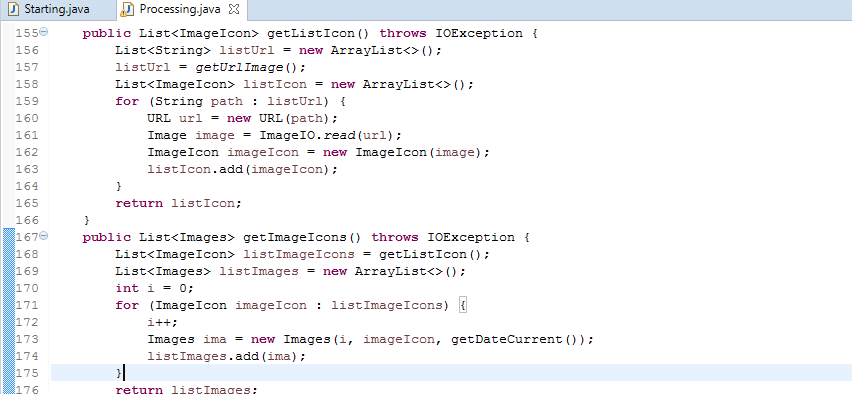


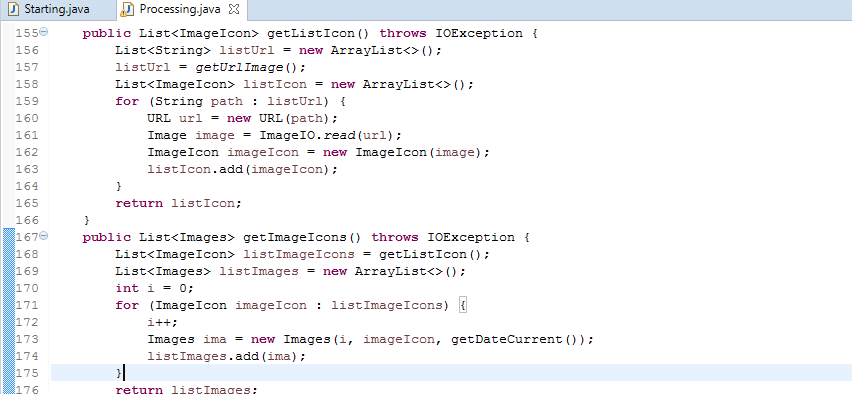
*3.3.Đoạn code để làm việc lưu file như sau:*

**4.Truy xuất dữ liệu ảnh.**

*4.1.Truy xuất dữ liệu ảnh và hiển thị lên table.*



 *4.2.Đoạn code lấy dữ*

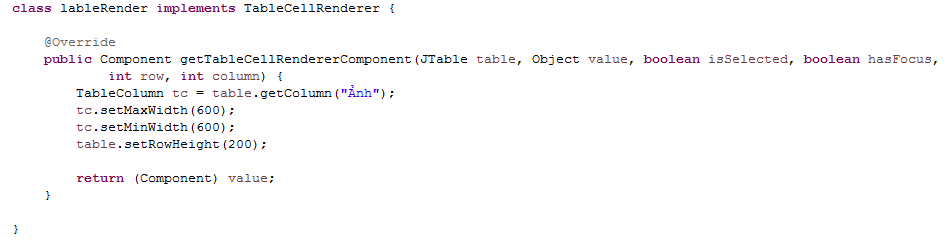


4.3.Đưa ảnh lên table.

