



BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HỒ CHÍ MINH

ĐỒ ÁN TỐT NGHIỆP

ỨNG DỤNG AI XÂY DỰNG HỆ THỐNG BÁN HÀNG THÔNG MINH

Ngành: **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **CÔNG NGHỆ PHẦN MỀM**

Giảng viên hướng dẫn: **Th.S Võ Tấn Dũng**

Sinh viên thực hiện: **Nguyễn Văn Huỳnh**

MSSV: 1911066137 Lớp: 19DTHD4

TP. Hồ Chí Minh, 2023

ĐỒ ÁN TỐT NGHIỆP

ỨNG DỤNG AI XÂY DỰNG HỆ THỐNG BÁN HÀNG THÔNG MINH

Ngành: **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **CÔNG NGHỆ PHẦN MỀM**

Giảng viên hướng dẫn: **Th.S Võ Tấn Dũng**

Sinh viên thực hiện : **Nguyễn Văn Huỳnh**

MSSV: 1911066137 **Lớp: 19DTHD4**

TP. Hồ Chí Minh, 2023

LỜI CAM ĐOAN

Em xin cam đoan đồ án tốt nghiệp với đề tài “**Ứng dụng AI xây dựng hệ thống bán hàng thông minh**” là đề tài nghiên cứu của em dưới sự hướng dẫn của ThS.Võ Tấn Dũng. Những nhận định được nêu ra trong đồ án cũng là kết quả từ sự nghiên cứu trực tiếp, nghiêm túc, độc lập dựa vào các cơ sở tìm kiếm, hiểu biết và nghiên cứu tài liệu khoa học có liên quan. Đồ án vẫn sẽ giúp đảm bảo được tính khách quan, trung thực và khoa học.

MỤC LỤC

Chương 1. TỔNG QUAN.....	1
1.1 Đặt vấn đề.....	1
1.2 Các nghiên cứu liên quan.....	2
1.2.1 Các bài viết liên quan dẫn chứng số liệu về hệ thống đề xuất.....	2
1.2.2 Các nghiên cứu liên quan xử lý ngôn ngữ tự nhiên NLP trong tìm kiếm, đề xuất.....	3
1.3 Nhiệm vụ của đồ án.....	5
1.3.1 Tính cấp thiết và lý do chọn đề tài.....	5
1.3.2 Mục tiêu đề tài.....	6
1.3.3 Đối tượng và phạm vi nghiên cứu của đề tài.....	6
1.3.4 Phương pháp nghiên cứu.....	6
1.4 Kết cấu của đề tài.....	7
1.5 Tóm tắt chương I.....	7
Chương 2. CƠ SỞ LÝ THUYẾT	8
2.1 Các khái niệm.....	8
2.1.1 Hệ thống đề xuất là gì?.....	8
2.1.2 Bài toán đề xuất.....	8
2.1.3 Các cách tiếp cận phổ biến.....	9
2.1.4 Lựa chọn phương pháp tiếp cận để giải quyết vấn đề.....	9
2.2 Phương pháp giải quyết vấn đề.....	10
2.2.1 Vấn đề đặt ra trong việc tìm kiếm, đề xuất bằng NLP.....	10
2.2.2 Các phương pháp đề xuất áp dụng trong NLP.....	11
2.2.3 Công nghệ phát triển AI.....	12
2.2.4 Công nghệ phát triển Web.....	14
2.3 Tóm tắt chương 2.....	23
Chương 3. THỰC TRẠNG VÀ PHÂN TÍCH VẤN ĐỀ	24
3.1 Các bước thực hiện đề tài.....	24
3.1.1 Phát triển website bán sách.....	24
3.1.2 Xây dựng triển khai mô hình sử dụng NLP cho website bán sách.....	37
3.2 Các kết quả nghiên cứu thực nghiệm.....	40
3.2.1 Các kết quả thực nghiệm áp dụng NLP vào bán hàng.....	40
3.2.2 Các chức năng của website bán hàng.....	42

3.3 Tóm tắt chương 3	53
Chương 4. GIẢI PHÁP GIẢI QUYẾT.....	54
4.1 Kết luận chung	54
4.2 Đề xuất mới	54
TÀI LIỆU THAM KHẢO.....	55

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Từ viết tắt	Ý nghĩa
NLP	Natural language processing
BE	Backend
FE	Frontend
CB	Content based
CF	Collaborative filtering
IE	Information extraction

DANH MỤC CÁC BẢNG

Bảng 1 User	27
Bảng 2 OrderDetails	28
Bảng 3 Order	28
Bảng 4 Product	29
Bảng 5 Trademark	30
Bảng 6 Category	30
Bảng 7 Meta_product	31
Bảng 8 Comment	31
Bảng 9 WebHistory	32
Bảng 10 Post.....	32
Bảng 11 Topic	33
Bảng 12 Contact	33
Bảng 13 Role	34
Bảng 14 Notification	34
Bảng 15 Banner	35

DANH MỤC HÌNH ẢNH

Hình 1 Khảo sát 1 số website bán hàng.....	1
Hình 2 Các thống kê về hệ thống đề xuất trong các hệ thống lớn.....	2
Hình 3 Các bằng chứng về sử dụng NLP trong sơ đồ tri thức có hệ thống đề xuất.....	4
Hình 4 Mô hình đại học Oxford báo cáo về NLP đề xuất.....	5
Hình 5 Thang điểm đánh giá về độ chính xác NLP đạt được	5
Hình 6 Các phương pháp tiếp cận trong đề xuất	9
Hình 7 Phương pháp Content based	10
Hình 8 Phương pháp xử lý ngôn ngữ tự nhiên NLP.....	11
Hình 9 Thư viện Scikit-learn học sâu	14
Hình 10 React	14
Hình 11 React function component vs class component.....	16
Hình 12 Vòng React component	17
Hình 13 Ví dụ về React.createClass	18
Hình 14 Nodejs	19
Hình 15 Các module trong Nodejs	20
Hình 16 Cấu trúc ExpressJS	21
Hình 17 MongoDB	22
Hình 18 Sơ đồ ERD hệ thống shop	27
Hình 19 Sơ đồ BFD	35
Hình 20 Sơ đồ Usecase hệ thống.....	36
Hình 21 Sơ đồ Classdiagram	36
Hình 22 Count Vectorizer	39
Hình 23 Minh họa đồng dạng Cosine.....	39
Hình 24 Công thức tính độ đồng dạng Cosine	39
Hình 25 Trang chủ website.....	40
Hình 26 Tìm kiếm sử dụng đề xuất lịch sử mua hàng.....	41
Hình 27 Đề xuất sách bằng NLP	41
Hình 28 Kết quả đề xuất sách.....	42
Hình 29 Đăng ký.....	42
Hình 30 Đăng nhập.....	43

Hình 31 Đăng nhập Google.....	43
Hình 32 Hồ sơ cá nhân	44
Hình 33 Thay đổi ảnh đại diện	44
Hình 34 SEO URL.....	44
Hình 35 Xem lịch sử mua hàng	45
Hình 36 Xem lịch sử thao tác trên website.....	45
Hình 37 Thông báo thêm vào giỏ hàng thành công	46
Hình 38 Xem chi tiết giỏ hàng	46
Hình 39 Tiến hành đặt hàng, có áp dụng mã giảm giá.....	46
Hình 40 Nhập thông tin đặt hàng	47
Hình 41 Thanh toán online bằng MoMo	47
Hình 42 Thông báo thanh toán thành công	47
Hình 43 Khách xem trang thái đơn hàng của mình.....	48
Hình 44 Trang quản lý.....	48
Hình 45 Quản lý đơn hàng	48
Hình 46 Xử lý tình trạng đơn hàng.....	49
Hình 47 Xem chi tiết đơn hàng quản lý.....	49
Hình 48 Quản lý khách hàng	50
Hình 49 Xem các đơn hàng của mỗi người.....	50
Hình 50 Quản lý sản phẩm	51
Hình 51 Cập nhật thông tin sản phẩm (Sách).....	51
Hình 52 Tài khoản hệ thống	52
Hình 53 Thống kê	52

Chương 1. TỔNG QUAN

1.1 Đặt vấn đề



Hình 1 Khảo sát 1 số website bán hàng

Công nghệ thông tin ngày càng phát triển, đặc biệt là **AI** đã tạo bước ngoặt mới cho sự triển kinh tế xã hội toàn cầu. Chính nền tảng đó, một hình thức thương mại mới đã xuất hiện và phát triển nhanh chóng, đó là **thương mại điện tử - bán hàng trực tuyến**. Thương mại điện tử chính là một công cụ hiện đại sử dụng mạng Internet giúp cho các doanh nghiệp có thể thâm nhập vào thị trường thế giới, thu thập thông tin nhanh hơn, nhiều hơn và chính xác hơn.

Với thương mại điện tử áp dụng **AI** vào hệ thống, các doanh nghiệp có thể đưa các thông tin về sản phẩm của mình đến các đối tượng khách hàng tiềm năng khác nhau ở mọi nơi trên thế giới với chi phí thấp hơn nhiều so với các phương pháp truyền thống.

Với khả năng của hệ thống đề xuất thông minh, các doanh nghiệp sử dụng chúng để giới thiệu sản phẩm tới người tiêu dùng, giúp gia tăng doanh số nhờ các ưu đãi, sản phẩm, dịch vụ được khuyến nghị một cách cá nhân hóa, làm nâng cao trải nghiệm khách hàng. Điều này cải thiện lợi thế cạnh tranh của doanh nghiệp và giảm thiểu tỉ lệ khách hàng rời bỏ và đến với đối thủ cạnh tranh khi họ nhận thấy doanh nghiệp hiểu nhu cầu của họ và cung cấp cho họ những thứ họ muốn.

