Journal of educational equipment: Applied research, Volume 2, Issue 293 (July 2023)
ISSN 1859 - 0810

Phát triển kĩ năng phân tích thị trường chứng khoán cho sinh viên khối ngành Kinh tế qua hệ thống tự động trích xuất thông tin

Trịnh Thanh Bình*

**Trường Đại học Phenikaa*

Received: 19/6/2023; Accepted: 23/6/2023; Published: 26/6/2023

Abstract: *With the rapid advancement of information technology, there is a growing demand for digital data research among organizations, companies, and individuals. This includes the stock market, where access to stock data is essential. To address this need, we have developed a system that employs automatic data extraction techniques, facilitating user access to stock data from the three major stock exchanges: HNX, HOSE, and UPCOM. The system applies the principles of Extract, Transform, and Load (ETL) to extract data from various sources, primarily websites and open data, as accessing the API of the Vietnam stock exchange is limited. By extracting data from multiple sources, the system ensures data authenticity and provides users with the latest information promptly.*

Keywords: *Automatic extraction, stock exchange, authencity*.

1. Đặt vấn đề

Công nghệ thông tin (CNTT) đóng vai trò vô cùng quan trọng trong sự phát triển kinh tế-xã hội, là chìa khóa để mở ra cánh cửa kinh tế tri thức. Sự phát triển của CNTT đã thay đổi cơ bản cấu trúc kinh tế, tổ chức và quy trình sản xuất, cách tiếp cận tri thức, giải trí, tư duy, giải quyết công việc và các mối quan hệ xã hội. Do sự phát triển này, nhu cầu tìm kiếm nguồn thu nhập thụ động từ CNTT ngày càng tăng.

Chứng khoán (CK) là một chứng từ có giá dài hạn hoặc bút toán ghi sổ xác nhận các quyền, lợi ích hợp pháp của người sở hữu đối với vốn hoặc tài sản của tổ chức phát hành. CK là hàng hóa của thị trường CK. CK có giá là hình thức biểu hiện của tư bản giả, bản thân không có giá trị độc lập, là những bản sao bằng giấy tờ của tư bản thực. Những CK có giá mang lại thu nhập cho người sỡ hữu nên nó cũng là đối tượng mua bán và có giá cả. Thông thường trên CK có giá không ghi tên người sở hữu, do đó có thể chuyển nhượng tự do từ người này sang người khác mà không cần có chữ ký của người chuyển nhượng. Trong lịch sử phát triển thị trường CK, lúc đầu CK được in bằng giấy nhưng dần dần được thể hiện dưới hình thức phi vật thể thông qua nghiệp vụ ghi chép kế toán bằng phương tiện điện tử.

Hiện nay, thị trường CK đang thu hút mối quan tâm của nhiều nhà đầu tư, và là một môn học bắt buộc đối với sinh viên (SV) khối ngành kinh tế, vì vậy, vấn đề thu thập, tự động trích xuất thông tin sàn CK, nhằm cung cấp thông tin mới nhất và hữu ích nhất hỗ trợ cho SV trong học tập, nghiên cứu thị trường là thực sự cần thiết. Dựa trên nghiên cứu và phân tích mô hình nghiệp vụ của ứng dụng tự động trích xuất thông tin sàn CK, nền tảng công nghệ xây dựng ứng dụng web, trong bài báo này tác giả đề xuất gải pháp xây dựng ứng dụng tự động trích xuất thông tin sàn CK.

2. Nội dung nghiên cứu

2.1. Thị trường chứng khoán

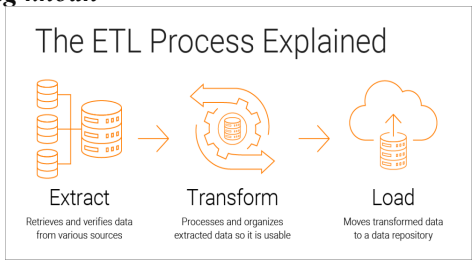
Thị trường CK là một thị trường mà ở đó người ta mua bán, chuyển nhượng, trao đổi CK nhằm mục đích kiếm lời. Thị trường CK trong nền kinh tế hiện đại được quan niệm là nơi diễn ra các hoạt động mua bán CK trung và dài hạn. Việc mua bán này được tiến hành ở thị trường sơ cấp khi người mua mua được CK lần đầu từ những người phát hành, và ở những thị trường thứ cấp khi có sự mua đi bán lại các CK đã được phát hành ở thị trường sơ cấp. Như vậy, xét về mặt hình thức, thị trường CK chỉ là nơi diễn ra các hoạt động trao đổi, mua bán, chuyển nhượng các loại CK, qua đó thay đổi chủ thể nắm giữ CK.