*a)Đưa ảnh lên table.*

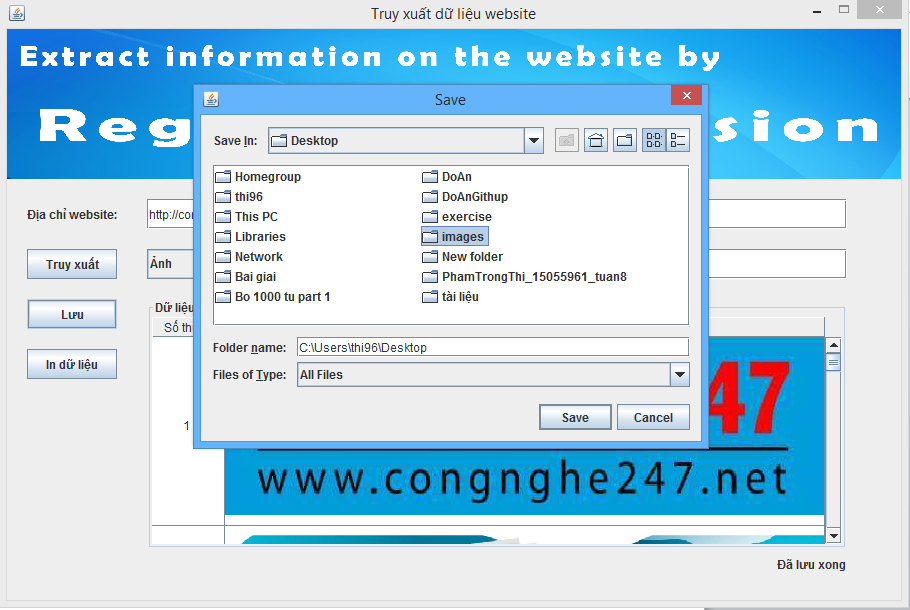


*b)Hiển thị ảnh trên table.*

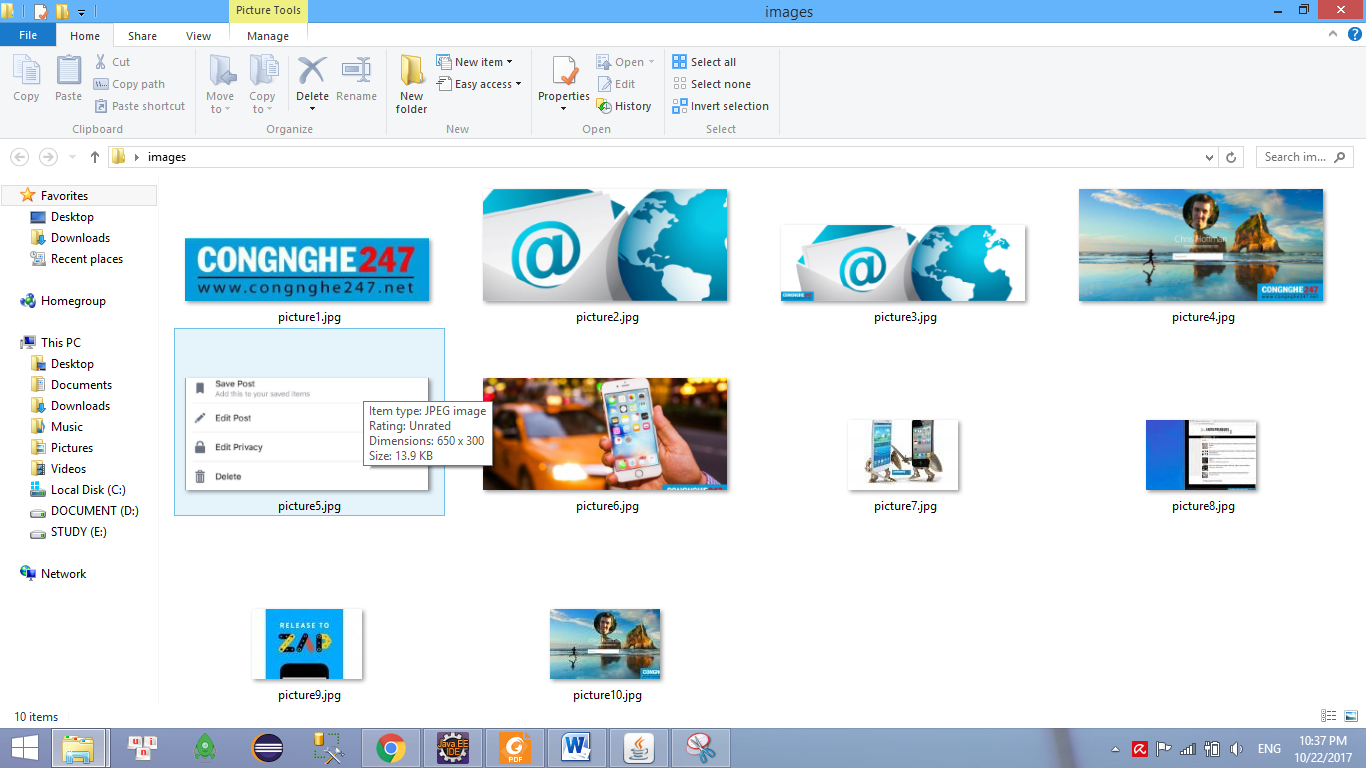


**5.Lưu hình ảnh.**

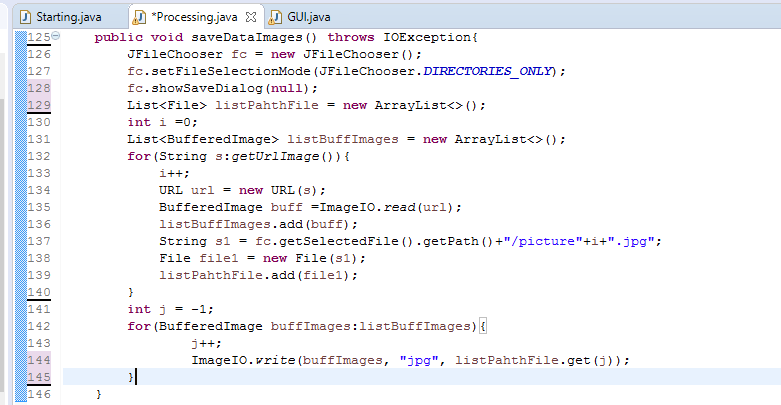
*5.1.Chọn folder cần lưu.*



*5.2. Sau khi lưu, ta có được ảnh sau khi lưu như sau.*