Các công ty lớn như Amazon, Netflix và Vinabook đều sử dụng hệ thống đề xuất sản phẩm để giúp khách hàng tìm kiếm sản phẩm một cách nhanh chóng và thuận tiện hơn. Các hệ thống đề xuất hiện đại dựa trên **Deep learning** đang thúc đẩy sự phát triển của những gã khổng lồ trực tuyến. Hiện tại, với tính sẵn sàng của các hệ thống xử lý dữ

liệu hiệu năng cao qua GPU, bất cứ doanh nghiệp nào cũng có thể áp dụng được hệ thống này.

Với sự phát triển mạnh mẽ của các hệ thống, ứng dụng áp dụng AI hiện nay, chúng ta có thể nắm bắt được cơ hội này để có thể nghiên cứu, đề xuất các phương pháp tốt hơn trong lĩnh vực bán hàng và một trong số đó chúng ta có thể áp dụng AI đó là giải quyết vấn đề: “*Xây dựng hệ thống tìm kiếm sách đề xuất cho khách hàng dựa trên xử lý ngôn ngữ tự nhiên NLP.*”. Đó cũng chính là nội dung bài toán mà đề tài này hướng đến giải quyết thay cho việc tìm kiếm thông thường như các trang web bán hàng hiện nay.

1.2 Các nghiên cứu liên quan

1.2.1 Các bài viết liên quan dẫn chứng số liệu về hệ thống đề xuất

Theo tạp chí quốc tế Forbes bài viết của tác giả Marshall (là VP Product tại SambaNova) đã nghiên cứu và đánh giá xác nhận rằng hiện nay trên web xem phim trực tuyến lớn nhất thế giới là Netflix – họ đã áp dụng thành công hệ thống đề xuất các bộ phim đến người xem với nhiều tiêu chí người xem khác nhau, hệ thống hoạt động tốt với gần 5000 bộ phim và 167 triệu người dùng. Còn đối với website nghe nhạc nổi tiếng nhất thế giới là Spotify họ cũng đã áp dụng thành công hệ thống đề xuất với 60 triệu track nhạc và 320 triệu người dùng.[9]

Mỗi phần nội dung có hàng trăm thuộc tính và mỗi hồ sơ người dùng cũng vậy. Đưa ra các đề xuất chính xác ở cấp độ người dùng cá nhân, tất cả đều trong thời gian thực và với tất cả dữ liệu đó, là một công việc cực kỳ chuyên sâu về điện toán, đó là lý do tại sao ngày nay nó phần lớn được tự động hóa bởi các hệ thống hỗ trợ bởi AI và ML.

Youtube	Amazon	Viblo
<ul style="list-style-type: none">- 300 hours of video are uploaded to YouTube every minute!- Almost 5 billion videos are watched on Youtube every single day.- The total number of people who use YouTube 1,300,000,000 users	<ul style="list-style-type: none">- Sells more than 119,928,851 products- In the U.S. alone, Amazon has over 150 million monthly unique visitors.	<ul style="list-style-type: none">- Over 26,000 verified users- Over 21,000 published posts- 1.5M Pageviews per month

Hình 2 Các thống kê về hệ thống đề xuất trong các hệ thống lớn

Theo bài viết của tạp chí khoa học Sunews của tác giả Phạm Hoàng Anh (thuộc đơn vị R&D) đã công bố số liệu thống kê được trên 3 nền tảng Youtube, Amazon, Viblo [12] với thực tế nghiên cứu được là vào năm 2019 dựa trên Youtube là chúng ta có thể thấy được tầm quan trọng của hệ thống đề xuất và 1 số mục tiêu mà hệ thống đề xuất nên hướng đến đạt hiệu quả tốt hơn:

1. **Sự liên quan:** Người dùng sẽ truy cập/ sử dụng/ mua những thứ liên quan tới sở thích, thói quen của họ.
2. **Sự mới lạ:** Việc gợi ý những sản phẩm hot, trending sẽ ít mang lại cảm giác mới lạ cho người sử dụng và dẫn tới việc mất cân bằng và sự phong phú về mặt sản phẩm của hệ thống.
3. **Sự bất ngờ:** Một mức độ cao hơn của mới lạ, khi mà sản phẩm được gợi ý không chỉ chưa từng được người dùng biết đến mà còn gây bất ngờ cho họ. Có thể hiểu đơn giản hơn là những sự liên quan mà người dùng không hề biết tới. Ví dụ như đối với một người thích và thường xuyên ăn đồ ăn Trung Quốc, việc hệ thống gợi ý cho họ một cửa hàng đồ ăn Trung Quốc mà họ chưa từng tới được đánh giá là **mới lạ** nhưng sẽ không có **bất ngờ**. Trong trường hợp nếu hệ thống gợi ý có thể đưa ra đề xuất về một cửa hàng chuyên về các món ăn truyền thống của Việt Nam sẽ được gọi là có **sự bất ngờ**. Tất nhiên, sự bất ngờ ở đây cũng phải được tính toán rất cẩn thận thông qua các thông tin để đảm bảo rằng người dùng sẽ có khả năng cao có hứng thú với sản phẩm này.
4. **Sự phong phú:** Trước đây, các hệ thống gợi ý thường sẽ đưa ra một danh sách Top-K các sản phẩm **rất liên quan** tới nhau để gợi ý cho người dùng. Điều này có nguy cơ dẫn tới một khả năng là người dùng sẽ **không thích bất cứ một cái nào** ở trong danh sách trên. Nói cách khác, **sự phong phú** tăng xác suất trong việc người dùng chắc chắn sẽ thích ít nhất một sản phẩm từ danh sách đó, không gây nhàm chán cho họ bởi những sản phẩm giống nhau lặp đi lặp lại.

1.2.2 Các nghiên cứu liên quan xử lý ngôn ngữ tự nhiên NLP trong tìm kiếm, đề xuất

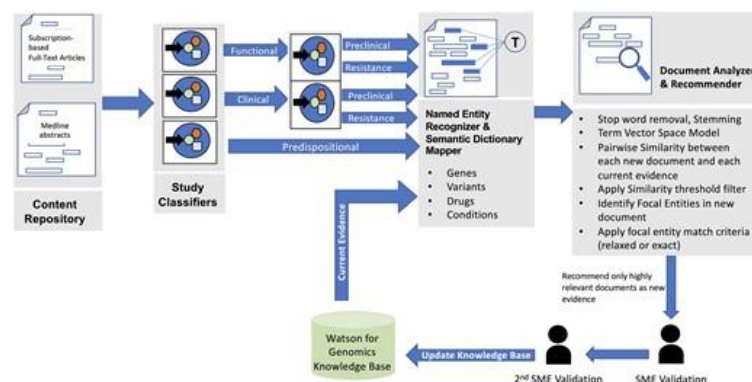
Theo bài viết khoa học đại học Oxford của tác giả Bhuvan Sharma và công sự đã nghiên cứu về cách áp dụng xử lý ngôn ngữ tự nhiên NLP và học máy để xây dựng hệ thống đề xuất cho sơ đồ tri thức Sinh học bao gồm 700 gen, 100 triệu biến thể, 1200 loại

thuốc và 300 tình trạng liên kết với đạt được điểm xác suất chính xác với con số từ 88% đến 94% khi đang thử nghiệm mô hình.

Loại hiệp hội	Thực thể liên kết	Ví dụ	Chứng cứ
chức năng	Gen và biến thể, hoặc gen tổng hợp	Gen tổng hợp TTVH3-BRAF	<ul style="list-style-type: none"> PMID: 31558800 Nghiên cứu này mô tả việc phát hiện ra một gen dung hợp TTVH3-BRAF mới, rất khác thường chứa protein BRAF có chiều dài gần như đầy đủ bao gồm miền tự ức chế nhưng vẫn có khả năng gây bệnh hoàn toàn.
lâm sàng	Gen, biến thể, liệu pháp, tình trạng	Đột biến EGFR L858R và phản ứng với afatinib trong NSCLC	<ul style="list-style-type: none"> PMID: 23816960 Đột biến L858R trong gen EGFR đại diện cho đột biến kích hoạt phổ biến nhất trong NSCLC và có liên quan đến phản ứng với các chất ức chế EGFR nhằm mục tiêu. Bàn thảo mô tả hiệu quả lâm sàng của thuốc ức chế EGFR thế hệ thứ hai (afatinib) ở bệnh nhân NSCLC bị đột biến này.
tiền lâm sàng	Gen, biến thể, liệu pháp, tình trạng	Đồng nhắm mục tiêu của JAK2 và HDAC trong MPN bằng cách sử dụng ruxolitinib và vorinuler	<ul style="list-style-type: none"> IMID: 31227936 Nghiên cứu này đề cập đến một nhu cầu y tế chưa được đáp ứng để xuất một liệu pháp kết hợp ruxolitinib với chất ức chế HDAC (vorinostat) để tăng đáp ứng và thời gian điều trị ở bệnh nhân MPN bị đột biến gen JAK2. Bảng chứng tiền lâm sàng được trình bày trong nghiên cứu này cho thấy kết quả đầy hứa hẹn và hiệp hội đã được thêm vào biểu đồ kiến thức của chúng tôi để khuyến nghị liệu pháp kép này nếu có sẵn các lựa chọn điều trị nghiên cứu.
Sức chống cự	Gen, biến thể, liệu pháp	Đột biến BCL2A1 G101V như một cơ chế kháng venetoclax	<ul style="list-style-type: none"> PMID: 30514704 Nghiên cứu này mô tả một cơ chế kháng thuốc tái phát mới đối với việc điều trị bằng venetoclax đã được phát hiện ở những bệnh nhân vài tháng trước khi bệnh tiến triển. Điều này có thể thông báo cho bác sĩ về sự tái phát sắp xảy ra của bệnh nhân.
khuyh hướng	Gen, biến thể, điều kiện	Gene ETV6, đột biến R418G và tình trạng TẮT CẢ	<ul style="list-style-type: none"> PMID: 27365488, 25807284 Khuyh hướng ung thư có thể được phát hiện bằng cách giải trình tự thế hệ tiếp theo. Trong nghiên cứu này, đột biến tên lửa R418G đã được mô tả là đột biến đồng mầm trong các gia đình bị rối loạn tiểu cầu có tính chất gia đình với khuyh hướng mắc bệnh bạch cầu.