Có ba cách để phân loại cơ bản thị trường CK, đó là căn cứ vào sự luân chuyển của các nguồn vốn, căn cứ vào hàng hoá trên thị trường và căn cứ vào phương thức hoạt động của thị trường. Căn cứ vào sự luân chuyển của các nguồn vốn, có hai loại là thị trường sơ cấp và thị trường thứ cấp. Thị trường CK sơ cấp là nơi duy nhất mà CK đem lại vốn cho người phát hành. Giá chứng khoán trên thị trường sơ cấp

(giá phát hành) do tổ chức phát hành quyết định. Những người bán trên thị trường sơ cấp thường là kho bạc, ngân hàng nhà nước, công ty phát hành, tập đoàn bảo lãnh phát hành. Thị trường thứ cấp không trực tiếp mang lại vốn cho người đầu tư sản xuất kinh doanh. Giao dịch trên thị trường thứ cấp phản ánh nguyên tắc tự do, cạnh tranh tự do. CK trên thị trường thứ cấp có thể được mua bán nhiều lần.

Căn cứ vào hàng hoá trên thị trường, có ba loại thị trường là thị trường cổ phiếu, thị trường trái phiếu và thị trường các công cụ CK phái sinh. Thị trường cổ phiếu là thị trường giao dịch và mua bán các loại cổ phiếu, bao gồm cổ phiếu thường và cổ phiếu ưu đãi. Thị trường trái phiếu là thị trường giao dịch và mua bán các trái phiếu đã được phát hành, các trái phiếu này bao gồm các trái phiếu công ty, trái phiếu đô thị và trái phiếu chính phủ. Thị trường các công cụ CK phái sinh là thị trường phát hành và mua đi bán lại các chứng từ tài chính khác ví dụ như là quyền mua cổ phiếu, chứng quyền và hợp đồng quyền chọn.

2.2. Trích xuất và lưu trữ thông tin thị trường chứng khoán



Hình 1. Quy trình ETL

Hệ thống trích xuất tự động thông tin sàn CK tuân theo quy tắc ETL. ETL là quy trình chung sao chép dữ liệu từ một hoặc nhiều nguồn vào hệ thống đích đại diện cho dữ liệu khác với nguồn. Quá trình ETL đã trở thành một khái niệm phổ biến trong những năm 1970 và thường được sử dụng trong kho dữ liệu. Extract/Trích xuất là quá trình đọc dữ liệu từ nhiều nguồn, ví dụ như cơ sở dữ liệu mở, các trang web có dữ liệu mong muốn. Trong giai đoạn này, dữ liệu được thu thập. Transform/Biến đổi là quá trình chuyển đổi dữ liệu được trích xuất từ biểu mẫu trước đó thành biểu mẫu cần có để có thể được đặt vào cơ sở dữ liệu khác. Chuyển đổi xảy ra bằng cách sử dụng các quy tắc hoặc bảng tra cứu hoặc bằng cách kết hợp dữ liệu này với dữ liệu khác. Load/Tải là quá trình ghi chép dữ liệu vào cơ sở dữ liệu đích. Vì việc trích xuất dữ liệu cần có thời gian, nên thường thực hiện song song ba giai đoạn. Trong khi dữ liệu đang được trích xuất, một quá trình chuyển đổi khác sẽ thực thi trong khi xử lý dữ liệu đã nhận và chuẩn bị để tải trong khi quá trình tải dữ liệu bắt đầu mà không cần chờ hoàn thành các giai đoạn trước. Các giai đoạn của quá trình ETL bao gồm 3 giai đoạn là giai đoạn trích xuất, giai đoạn chuyển đổi và cuối cùng là giai đoạn tải.