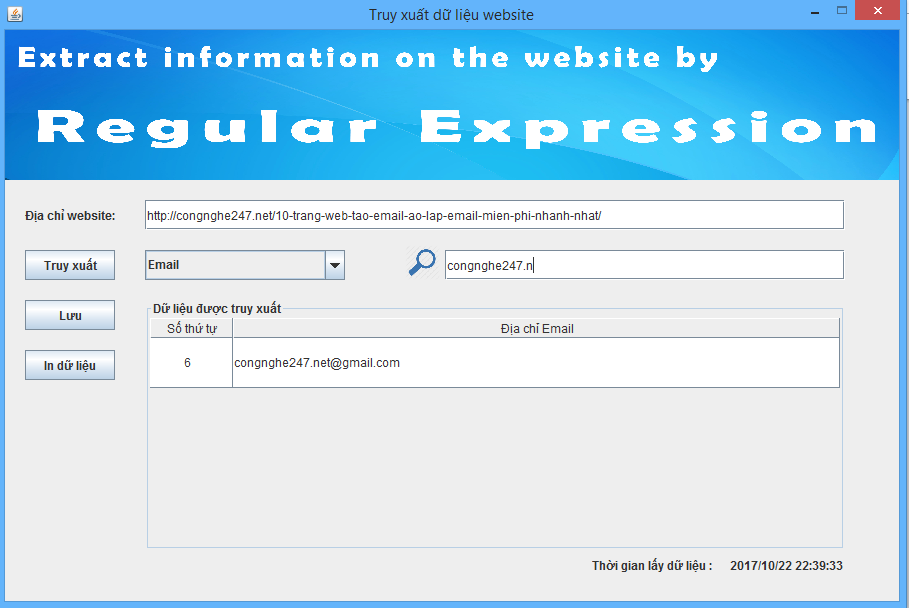


*5.3.Đoạn code để thực hiện công việc trên như sau:*

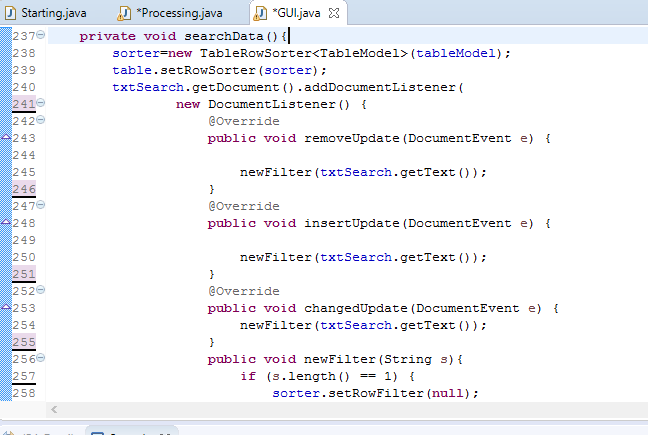


**6.Tìm kiếm dữ liệu:**

*6.1.Ta ghi kí tự đầu tiên thì dữ liệu dần dần hiện lên*

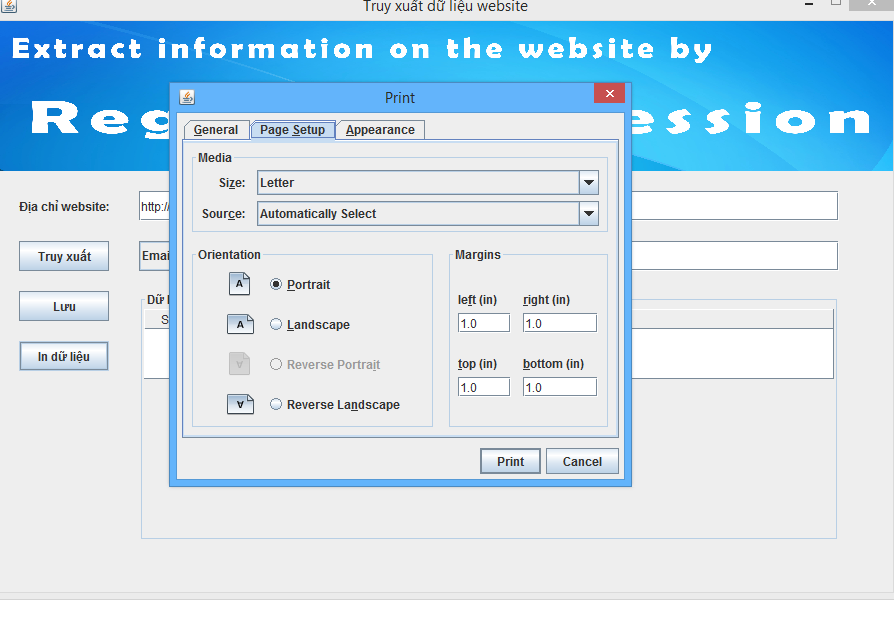