Hình 3 Các bằng chứng về sử dụng NLP trong sơ đồ tri thức có hệ thống đề xuất

Hệ thống mà họ đã nghiên cứu áp dụng có kiến trúc là PARSe của IBM, các thành phần của hệ thống gồm có: Kho lưu trữ nội dung, Bộ phân loại nghiên cứu, Bộ nhận dạng thực thể được gán nhãn và & Trình ánh xạ tìm kiếm Semantic searching và cuối cùng là Bộ phân tích & đề xuất tài liệu.



Hình 4 Mô hình đại học Oxford báo cáo về NLP đề xuất

Phương pháp đánh giá về độ chính xác dựa trên thang điểm F với kho dữ liệu 500 tài liệu được gán nhãn từ biểu đồ tri thức WfG.

Kết quả nghiên cứu hiệu xuất phân loại nằm trong tầm 88% đến 94%.

Study type	Precision	Recall	F-score
Clinical	0.96	0.92	0.94
Functional	0.93	0.91	0.92
Predisposition	0.93	0.89	0.91
Resistance	0.91	0.86	0.88
Preclinical	0.90	0.86	0.88

Hình 5 Thang điểm đánh giá về độ chính xác NLP đạt được

1.3 Nhiệm vụ của đề án

1.3.1 Tính cấp thiết và lý do chọn đề tài

Với xu hướng khách hàng tìm kiếm các sản phẩm trên các trang thương mại điện tử cho mình một cách nhanh chóng, với lượng data không lồ có thể kể đến ở nổi tiếng trên thế giới Amazon, Ebay, Alibaba hay ở Việt Nam chúng ta quá quen thuộc với các trang thương mại điện tử như Shopee, Tiki, Lazada, Chợ tốt, Thế giới di động, ... đòi hỏi các hệ thống này cần phải có các biện pháp nghiên cứu áp dụng AI để đề xuất các sản phẩm tới khách hàng ngay trên chính dữ liệu của chính mình.

Hệ thống đề xuất rất tiềm năng không chỉ áp dụng cho mỗi kênh bán hàng mà có thể áp dụng vào rất nhiều lĩnh vực khác trong đời sống xã hội như đề xuất bản tin, đề xuất video, đề xuất ca khúc mà người dùng hay nghe,...

Do đó đề tài “Ứng dụng AI xây dựng hệ thống bán hàng thông minh” mong muốn có thể giúp ích phần nào giải quyết được bài toán đề xuất sản phẩm mà ngày nay các ứng dụng lớn đã áp dụng cho các mạng xã hội như Facebook, Youtube, ... vào cho các hệ thống bán hàng truyền thống.

Hiện có rất nhiều phương pháp đề xuất khác nhau, mỗi phương pháp đều có ưu nhược điểm riêng. Trong quá trình nghiên cứu tìm hiểu tôi đã thấy đề tài của mình sẽ tập trung giải quyết bài toán, vấn đề trọng tâm là “*Xây dựng hệ thống tìm kiếm đề xuất cho khách hàng dựa trên xử lý ngôn ngữ tự nhiên NLP*” hiện thực việc ứng dụng AI vào hệ thống bán hàng.

1.3.2 Mục tiêu đề tài

Ứng dụng được xử lý ngôn ngữ tự nhiên **NLP** vào trong việc tối ưu phân tích tìm kiếm sản phẩm (sách) giúp nâng cao khả năng đề xuất gợi ý tới khách hàng.

1.3.3 Đối tượng và phạm vi nghiên cứu của đề tài

Đề tài hướng đến 3 nhóm người sử dụng hệ thống bán hàng thông minh gồm có:

Khách hàng: Có nhu cầu mua hàng, đặt hàng, tìm kiếm sản phẩm, sẵn sự kiện ưu đãi, nhận xét đánh giá các sản phẩm của shop bán, liên hệ mua hàng.

Nhân viên shop: Quản lý các thông tin liên quan đến sản phẩm, đơn hàng, kiểm tra số lượng sản phẩm, cập nhật hàng hóa, in báo cáo cho chủ shop.

Chủ cửa hàng shop - Admin: Có nhu cầu quản lý, quản trị tất cả các thông tin về sản phẩm, thông tin khách hàng, thông tin đơn hàng, các hoạt động chăm sóc khách hàng, chương trình sự kiện cho khách và các thiết lập cài đặt cho hệ thống shop.

1.3.4 Phương pháp nghiên cứu

1. Xây dựng và xử lý làm sạch, làm giàu dữ liệu học thuật
2. Xây dựng mô hình xử lý ngôn ngữ tự nhiên, áp dụng các thuật toán hợp lý cho việc cải thiện khả năng tìm kiếm, đề xuất.
3. Áp dụng, khai thác mô hình xử lý ngôn ngữ tự nhiên cho việc cải thiện khả năng tìm kiếm, đề xuất.

1.4 Kết cấu của đề tài

Báo cáo này gồm có lời mở đầu và 4 chương chính và phần tham khảo, 4 chương gồm có:

CHƯƠNG 1: TỔNG QUAN

Nêu các vấn đề liên quan đến việc đề xuất sản phẩm, vai trò, tầm ảnh hưởng, các nghiên cứu trước đây và các ứng dụng thực tế bắt gặp trong cuộc sống hiện nay.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

Trình bày các khái niệm liên quan đến bài toán áp dụng NLP để cải thiện tìm kiếm, đề xuất sách, các công nghệ về AI cũng như công nghệ ứng dụng Web có sử dụng để thực nghiệm hệ thống. Ngoài ra chương này còn có nêu rõ mô hình Model, thuật toán sẽ áp dụng, giải thích mô hình và thuật toán.

CHƯƠNG 3: THỰC TRẠNG VÀ PHÂN TÍCH VẤN ĐỀ

Chương này sẽ nêu rõ các công việc sẽ tiến hành liên quan đến cách thực hiện giải quyết bài toán tìm kiếm, đề xuất, kết quả trên lý thuyết, kết quả thực nghiệm đạt được.

CHƯƠNG 4: GIẢI PHÁP GIẢI QUYẾT

Nêu được kết luận chung về đề tài, khẳng định rõ vai trò, ý nghĩa những kết quả đạt được của đề tài cũng như đề xuất, đóng góp các ý tưởng, phương án mới nhằm cải thiện thêm chất lượng của đề tài.

1.5 Tóm tắt chương I

Tóm lại từ các bài viết khoa học về hệ thống đề xuất nói chung cũng như áp dụng xử lý ngôn ngữ tự nhiên **NLP** trong tìm kiếm đề xuất cho từng lĩnh vực nói riêng ta thấy được vai trò, tầm quan trọng nếu như không có các hệ thống này. Ý nghĩa thực tiễn được áp dụng trên quy mô dữ liệu lớn có khả năng đáp ứng được nhu cầu tìm kiếm chính sách hơn ở hệ thống bán hàng khi áp dụng vào so với các hệ thống bán hàng cũ truyền thống. Tuy việc áp dụng còn khá là manh mún nhưng nó cho thấy được tiềm năng thật sự giải quyết được bài toán tìm kiếm đề xuất.

Chương 2. CƠ SỞ LÝ THUYẾT

2.1 Các khái niệm

2.1.1 Hệ thống đề xuất là gì?

Hệ thống đề xuất (tiếng Anh là **Recommendation System**) là những hệ thống được thiết kế để hướng người dùng đến những đối tượng quan tâm, yêu thích, khi lượng thông tin quá lớn vượt quá khả năng xử lý của người dùng.

Theo tác giả Kantor và cộng sự của ông – đồng tác giả của cuốn sách Recommender Systems Handbook [100], hệ khuyến nghị là những công cụ phần mềm, kỹ thuật cung cấp những đề xuất các đối tượng có thể hữu ích với người dùng. Những đề xuất liên quan đến quyết định của người dùng như: sản phẩm nào nên mua, bài hát nào nên nghe, hay tin tức nào nên đọc.

2.1.2 Bài toán đề xuất

Đầu tiên ta xác định tập không gian người dùng và tập không gian đối tượng muốn đề xuất (ở đây là sách).

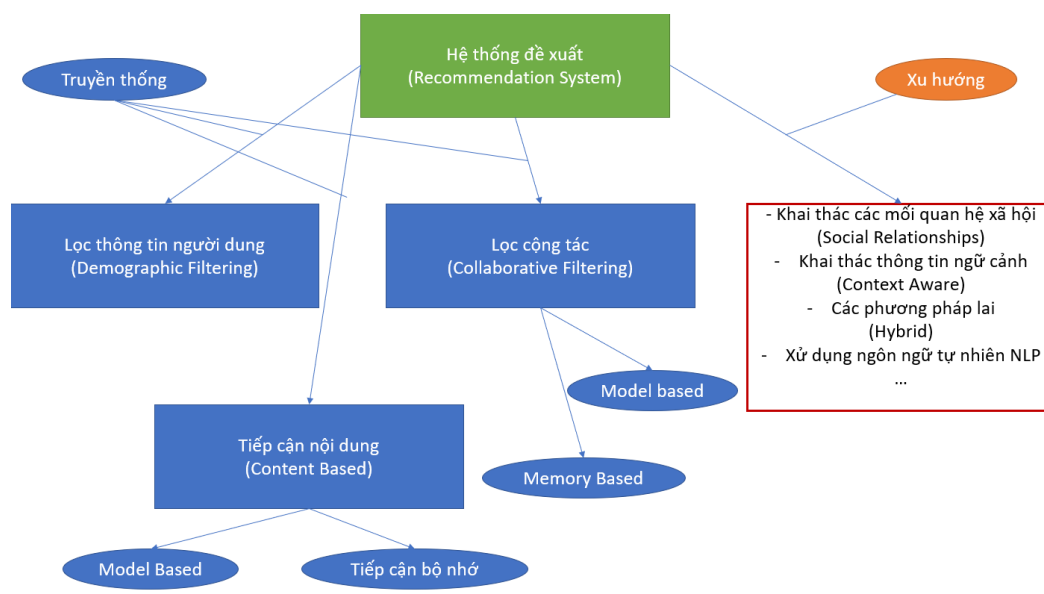
- $U = \{u_1, u_2, u_3, \dots, u_n\}$: không gian người dùng.
- $P = \{p_1, p_2, p_3, \dots, p_m\}$: không gian đối tượng đề xuất.