Giai đoạn trích xuất là phần đầu tiên của quy trình ETL, liên quan đến việc trích xuất dữ liệu từ các nguồn, để dữ liệu được chuyển đến một đích mới, trước tiên nó phải được trích xuất từ các nguồn. Mặc dù các dữ liệu này có thể xử lý thủ công, nhưng việc trích xuất dữ liệu được mã hóa bằng tay có thể tốn nhiều thời gian và dễ bị lỗi. Các công cụ ETL tự động hóa quá trình trích xuất và tạo ra một quy trình làm việc hiệu quả và đáng tin cậy hơn.

2.3. Một số công nghệ

React (hay còn được gọi là React.js hoặc ReactJS) là một thư viện JavaScript front-end mã nguồn mở và miễn phí để xây dựng giao diện người dùng dựa trên các thành phần UI riêng lẻ. Nó được phát triển và duy trì bởi Meta (trước đây là Facebook) và cộng đồng các nhà phát triển và công ty cá nhân. React có thể được sử dụng làm cơ sở để phát triển các ứng dụng SPA (Single-page), thiết bị di động hoặc ứng dụng được kết xuất bằng máy chủ với các thư viện khác như Next.js.

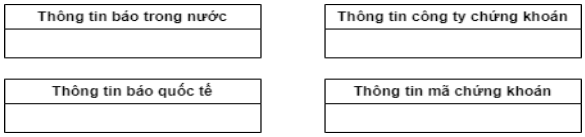
MariaDB là một sản phẩm mã nguồn mở tách ra từ mã mở do cộng đồng phát triển của hệ quản trị cơ sở dữ liệu quan hệ MySQL nhằm theo hướng không phải trả phí với GNU GPL. MariaDB được phát triển từ sự dẫn dắt của những nhà phát triển ban đầu của MySQL, do lo ngại khi MySQL bị Oracle Corporation mua lại. Những người đóng góp được yêu cầu chia sẻ quyền tác giả của họ với MariaDB Foundation.

Spring Boot là một dự án phát triển bởi ngôn ngữ Java trong hệ sinh thái Spring framework. Spring Boot giúp lập trình viên đơn giản hóa quá trình lập trình một ứng dụng với Spring và tập trung vào phát triển với kiến trúc dịch vụ siêu nhỏ (microservices). Sơ qua về dịch vụ siêu nhỏ thì đây là một kiến trúc với các module được chia thành các dịch vụ rất nhỏ. Mỗi dịch vụ sẽ được đặt trên một máy chủ riêng nên dễ dàng để nâng cấp và phát triển ứng dụng.

Jsoup là một thư viện Java để làm việc với HTML trong thời gian thực. Jsoup cung cấp một API rất thuận tiện để tìm nạp URL, trích xuất và thao tác dữ liệu sử dụng các phương thức HTML5 và bộ chọn CSS tốt nhất.

2.4. Các yêu cầu hệ thống

Hệ thống trích xuất tự động thông tin sàn CK là hệ thống được xây dựng với mục đích cung cấp thông tin về CK một cách tự động đối với những người quan tâm đến CK. Hệ thống này giúp tiết kiệm thời gian và công sức; đồng thời người dùng sẽ có thông tin liên quan đến CK một cách nhanh nhất: là thông tin thế giới, trong nước, thông tin về các công ty đăng ký chứng khoán hay là thông tin về số liệu cuối ngày của chứng khoán. Hệ thống sẽ đáp ứng được các chức năng cơ bản như: đọc thông tin thế giới và thông tin trong nước, xem thông tin về các công ty đăng ký CK, xem thông tin khi sàn CK đóng của các mã giao dịch CK.

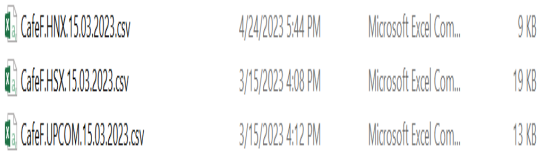


Hình 2. Các thành phần hệ thống

2.5. Dịch vụ phần mềm

Hệ thống trích xuất tự động sàn CK là hệ thống được xây dựng với quy mô với đối tượng là những người quan tâm đến sàn CK. Mục đích của hệ thống là cung cấp thông tin CK một cách nhanh và chính xác nhất, với đơn vị cập nhật thông tin theo ngày. Hệ thống có backend được xây dựng bằng framework Spring Boot, frontend là ReactJs. Hệ quản trị cơ sở dữ liệu Mariadb, trích xuất tự động dựa theo phương pháp ETL.