*6.2.Đoạn code như sau:*

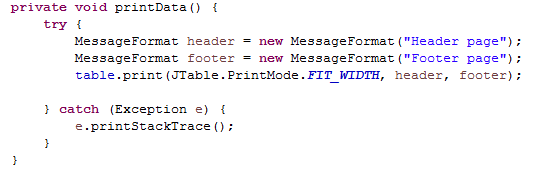


**7.In dữ liệu.**

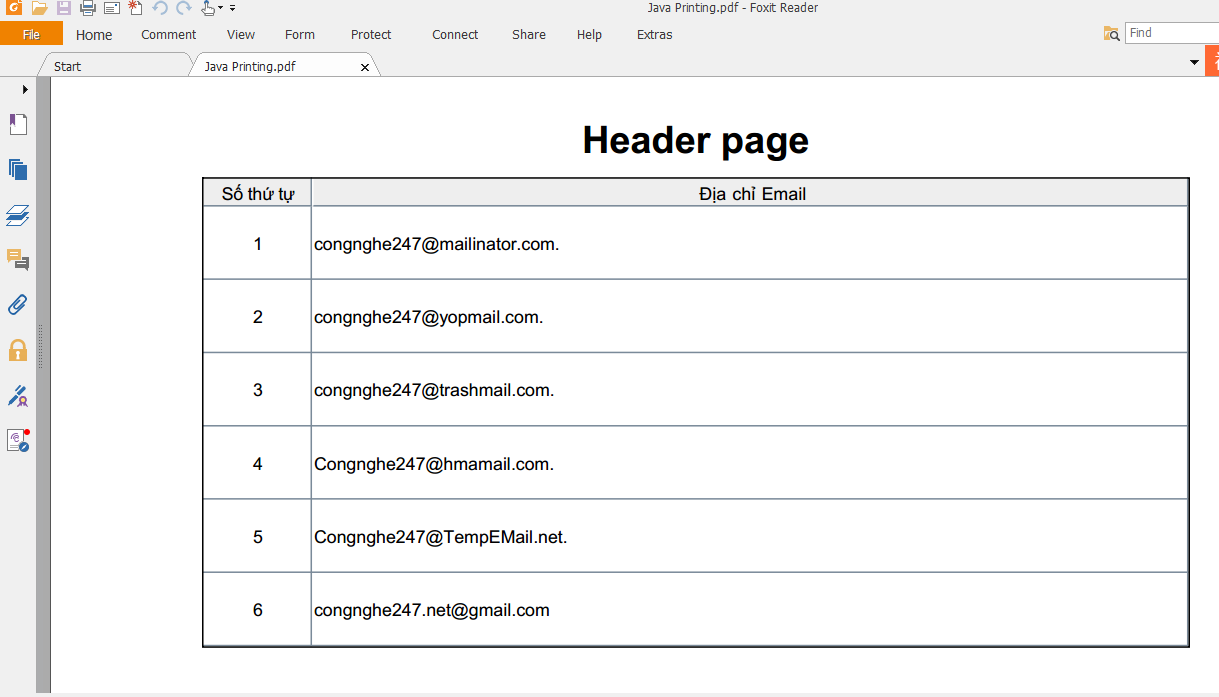
*7.1.Nhấn button in và chọn thuộc tính để in dữ liệu.*

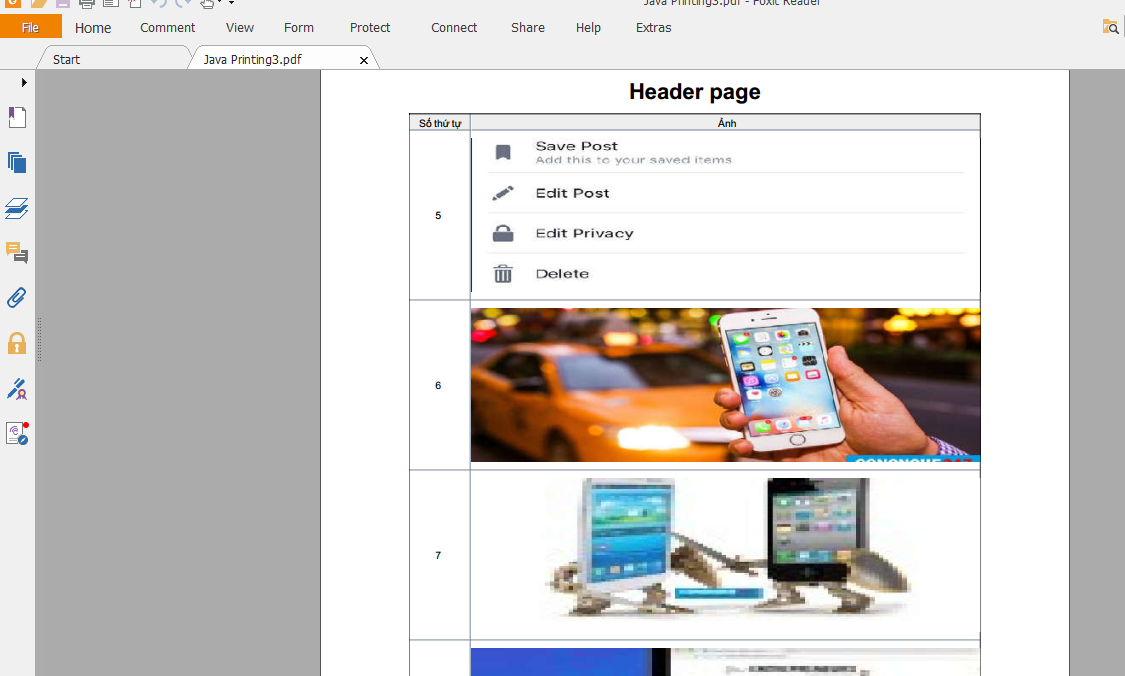


*7.2 Đoạn code như sau:*



*7.3 Ta có được kết quả file pdf như sau:*





# PHẦN III ĐÁNH GIÁ KẾT QUẢ

**I.Đánh giá kết quả cả nhóm:Hoàn thành được 100% đồ án.**

**1.Phần làm được:**

1.1.Nhóm đã lấy được dữ liệu xuống từ trang web,dữ liệu lấy được là email,số điện thoại, hình ảnh.