Hàm đánh giá hữu ích f là ánh xạ $f: U \times P \rightarrow R$, dùng để ước lượng mức độ hữu ích của $p \in P$ với $u \in U$. Với R là tập có thứ tự các số nguyên hoặc thực trong một khoảng nhất định.

Mục đích của hệ thống đề xuất là đi tìm hàm đánh giá hữu ích f , ước lượng giá trị của $f(u, p)$ (với $u \in U, p \in P$). Giá trị của $f(u, p)$ giúp tiên đoán u sẽ thích p nhiều hay ít, hay p hữu ích đối với u như thế nào. Đối với mỗi người dùng $u \in U$, hệ thống đề xuất cần chọn $TopN$ đối tượng $p \in P$ hữu ích nhất đối với người dùng u để đề xuất, $P_{TopN} = \langle p_1, p_2, \dots, p_{TopN} \rangle$, (với $TopN \ll m$). Việc chọn $TopN$ bao nhiêu là tùy thuộc vào nhu cầu thông tin của người dùng, cũng như mục đích cung cấp thông tin của hệ thống đề xuất.

Việc xây dựng hàm đánh giá hữu ích f và ước lượng giá trị hữu ích của các đối tượng đề xuất $p \in P$ với những người dùng $u \in U$ có thể thực hiện bằng nhiều phương pháp khác nhau như: dựa vào kinh nghiệm (heuristics), máy học, lý thuyết xấp xỉ, ...

2.1.3 Các cách tiếp cận phổ biến



Hình 6 Các phương pháp tiếp cận trong đề xuất

Với cách tiếp cận Content Based ta có các mặt hạn chế sau:

- Các khó khăn liên quan đến phân tích nội dung.
- Không thể đa dạng trong khuyến nghị (các đối tượng khuyến nghị ngoài lĩnh vực quan sát).
- Người dùng mới (mới sử dụng hệ thống lần đầu).

Với cách tiếp cận Collaborative Filtering ta có các mặt hạn chế sau:

- Ma trận đánh giá thưa (Matrix Factorization).
- Người dùng mới (cũng giống như Content Based là khi có người dùng mới hệ thống không đề xuất được – bắt buộc phải sử dụng hệ thống 1 thời gian).

2.1.4 Lựa chọn phương pháp tiếp cận để giải quyết vấn đề

Đề tài này áp dụng phương pháp trong xu hướng mới là xử lý ngôn ngữ tự nhiên NLP vào hệ thống đề xuất mặc dù phương pháp này nổi tiếng ở những vấn đề liên quan đến ngôn ngữ như nhận dạng chữ viết, văn bản, ... nên đối với xử lý ngôn ngữ tự nhiên NLP vào hệ thống tìm kiếm đề xuất khá mới mẻ, đầy tiềm năng và cũng đạt được kết quả ấn tượng như ở CHƯƠNG I đã trình bày về ứng dụng trong sơ đồ tri thức.

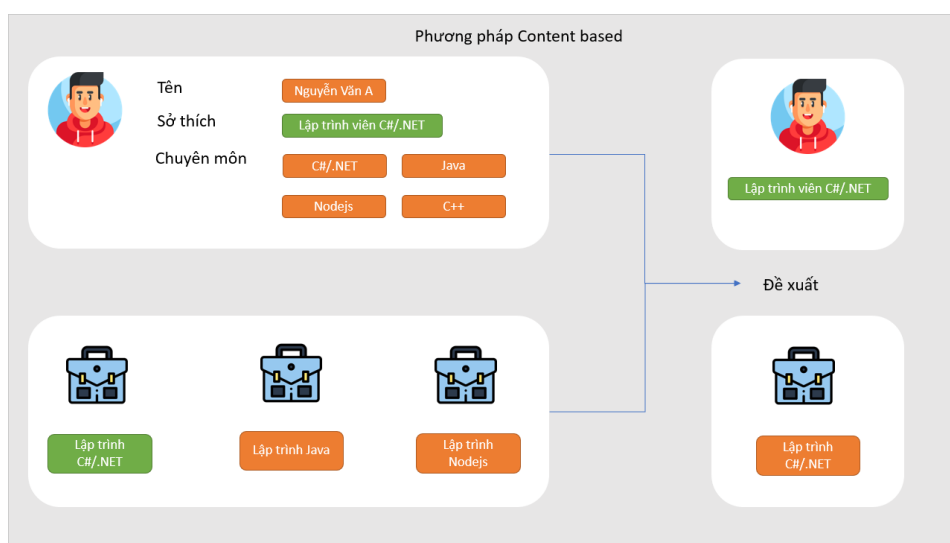
2.2 Phương pháp giải quyết vấn đề

2.2.1 Vấn đề đặt ra trong việc tìm kiếm, đề xuất bằng NLP

Xử lý ngôn ngữ tự nhiên, hay **NLP**, xử lý tốt văn bản thuần túy và lời nói hằng ngày của con người. Chúng ta có thể tìm thấy vô số trường hợp phân tích tình cảm hoặc xử lý tài liệu dựa trên NLP để giải quyết nhiệm vụ làm việc với ngôn ngữ viết. Những khả năng này cũng có thể được áp dụng cho các đề xuất nếu chúng ta hiểu đúng đầu vào và đầu ra của hệ thống bán hàng của mình để áp dụng vào.

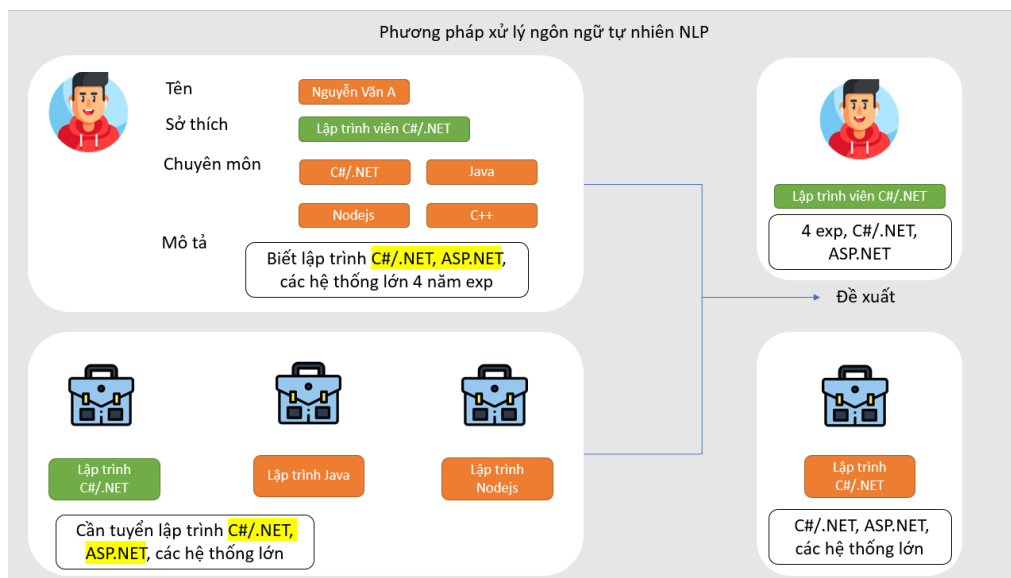
Bất kỳ hệ thống đề xuất nào cũng thực hiện một chức năng cơ bản cho người dùng. Nó phù hợp với mong đợi của người dùng với nội dung được phát hiện, bất kể đó có phải là yêu cầu dự định hay không. Các đề xuất được hình thành bằng cách tìm hiểu hoạt động trước đây của người dùng và chỉ định các danh mục cho các phần nội dung sẽ gọi là “bộ lọc”.

Một thực tế mà chúng ta đã biết là đầu vào của bất kỳ mô hình AI nào đều là dữ liệu số. Rõ ràng, khi nói đến dữ liệu như tuổi của người dùng hoặc số năm kinh nghiệm, các mô hình AI có thể dễ dàng hoạt động với các giá trị này. Để làm việc với loại dữ liệu này, cần thực hiện theo các bước theo phương pháp CRISP-DM, cụ thể là phân tích dữ liệu, hiểu cách tốt nhất để chuyển đổi dữ liệu thành định dạng đào tạo và thực hiện chuyển đổi. Ví dụ, hãy nghĩ về một mạng chuyên nghiệp như LinkedIn nơi người dùng áp dụng các bộ lọc để tìm kiếm cơ hội nghề nghiệp. Cách hoạt động của tìm kiếm, đề xuất dựa trên NLP



Hình 7 Phương pháp Content based

Nhưng, làm thế nào chúng ta có thể xử lý dữ liệu văn bản? Làm thế nào chúng ta có thể chuyển đổi từ thành dữ liệu số? Mô tả, bình luận và thông điệp lời nói thông tục có thể chỉ định các điểm dữ liệu như số năm kinh nghiệm với một công nghệ cụ thể có thể rất quan trọng để đối sánh chính xác.