Chức năng trích xuất và xử lý dữ liệu thô từ các nguồn: Đây là một quá trình được xử lý ở backend 1 cách tự động, không có ui/ux nhằm trích xuất dữ liệu và xử lý dữ liệu thô một cách tự động từ các nguồn một cách nhanh nhất. Hình 3 mô tả cho các dữ liệu về công ty CK được trích xuất một cách tự động khi hệ thống đang chạy.



Hình 3. Dữ liệu được trích xuất sau khi giải nén.

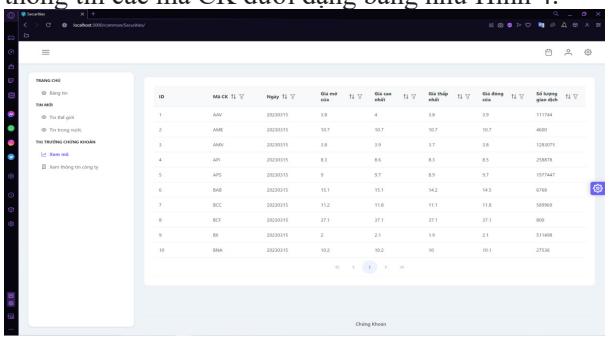
Đặc điểm chung của công việc tiến hành trích xuất là nhận ra điểm giống nhau của đường dẫn tới các trang. Ví dụ như trang Vnexpress sẽ có 30 bài báo 1 trang. Mỗi trang sẽ có đường dẫn tương ứng là p-i, với i là 1 số nguyên dương từ bé tới lớn úng với từ mới nhất tới cũ nhất.

Chức năng đọc tin tức về kinh tế quốc tế cho phép người dùng truy cập vào để xem thông tin thế giới. Màn hình sẽ hiển thị các bài báo dưới dạng tiêu đề và tóm tắt. Một trang tin tức sẽ có gồm 10 bài báo và bộ chuyển trang ở cuối trang.

Chức năng đọc tin tức về CK trong nước tương tự như màn hình xem tin tức kinh doanh quốc tế, màn hình xem CK trong nước cũng hiển thị các bài báo dưới dạng tóm tắt và tiêu đề. Một trang tin tức sẽ có 10 bài báo và bộ chuyển trang ở cuối trang.

Chức năng xem thông tin về các công ty đăng ký CK. Mô tả: trang hiển thị thông tin về các công ty CK theo dạng bảng với đầy đủ các thông tin. Ngoài ra, người dùng có thể tìm kiếm theo các trường thông tin mình mong muốn.

Chức năng xem thông tin về một mã CK. Mô tả: trang hiển thị thông tin về mã CK theo dạng bảng với đầy đủ các thông tin. Thao tác: người dùng chọn mục xem mã chứng khoán. Lúc này trang web sẽ hiển thị thông tin các mã CK dưới dạng bảng như Hình 4.



Hình 4. Chức năng xem thông tin mã chứng khoán

3. Kết luận

Hệ thống trích xuất tự động thông tin CK hiện tại có khả năng lấy thông tin từ nhiều nguồn uy tín khác nhau ở trong và ngoài nước như: Vnexpress, Việt Nam VSD, cafef... Người dùng có thể thấy thông tin được lọc theo nhiều cách khác nhau: thấy được top 10 CK hot trong ngày, xem lịch sử giao dịch, giá... của một mã CK.

Hệ thống trình bày trong bài báo đã cung cấp một giải pháp đơn giản, nhanh chóng, linh hoạt trong quá trích xuất thông tin thị trường CK, từ đó giúp cho các nhà đầu tư có thể dễ dàng nắm bắt thông tin, hỗ trợ đưa ra các quyết định chính xác.

Tài liệu tham khảo

[1]Https://luatminhkhue.vn/thi-truong-chung-khoanla-gi.aspx (2020) Thị trường chứng khoán là gì.

[2] Jordan Walke, React docs (2021). Accessed on: April 20, 2021. [Online]

[4] Ryan Dahl, Nodejs docs (2021). Accessed on 2021.