1.2.Lưu dữ liệu email,số điện thoại xuống file text, dữ liệu ảnh thì được lưu vào trong folder của máy tính.

1.3.Export dữ liệu ra file pdf, thuận tiện cho việc in.

1.4.Chức năng tìm kiếm dữ liệu giúp cho người dùng tìm kiếm dữ liệu một cách nhanh chóng.

**2.Khó khăn của nhóm:**

2.1.Nhóm chưa có được sự nhất quán các ý kiến với nhau.

2.2.Một vài thành viên thì làm không hết mình, làm cho có làm, đến lúc cần giao kết quả thì lại không giao đúng như yêu cầu ban đầu.

2.2.Một vài thành viên chưa có trách nhiệm trong công việc mình đã nhận.

**II.Đánh giá kết quả từng thành viên:**

**1.Trần Nhật Lâm :**

1.1.Sử dụng đồ họa để tạo ảnh phục vụ cho giao diện,viết giao diện như Label,Button, Textfield,Table,xác định mục tiêu dữ liệu cần lấy để tạo ra những chức năng cần thiết sao cho hợp lý.

1.2.Test những website khác nhau kiểm tra lỗi thiếu sót còn gặp phải, đề ra phương án giải quyết cho phần mềm.

1.3.Đưa dữ liệu lên table một cách trực quan cho người dùng phổ thông có thể dễ dàng sử dụng phần mềm.

1.4. Đưa thời gian truy xuất lên phần mềm.

1.5 Lưu dữ liệu số điện thoại xuống file text.

1.6. Viết báo cáo.

**2.Quan Văn Sang:**

2.1.Lưu dữ liệu email xuống file text.

2.2.Test chương trình,tìm kiếm lỗi,tìm kiếm website.

2.3.Viết hàm lấy được ngày lấy dữ liệu.

2.4.Viết được chức năng tìm kiếm dữ liệu.

2.5.Chỉnh sửa lại giao diện.

2.6.Viết được biểu thức chính quy của email.

**3. Nguyễn Ngoc Vỹ:**

3.1.Viết hàm lấy được số điện thoại.

3.2.Viết báo cáo tổng hợp để bàn giao cho Thi gửi lên githup.

3.3.Tìm kiếm các trang web có email, số điện thoại để truy xuất.

3.4.Đưa dữ liệu ảnh lên table.

3.5.Viết được biểu thức chính quy của số điện thoại.

**4. Phạm Trọng Thi:**

4.1.Phân tích website,tìm những dữ liệu có thể lấy được như email,số điện thoại,hình ảnh.

4.2.Kết nối phần mềm với website

4.3.Xử lý button truy xuất để lấy dữ liệu xuống phần mềm.

4.4.Viết hàm lấy được email, viết được hàm đưa ảnh vào giao diện

4.5.Tích hợp các thành phần lại với nhau thành một chương trình hoàn chỉnh: giao diện, kết nối website với phần mềm, truy xuất dữ liệu xuống phần mêm, lưu dữ liệu ra file text.

4.6.Báo cáo kết quả từng tuần đưa lện trang githup.

4.7.Viết được biểu thức chính quy của nguồn ảnh.

4.8.Lấy dữ liệu ảnh từ website xuống phần mềm.

4.9.Viết được chức năng lưu dữ liệu ảnh.

4.10.Viết được chức năng in tất cả dữ liệu ra file pdf.

**III. Hướng phát triển**

1. Nhóm có mong muốn sẽ phát triển phần mềm này về sau này, đặc biệt là đồ án tiếp theo. Để tạo ra phần mềm dễ dàng sử dụng và đưa vào ứng dụng thực tế.