Hình 8 Phương pháp xử lý ngôn ngữ tự nhiên NLP

Các phương pháp NLP có thể giúp người dùng cung cấp đầu vào tìm kiếm ở dạng miễn phí và không có bất kỳ hạn chế nào, theo yêu cầu của hệ thống, để nhận được câu trả lời cho yêu cầu mong muốn. Thông tin văn bản có thể lưu trữ một lượng lớn dữ liệu mà các bộ lọc thông thường không thể che được, như trong ví dụ trên.

2.2.2 Các phương pháp đề xuất áp dụng trong NLP

2.2.2.1 Phương pháp nhận dạng thực thể được đặt tên

Phương pháp *Nhận dạng thực thể được đặt tên* có bản chất là tìm các thực thể được đặt tên trong văn bản, chẳng hạn như: địa điểm, tổ chức, con người, v.v. Miễn là chúng ta nhắm mục tiêu các từ khóa nhất định, chúng tôi muốn đảm bảo mô hình chỉ ra các thực thể được đặt tên một cách chính xác để hệ thống đề xuất không nhầm tên tổ chức với vị trí hoặc công nghệ với dự án.

2.2.2.2 Phương pháp xét văn bản giống nhau

Phương pháp *tương tự văn bản* cung cấp hệ số tương tự của hai văn bản, so sánh các vectơ của cả hai văn bản. Trong trường hợp này, tính năng sẽ được sử dụng khi xây dựng mô hình AI là giá trị của hệ số này. Dưới đây là một ví dụ về công việc của mô

hình văn bản tương tự, trong đó điểm tương đồng là hệ số tương đồng của mô tả công việc với mô tả kỹ năng của ứng viên.

2.2.2.3 Phương pháp khai thác từ khóa

Phương pháp *khai thác từ khóa* sẽ giúp kiểm tra sự hiện diện của các từ khóa quan trọng nhất trong văn bản đầu vào, như hình dưới đây. Trích xuất từ khóa tự động rất hữu ích để phân tích cú pháp văn bản và biểu thị phần nào của câu hoặc các từ riêng biệt là phần chính. Hơn nữa, chúng ta có thể so sánh có bao nhiêu từ khóa được đáp ứng trong một văn bản mục tiêu giống như từ khóa được hiển thị trong ví dụ bên dưới.

2.2.3 Công nghệ phát triển AI

2.2.3.1 Numpy

Numpy là một thư viện mã nguồn mở để làm việc hiệu quả với các mảng. Được phát triển vào năm 2005 bởi Travis Oliphant, cái tên này là viết tắt của Numerical Python. Là một thư viện khoa học dữ liệu quan trọng trong Python, nhiều thư viện khác phụ thuộc vào nó.

NumPy cực kỳ phổ biến vì nó cải thiện đáng kể sự dễ dàng và hiệu suất làm việc với các mảng đa chiều.

Một số ưu điểm của Numpy:

1. Các phép toán trên các đối tượng ndarray của NumPy nhanh hơn tới 50 lần so với việc lặp qua các danh sách Python gốc bằng các vòng lặp. Hiệu quả đạt được chủ yếu là do NumPy lưu trữ các phần tử mảng ở một vị trí duy nhất được sắp xếp theo thứ tự trong bộ nhớ, loại bỏ sự dư thừa bằng cách có tất cả các phần tử cùng loại và tận dụng tối đa các CPU hiện đại. Lợi thế về hiệu quả trở nên đặc biệt rõ ràng khi hoạt động trên các mảng có hàng nghìn hoặc hàng triệu phần tử, đây là tiêu chuẩn khá phổ biến trong khoa học dữ liệu.
2. Nó cung cấp một cú pháp lập chỉ mục để dễ dàng truy cập các phần dữ liệu trong một mảng.
3. Nó chứa các hàm tích hợp giúp cải thiện chất lượng cuộc sống khi làm việc với mảng và toán học, chẳng hạn như các hàm dành cho đại số tuyến tính, phép biến đổi mảng và phép toán ma trận.

4. Nó yêu cầu ít dòng mã hơn cho hầu hết các hoạt động toán học so với danh sách Python gốc.

2.2.3.2 Pandas

Pandas có rất nhiều công dụng nên có thể liệt kê những thứ nó không thể làm thay vì những thứ nó có thể làm.

Công cụ này về cơ bản là ngôi nhà của dữ liệu của bạn. Thông qua gấu trúc, bạn làm quen với dữ liệu của mình bằng cách làm sạch, biến đổi và phân tích dữ liệu đó.

Ví dụ: giả sử bạn muốn khám phá tập dữ liệu được lưu trữ trong CSV trên máy tính của mình. Pandas sẽ trích xuất dữ liệu từ CSV đó vào DataFrame - về cơ bản là một bảng - sau đó cho phép bạn thực hiện những việc như:

- Tính toán số liệu thống kê và trả lời các câu hỏi về dữ liệu, như
 - Giá trị trung bình, trung bình, tối đa hoặc tối thiểu của mỗi cột là bao nhiêu?
 - Cột A có tương quan với cột B không?
 - Phân phối dữ liệu trong cột C trông như thế nào?
- Làm sạch dữ liệu bằng cách thực hiện những việc như xóa các giá trị bị thiếu và lọc các hàng hoặc cột theo một số tiêu chí
- Trực quan hóa dữ liệu với sự trợ giúp từ Matplotlib. Vẽ thanh, đường, biểu đồ, bong bóng, v.v.
- Lưu trữ dữ liệu đã được làm sạch, chuyển đổi trở lại thành CSV, tệp hoặc cơ sở dữ liệu khác

2.2.3.3 Scikit-learn



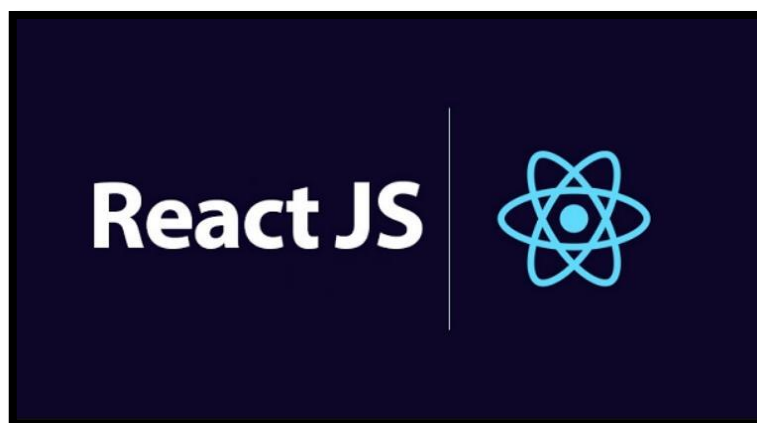
Hình 9 Thư viện Scikit-learn học sâu

Scikit-learn (Sklearn) là thư viện mạnh mẽ nhất dành cho các thuật toán học máy được viết trên ngôn ngữ Python. Thư viện cung cấp một tập các công cụ xử lý các bài toán machine learning và statistical modeling gồm: classification, regression, clustering, và dimensionality reduction.

Thư viện được cấp phép bản quyền chuẩn FreeBSD và chạy được trên nhiều nền tảng Linux. Scikit-learn được sử dụng như một tài liệu để học tập.

2.2.4 Công nghệ phát triển Web

2.2.4.1 ReactJS



Hình 10 React

ReactJS là một thư viện JavaScript có tính hiệu quả và linh hoạt để xây dựng các thành phần giao diện người dùng (UI) có thể sử dụng lại. ReactJS giúp phân chia các UI phức tạp thành các thành phần nhỏ (được gọi là component). Nó được tạo ra bởi Jordan Walke, một kỹ sư phần mềm tại Facebook. ReactJS ban đầu được phát triển và duy trì bởi Facebook sau đó được sử dụng trong các sản phẩm của mình như WhatsApp & Instagram.

ReactJS được dùng để xây dựng các ứng dụng Single Page Application (SPA). Một trong những điểm hấp dẫn của ReactJS là nó không chỉ được xây dựng bên phía client mà còn được sử dụng bên phía server.

Những khái niệm cơ bản của ReactJS:

Khi bắt đầu làm quen với ReactJS, chúng ta nên làm quen với các khái niệm cơ bản của nó trước, bởi các khái niệm này sẽ đi cùng với chúng ta trong suốt quá trình học tập và làm việc với ReactJS sau này.

Virtual Dom:

DOM là một Document Object Model và là cấu trúc trừu tượng của text. Các đoạn mã HTML được gọi là HTML DOM. Mỗi elements trong HTML là các nodes của DOM đó.

Virtual DOM được xây dựng dựa trên DOM thật, có một vài thuộc tính của DOM thật nhưng khi thay đổi Virtual DOM sẽ không thực hiện thay đổi trên màn hình giống như DOM thật.

Khi chúng ta làm việc với một DOM, khi một nodes thay đổi thì tất cả các nodes cũng phải thay đổi theo. Giả sử, chúng ta có một danh sách gồm 10 items, nếu chúng ta thay đổi 1 item thì DOM cũng thay đổi 9 items còn lại về trạng thái ban đầu của nó. Điều này sẽ làm cho tốc độ xử lý của SPA khá là chậm khi thay đổi các DOM liên tục và không khả thi khi xây dựng các ứng dụng lớn.

Khi chúng ta thực hiện render một JSX elementm, mỗi Virtual DOM object sẽ được cập nhật, khi Virtual DOM được cập nhật, ReactJS sẽ so sánh Virtual DOM với Virtual DOM trước đó để kiểm tra trước khi thực hiện việc cập nhật và sau đó cập nhật trên một thành phần DOM thật. Thay đổi DOM thật sẽ được hiển thị ra màn hình.

JSX:

JSX là viết tắt của JavaScript XML, nó cho phép bạn viết các đoạn mã HTML trong React một cách dễ dàng và có cấu trúc hơn. Về cú pháp cũng gần tương tự như HTML, giả sử ta có 1 đoạn mã HTML như sau:

```
1 | <p class="text">Đây là đoạn mã HTML</p>
```

Trong JSX thì sẽ được thay được viết như sau:


```
1 | <p className="text">Đây là đoạn mã HTML</p>
```

Component:

Khi ta làm việc với dự án lớn, UI có độ phức tạp cao chia các thành phần khác nhau. Việc chia nhỏ các thành phần trong UI là một điều cần thiết, các phần nhỏ như vậy được gọi là component, cho phép render các đoạn mã HTML, ... Trong ReactJS cách viết components được chia thành 2 loại:

Class Component

Function Component

Mỗi loại sẽ có ưu và nhược điểm khác nhau. Sự khác biệt rõ nhất là cú pháp. Một functional component thực tế chỉ là một hàm JavaScript đơn giản chấp nhận các props như là một argument và trả về một phần tử React.

Một class component yêu cầu chúng ta phải cần kế thừa từ phần tử React Component và tạo một function render trả về một phần tử React. Điều này sẽ đòi hỏi chúng ta phải code nhiều hơn nhưng sẽ cung cấp một số lợi ích mà chúng ta thấy sau này.

Props

Chúng ta cần sử dụng props khi sử dụng cùng một Component nhưng với những thông số khác nhau.

```
// Functional
const Hello = (props) => {
  return (
    <div className="App">
      Hello {props.name}
    </div>
  );
}

// Class component
class Hello extends React.Component {
  render() {
    return (
      <div className="App">
        Hello {this.props.name}
      </div>
    );
  }
}

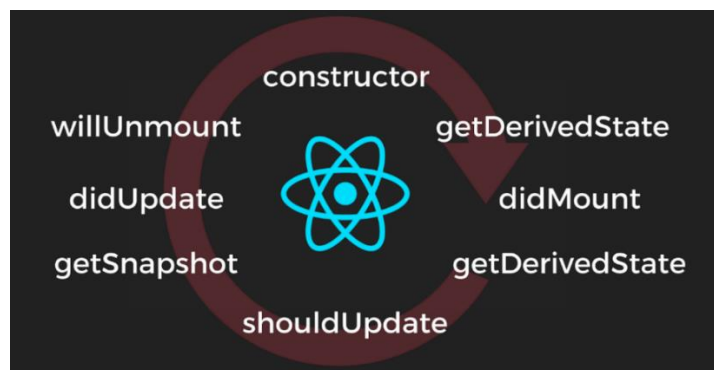
// Sử dụng
<Hello name="You" />
```

Hình 11 React function component vs class component

Với props trong Class component được xem như giá trị truyền vào cho hàm khởi tạo class. Còn props trong Function Component thì được xem như là giá trị truyền vào hàm pure function. Tuy nhiên phiên bản react 16.8 xuất hiện thuật ngữ hoàn toàn mới đó là hooks. Với việc xuất hiện hook thì việc sử dụng state trong functional hoàn toàn dễ dàng.

React Lifecycle:

React Lifecycle là một vòng đời của component, khi chúng ta tiến hành render một component thì ReactJS thực hiện nhiều tiến trình khác nhau, các tiến trình này được lặp đi lặp lại đối với các component.



Hình 12 Vòng React component

Giả sử khi một component được gọi trước tiên nó sẽ cài đặt các props và state, sau đó tiến hành mounting, update, unmounting, ... việc tham gia vào quá trình này bạn cần sử dụng đến các hàm hỗ trợ của lifecycle.

Ưu điểm:

ReactJS cực kì hiệu quả: ReactJS tạo ra cho chính nó DOM ảo – nơi mà các component thực sự tồn tại trên đó. Điều này sẽ giúp cải thiện hiệu suất rất nhiều. ReactJS cũng tính toán những thay đổi nào cần cập nhật lên DOM và chỉ thực hiện chúng, Điều này giúp ReactJS tránh những thao tác cần trên DOM mà nhiều chi phí.

Ví dụ:

```
<title>Hello React</title>
<script src="https://fb.me/react-0.13.2.js"></script>
<script src="https://fb.me/JSXTransformer-0.13.2.js"></script>
<script src="https://code.jquery.com/jquery-2.1.3.min.js"></script>
<div id="content"></div>
<script type="text/jsx">
  var Hoge = React.createClass({
    getInitialState(){
      return{
        style:{
          color: "#ccc",
          width: 200,
          height: 100
        }
      };
    },
    onChange(){
      var style = _.clone(this.state.style);
      style.color = "#ddd";
      this.setState({ style: style});
    },
    render(){
      return(
        <div style={this.state.style} onClick={this.onChange}>xxx</div>
      );
    }
  });
  React.render(
    <Hoge />,
    document.getElementById('content')
  );
</script>
```

Hình 13 Ví dụ về `React.createClass`

Ở ví dụ trên, chúng ta đã định nghĩa một component Hoge bằng method `React.createClass(...)`, sau đó render ra View:

```
React.render(, document.getElementById('content'));
```

Khi đó ta có thể viết các methods, các functions tác động lên component Hoge (thay đổi Model data), View sẽ lập tức được cập nhật cho dù ta không phải động chạm trực tiếp gì vào phần tử DOM trên View.

ReactJS giúp việc viết các đoạn code JS dễ dàng hơn: Nó dùng cú pháp đặc biệt là JSX (JavaScript mở rộng) cho phép ta trộn code HTML và JavaScript. Ta có thể thêm vào các đoạn HTML vào trong hàm render không cần phải nối chuỗi. Đây là đặc tính thú vị của ReactJS. Nó sẽ chuyển đổi các đoạn HTML thành các hàm khởi tạo đối tượng HTML bằng bộ biến đổi JSX.

Nó có nhiều công cụ phát triển

Render tầng server

Làm việc với vấn đề test giao diện

Hiệu năng cao đối với các ứng dụng có dữ liệu thay đổi liên tục, dễ dàng cho bảo trì và sửa đổi.

Nhược điểm:

ReactJS chỉ phục vụ cho tầng View. React chỉ là View Library nó không phải là một MVC framework như những framework khác. Đây chỉ là thư viện của Facebook giúp render ra phần view. Vì thế React sẽ không có phần Model và Controller, mà phải kết hợp với các thư viện khác. React cũng sẽ không có two-way binding hay là Ajax.

Tích hợp ReactJS vào các framework MVC truyền thống yêu cầu cần phải cấu hình lại.

Khó tiếp cận với người mới học Web.

2.2.4.2 NodeJS



Hình 14 Nodejs

Khái niệm:

NodeJS là một nền tảng (Platform) phát triển độc lập được xây dựng trên V8 JavaScript Engine – trình thông dịch thực thi mã JavaScript giúp chúng ta có thể xây dựng được các ứng dụng web.

NodeJS có thể chạy trên nhiều nền tảng hệ điều hành khác nhau từ Window cho tới Linux, OS x nên đó cũng là một lợi thế. NodeJS cung cấp các thư viện phong phú ở dạng Javascript Module khác nhau giúp đơn giản hóa việc lập trình và giảm thời gian ở mức thấp nhất.

Ý tưởng chính của NodeJS là sử dụng non-blocking, hướng sự vào ra dữ liệu thông qua các tác vụ thời gian thực một cách nhanh chóng. NodeJS đưa ra là sử dụng luồng đơn (Single - Threaded), kết hợp non – blocking I/O để thực thi các request, cho phép hỗ trợ hàng chục ngàn kết nối đồng thời. Do đó, NodeJS có khả năng mở rộng nhanh chóng, khả năng xử lý một số lượng lớn cá kết nối đồng thời bằng thông lượng cao.

Cấu trúc:

Module:

Module giống như các thư viện JavaScript sử dụng trong ứng dụng Node.js application bao gồm một bộ chức năng. Để đưa một module vào ứng dụng Node.js thì phải sử dụng hàm `require()` function với dấu ngoặc đơn chứa tên các module.

NodeJS có rất nhiều module cung cấp các chức năng cơ bản cần thiết cho ứng dụng web. Ví dụ như các module trong bảng sau:

Core Modules	Description
http	Includes classes, methods and events to create Node.js http server
util	Includes utility functions useful for developers
fs	Includes events, classes, and methods to deal with file I/O operations
url	Includes methods for URL parsing
querystring	Includes methods to work with query string
stream	Includes methods to handle streaming data
zlib	Includes methods to compress or decompress files

Hình 15 Các module trong Nodejs

Ưu điểm:

Có tốc độ xử lý nhanh nhờ cơ chế xử lý bất đồng bộ (non-blocking).

Giúp bạn dễ dàng mở rộng khi có nhu cầu phát triển website.

Nhận và xử lý nhiều kết nối chỉ với một single-thread. Nhờ đó, hệ thống xử lý sẽ sử dụng ít lượng RAM nhất và giúp quá trình xử lý NodeJS nhanh hơn rất nhiều.

Có khả năng xử lý nhiều Request/s cùng một lúc trong thời gian ngắn nhất.

Có khả năng xử lý hàng ngàn Process cho hiệu suất đạt mức tối ưu nhất.

Phù hợp để xây dựng những ứng dụng thời gian thực như các ứng dụng chat, mạng xã hội...

Nhược điểm:

NodeJS gây hao tốn tài nguyên và thời gian. NodeJS được viết bằng C++ và JavaScript nên khi xử lý cần phải trải qua một quá trình biên dịch. Nếu cần xử lý những ứng dụng tốn tài nguyên CPU thì không nên sử dụng NodeJS.