2. Nhóm muốn trong đồ án tiếp theo, sẽ cải thiện phần mềm và nhúng phần mềm vào website, để người dùng có thể sử dụng trực tiếp ở trên website, không nhất thiết phải tải về mới dùng được.

## KẾT QUẢ TỪNG TUẦN CỦA NHÓM

|  |  |  |  |
| --- | --- | --- | --- |
| Thời gian | Người thực hiện | Công việc | Kết quả |
| Tuần 1 | -Sang  -Thi  -Vỹ  -Lâm | -Lập bản kế hoạch chi tiết và cụ thể trong 10 tuần làm việc của nhóm.  -Tìm tài liệu và tổng hợp chi tiết về Java ,OOP.  -Làm quen với Java cơ bản, làm một số bài tập đơn giản. | -Đã có bản kế hoạch thực hiện cho 10 tuần.  -Đã cài đặt và hoàn thành 1 số bài cơ bản về Java,OOP. |
| Tuần 2 | -Vỹ  -Lâm  -Thi  -Sang | -Cả nhóm làm một số bài tập về OOP cơ bản  -Nghiên cứu và tổng hợp tài liệu về Java Swing để làm phần mềm cho đề tài  -Tham khảo một số bài tập mẫu về Java Swing cơ bản và tập làm những phần mềm đơn giản giống các bài tập mẫu đã có. | https://giai-ma.blogspot.bg/2016/04/so-sanh-2-giao-thuc-tcp-va-udp.html |
| Tuần 3 | -Sang  -Thi  -Vỹ  -Lâm | -Tìm tài liệu và tổng hợp cụ thể về Regular Expression.  -Tìm hiểu về website, và một số tài liệu nói về URL để biết cách kết nối phần mềm với website.  -Cả nhóm tìm hiểu về Socket TCP. | https://yinyangit.wordpress.com/2011/06/22/socket-communication-with-tcp-client-server/  https://www.cise.ufl.edu/~amyles/tutorials/tcpchat/ |
| Tuần 4 | -Sang  -Vỹ  -Thi  -Lâm | -Tham khảo và nghĩ ra một vài giao diên đẹp cho phần mềm  -Viết một số giao diện tạm thời của nhóm bằng JavaSwing. |  |
| Tuân 5 | -Thi  -Sang  -Vỹ  -Lâm | -Cả nhóm thống nhất với nhau về một giao diện và cố gắng hoàn thiện phần giao diện,chức năng. |  |
| Tuần 6 | -Sang  -Vỹ  -Thi  -Lâm | -Đọc và tìm hiểu về biểu thức chính quy.  -Kết nối phần mềm với website thông qua địa chỉ website(có thể làm chưa kip được sẽ cố gắng hoàn thiện trong tuần tới). | Đã viết được 1 số biểu thức chính quy  Đã kết nối được phần mềm với website |
| Tuần 7 | -Thi  -Lâm  -Vỹ  -Sang | -Hoàn thiện kết nối website với phần mềm.truy xuất dữ liệu  -Quy định các biếu thức chính quy . |  |
| Tuần 8 | -Vỹ  -Sang  -Thi  -Vỹ | -Quy định biểu thức chính quy của ảnh và dựa vào đó lấy dữ liệu ảnh xuống.  -Hỗ trợ lấy dữ liệu ảnh và đưa dữ liệu ảnh lên phần mềm. |  |
| Tuần 9 | -Vỹ  -Sang  -Thi  -Lâm | -Sẽ viết chức năng in dữ liệu ra,chỉnh sửa lại giao diện.  -Viết chức năng lưu dữ liệu.  -Cả nhóm chạy chương trình và kiểm tra những lỗi sai nếu có. |  |
| Tuần 10 | -Vỹ  -Thi  -Sang  -Lâm | -Cả nhóm báo cáo |  |

**Tài liệu tham khảo**

1. <https://phamky.wordpress.com/l%E1%BA%ADp-trinh-oop/l%E1%BA%ADp-trinh-giao-di%E1%BB%87n-v%E1%BB%9Bi-swing/>
2. <https://www.tutorialspoint.com/java/java_regular_expressions.htm>
3. https://www.tutorialspoint.com/java/java\_url\_processing.htm