2.2.4.3 Express

Khái niệm:

Express là một framework giành cho nodejs. Nó cung cấp cho chúng ta rất nhiều tính năng mạnh mẽ trên nền tảng web cũng như trên các ứng dụng di động. Express hỗ trợ các phương thức HTTP và middleware tạo ra một API vô cùng mạnh mẽ và dễ sử dụng. Có thể tổng hợp một số chức năng chính của express như sau:

Thiết lập các lớp trung gian để trả về các HTTP request.

Định nghĩa các router cho phép sử dụng với các hành động khác nhau dựa trên phương thức HTTP và URL.

Cho phép trả về các trang HTML dựa vào tham số.

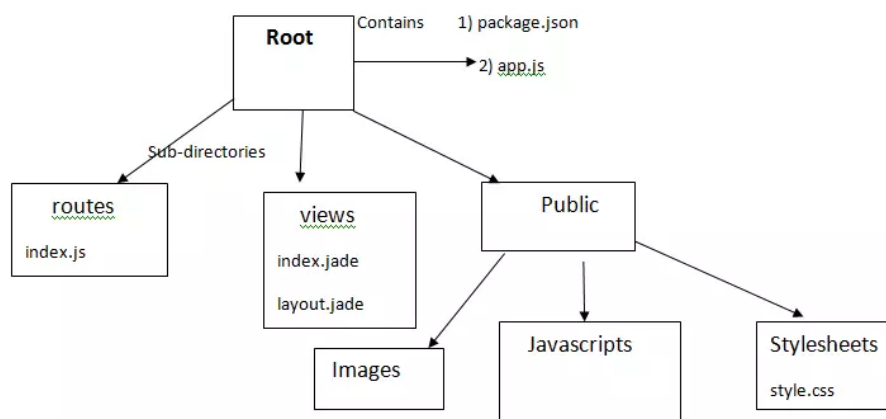
ExpressJS hay còn được viết là Express js, Express.js. Đây là một framework mã nguồn mở miễn phí cho node.js. Express.js được sử dụng trong thiết kế xây dựng các ứng dụng web một cách đơn giản và nhanh chóng.

ExpressJS là một framework nhỏ nhưng linh hoạt được xây dựng trên nền tảng của NodeJS. Nó cung cấp tính năng mạnh mẽ để phát triển web hoặc mobile.

ExpressJS có vô số các package hỗ trợ.

ExpressJS cung cấp thêm về các tính năng (feature) để lập trình tốt hơn.

Cấu trúc ExpressJS:



Hình 16 Cấu trúc ExpressJS

Root:

“app.js” chứa các thông tin về cấu hình, khai báo, các định nghĩa, ... để ứng dụng của chúng ta chạy ok.

“package.json” chứa các package cho ứng dụng chạy.

Folder “routes”: chứa các route có trong ứng dụng.

Folder “view”: chứa view/template cho ứng dụng.

Folder “public”: chứa các file css, js, images, ... cho ứng dụng.

2.2.4.4 MongoDB



Hình 17 MongoDB

MongoDB là một hệ quản trị cơ sở dữ liệu NoSQL mã nguồn mở đa nền tảng viết bằng C++.

Bản ghi trong MongoDB được lưu trữ dạng một dữ liệu văn bản (Document), là một cấu trúc dữ liệu bao gồm các cặp giá trị và trường tương tự như các đối tượng JSON.

2.2.4.5 Công cụ Postman

Postman là một loại công cụ cho phép người dùng có thể thao tác với API, mà trong đó phổ biến nhất là REST. Với thử nghiệm API thì Postman là một trong những công cụ phổ biến vì được thực nghiệm nhiều nhất. Nhờ Postman lập trình viên có thể gọi Rest API mà không cần phải viết bất kỳ dòng code nào.

Postman có khả năng hỗ trợ mọi phương thức HTTP bao gồm: POST, PUT, DELETE, PATCH, GET... Ngoài ra, Postman còn cho phép lập trình viên lưu lại lịch sử của các lần request nên vô cùng tiện lợi cho nhu cầu sử dụng lại.

Người dùng có thể post dữ liệu dưới dạng text, form, json...

Kết quả hiện về và trả về sẽ dưới dạng hình ảnh, text, XML, JSON...

Có thể hỗ trợ authorization.

Có thể thực hiện thay đổi header của các request.

2.2.4.6 Visual Studio Code (VS Code)

Là một trình biên tập lập trình code miễn phí dành cho Windows, Linux và macOS, Visual Studio Code được phát triển bởi Microsoft. Nó được xem là một sự kết hợp hoàn hảo giữa IDE và Code Editor.

Visual Studio Code hỗ trợ chức năng debug, đi kèm với Git, có syntax highlighting, tự hoàn thành mã thông minh, snippets, và cải tiến mã nguồn. Nhờ tính năng tùy chỉnh, Visual Studio Code cũng cho phép người dùng thay đổi theme, phím tắt, và các tùy chọn khác.

Các lý do nên sử dụng Visual Studio Code:

Hỗ trợ đa nền tảng: Windows, Linux, Mac.

Hỗ trợ đa ngôn ngữ: C/C++, C#, F#, Visual Basic, HTML, CSS, JavaScript, JSON.

Ít dung lượng. Tính năng mạnh mẽ. Intellisense chuyên nghiệp, giao diện thân thiện

Với việc cải tiến và áp dụng các công nghệ mới, Visual Studio Code đã chứng minh cho các lập trình viên thấy hiệu quả của mình. Ngoài các Developers, thì Testers, Data Administrators và tất cả những ai quan tâm đến lập trình đều có thể sử dụng.

2.3 Tóm tắt chương 2

Chương 2 của đề tài cung cấp các thông tin về khái niệm, bài toán đề xuất, giải pháp trong NLP, cách chúng hoạt động như thế nào? Ưu nhược điểm trong mỗi phương pháp xử lý NLP và ví dụ minh họa. Ngoài ra chương cũng cung cấp thông tin các công nghệ có mặt trong đề tài được sử dụng để làm ra sản phẩm thực nghiệm.

Chương 3. THỰC TRẠNG VÀ PHÂN TÍCH VẤN ĐỀ

3.1 Các bước thực hiện đề tài

Đề tài này thực nghiệm sẽ chia làm 2 giai đoạn chính là:

Giai đoạn 1: Phát triển các chức năng cho website bán sách.

Giai đoạn 2: Tích hợp xử lý ngôn ngữ tự nhiên NLP vào website bán sách.

3.1.1 *Phát triển website bán sách*

3.1.1.1 Khảo sát mô tả thực trạng hệ thống

Cửa hàng bán sách bán lẻ nhiều loại sách khác nhau. Nguồn sách do các nhà cung cấp trong và ngoài nước cung cấp. Khách hàng của cửa hàng là người tiêu dùng phù hợp với học sinh, sinh viên và các độc giả xa gần khác. Cửa hàng có nhiều loại sách như: truyện, sách giáo khoa, tiểu thuyết trinh thám,... Vì vậy cần một website bán hàng để tiếp cận được nhiều khách hàng và giúp cho việc quản lý cửa hàng được dễ dàng hơn. Các hoạt động chính của cửa hàng gồm bán hàng, quản lý khách hàng, thống kê.

3.1.1.2 Phân tích yêu cầu hệ thống

Các yêu cầu chức năng:

Website bán sách SmartShop có 23 chức năng:

- **Đăng ký:** Cho phép khách hàng đăng ký thành viên vào hệ thống.
- **Đăng nhập:** Cho phép khách hàng thành viên, nhân viên chăm sóc khách hàng, quản trị viên đăng nhập vào hệ thống (có tích hợp Google Login).
- **Đăng xuất:** Cho phép quản trị viên, nhân viên chăm sóc khách hàng, khách hàng thành viên đăng xuất khỏi hệ thống.
- **Quản lý tài khoản:** Cho phép khách hàng thành viên xem thông tin tài khoản, cập nhật thông tin cá nhân gồm: Họ tên, địa chỉ, giới tính, profile cá nhân.
- **Quản lý liên hệ (Form or bong bóng chat):** Cho phép khách hàng xem câu hỏi, khách hàng thành viên xem câu hỏi, đăng câu hỏi, sửa câu hỏi khi quản trị viên và nhân viên chăm sóc khách hàng chưa trả lời và xóa câu hỏi.
- **Đánh giá - Comment:** Cho phép khách hàng xem đánh giá, khách hàng thành viên xem đánh giá và thực hiện đánh giá.

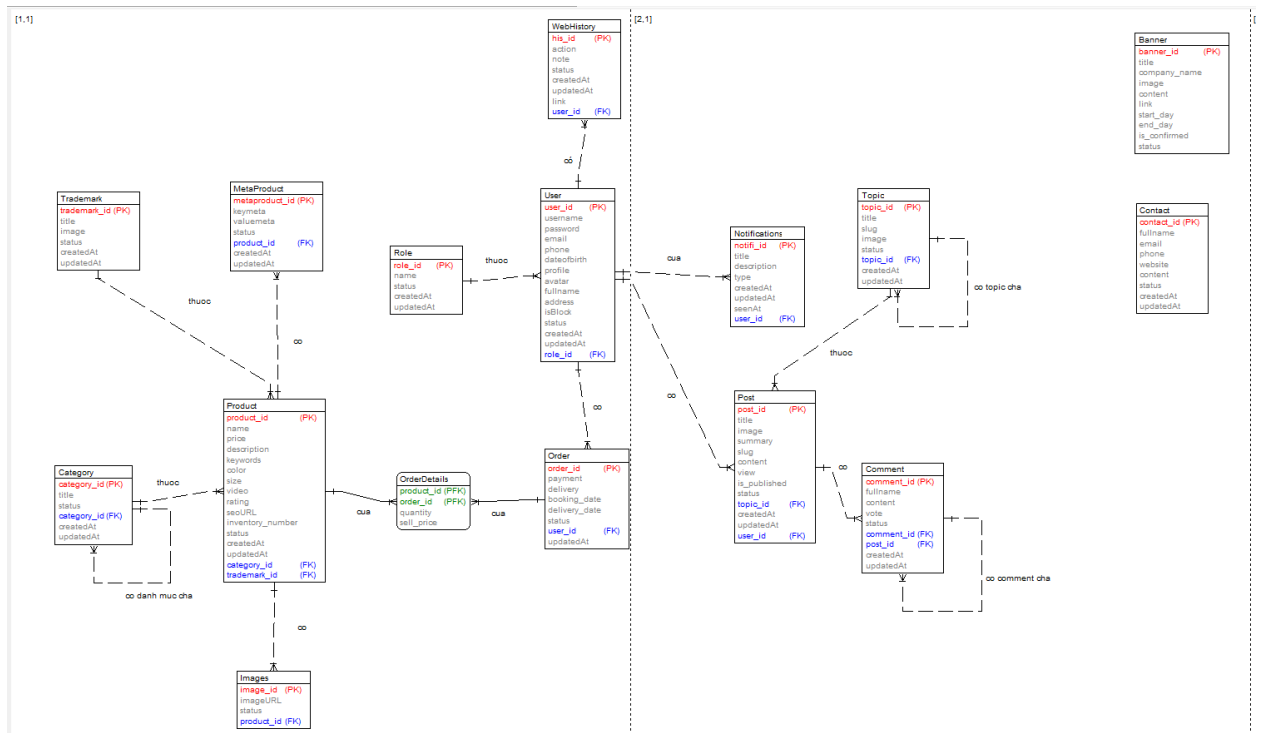
- **SEO URL Sản phẩm & Blog:** tối ưu hóa công cụ tìm kiếm với SEO Url cho đến nội dung.
- **Tìm kiếm:** Cho phép người dùng tìm kiếm sách, đề xuất sách theo lịch sử đã mua.
- **Đề xuất sách:** Đề xuất các cuốn sách theo NLP sẽ được trình bày trong giai đoạn số 2.
- **Quản lý lịch sử xem hàng:** Cho phép khách hàng thành viên xem lịch sử xem hàng.
- **Quản lý giỏ hàng:** Cho phép khách hàng, khách hàng thành viên xem giỏ hàng, thêm sản phẩm vào giỏ, cập nhật số lượng sản phẩm trong giỏ, xóa sản phẩm khỏi giỏ hàng.
- **Đặt hàng:** Cho phép khách hàng tiến hành đặt hàng các sản phẩm trong giỏ hàng.
- **Quản lý đơn đặt hàng:** Cho phép khách hàng thành viên xem tình trạng đơn đặt hàng như: Đơn đặt hàng chưa duyệt, đã duyệt, đang giao, đã nhận, đã hủy và thực hiện chức năng hủy đơn hàng khi đơn đặt hàng chưa được duyệt và chức năng xác nhận đã nhận được hàng.
- **Xử lý đơn đặt hàng:** Cho phép quản trị viên xem danh sách đơn đặt hàng chưa duyệt, đã duyệt nhưng chưa giao, đã hoàn thành và thực hiện chức năng duyệt đơn đặt hàng, giao hàng, tình trạng đơn hàng xác thực hay đã hủy.
- **Quản lý câu trả lời:** Cho phép Nhân viên chăm sóc khách hàng, quản trị viên thực hiện các chức năng: Xem danh sách hỏi đáp, trả lời câu hỏi, sửa câu trả lời.
- **Quản lý danh mục sản phẩm:** Cho phép quản trị viên xem danh sách danh mục sản phẩm và thực hiện chức năng tìm kiếm, thêm danh mục sản phẩm, cập nhật thông tin danh mục sản phẩm, khóa và kích hoạt danh mục sản phẩm.
- **Quản lý sản phẩm:** Cho phép quản trị viên xem danh sách sản phẩm và thực hiện chức năng tìm kiếm, thêm sản phẩm, cập nhật thông tin sản phẩm, khóa và kích hoạt sản phẩm.

- **Quản lý tài khoản hệ thống:** Cho phép quản trị viên xem danh sách tài khoản hệ thống và thực hiện chức năng tìm kiếm, thêm, cập nhật thông tin, khóa và kích hoạt tài khoản Admin.
- **Quản lý thể loại sách, nhà xuất bản:** Cho phép quản trị viên xem danh sách thể loại sách, nhà xuất bản cũng như nhu cầu và thực hiện chức năng tìm kiếm, thêm, cập nhật thông tin hãng, khóa và kích hoạt hãng.
- **Quản lý tin tức – blog – quảng cáo:** Cho phép quản trị viên viết bài giúp cho việc SEO website bán hàng cũng như đặt vị trí quảng cáo trong chi tiết bài viết.
- **Phân quyền:** Cho phép quản trị viên phân quyền cho loại tài khoản.
- **Thông kê:** Chức năng cho phép quản trị viên thực hiện thống kê hàng tồn, khách hàng tiềm năng, sản phẩm bán chạy, đơn hàng, nhà cung cấp sách,...
- **Đổi mật khẩu:** Cho phép người dùng đổi mật khẩu khi cần thiết.

3.1.1.3 Thiết kế hệ thống, sơ đồ liên quan

❖ **Sơ đồ ERD**

Sử dụng Case Studio 2 để thiết kế ERD (Entity – Relationship Diagram). Sử dụng Case Studio ta có thể generate dễ dàng các đoạn script SQL phục vụ cho nhiều loại database khác nhau (MySQL, SQLServer, DB2, Oracle,...).



Hình 18 Sơ đồ ERD hệ thống shop

Cơ sở dữ liệu kết hợp giữa hệ thống Database của MongoDB được thiết kế đáp ứng vừa đủ những chức năng mà hệ thống bán hàng Shop cần có (mục đích lưu trữ).

Model này sẽ được migrations giữa các loại database khác nếu Admin có cầu chuyển đổi sang ví dụ từ MongoDB chuyển qua, MySQL, MS SQL Server.

❖ Phân tích thành phần trong Database

Cơ sở dữ liệu được thiết kế và lưu trữ mặc định trên MongoDB.

Bảng 1 User

Tên thuộc tính	Định dạng (MongoDB)	Khóa chính	Khóa phụ	Mô tả
User_id	ObjectID	X		Mã người dùng
Username	String			Tên tài khoản
password	String			Mật khẩu
Email	String			Email
Phone	String			Số điện thoại

Dateofbirth	Date			Ngày sinh
Profile	String			Hồ sơ cá nhân tóm tắt
Avatar	String			ảnh đại diện
Fullname	String			Họ và tên
Address	String			Địa chỉ
isBlock	Boolean			Trạng thái khóa tài khoản
Status	Boolean			Trạng thái xóa (thùng rác)
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa
Role_id	ObjectID		X	Mã quyền

Bảng 2 OrderDetails

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Product_id	ObjectID	X	X	Mã sản phẩm
Order_id	ObjectID	X	X	Mã đơn hàng
Quantity	Number			Số lượng
Sell_price	Number			Giá bán

Bảng 3 Order

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Order_id	ObjectID	X		Mã đơn hàng
Payment	Boolean			Trạng thái thanh toán

Delivery	Boolean			Trạng thái giao hàng
Booking_date	Date			Ngày giao
Delivery_date	Date			Ngày nhận
Status	Boolean			Trạng thái xóa (thùng rác)
updatedAt	Date			Ngày sửa
User_id	ObjectID		X	Mã người dùng

Bảng 4 Product

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Product_id	ObjectID	X		Mã sản phẩm
Name	String			Tên sản phẩm
Price	Number			Giá
Description	String			Mô tả sản phẩm
Keywords	String			Từ khóa SEO
Colors	Array			Màu
Sizes	Array			Kích thước
Videos	Array			Video
Rate	Object			Đánh giá *
SeoURL	String			Link SEO
Inventory_number	Number			Số lượng tồn
Status	Boolean			Trạng thái xóa (thùng rác)

createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa
Category_id	ObjectID		X	Mã loại sản phẩm
Trademark_id	ObjectID		X	Mã thương hiệu

Bảng 5 Trademark

Tên thuộc tính	Định dạng	Khóa chính	Khóa ngoại	Mô tả
Trademark_id	ObjectID	X		Mã thương hiệu
Title	String			Tên thương hiệu
Image	String			Hình ảnh
Status	Boolean			Trạng thái xóa (thùng rác)
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 6 Category

Tên thuộc tính	Định dạng	Khóa chính	Khóa ngoại	Mô tả
Category_id	ObjectID	X		Mã loại sản phẩm
Title	String			Tên loại sản phẩm
Status	Boolean			Trạng thái xóa (thùng rác)

Parent_ category_id	ObjectID		X	Mã loại sản phẩm cha
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 7 Meta_product

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Metaproduct_id	ObjectID	X		Mã meta mở rộng sản phẩm
Keymeta	String			Mã
valuemeta	String			Giá trị key
Status	Boolean			Trạng thái xóa
Product_id	ObjectID		X	Mã sản phẩm
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 8 Comment

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Comment_id	ObjectID	X		Mã nhận xét
Fullname	String			Họ tên người comment
Content	String			Nội dung comment
Vote	Object			Đánh giá *
Status	Boolean			Trạng thái xóa
Parent_	ObjectID		X	Mã nhận xét cha

comment_id				
Post_id	ObjectID		X	Mã bài viết
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 9 WebHistory

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
His_id	ObjectID	X		Mã lịch sửa thao tác
Action	nvarchar(255)			Thao tác
Note	nvarchar(200)			Ghi chú
Status	varchar(100)			Trạng thái xóa lịch sử
Link	varchar(100)			Link liên kết
createdAt	smalldatetime			Ngày tạo
updatedAt	smalldatetime			Ngày sửa
User_id	bit		X	Mã người dùng

Bảng 10 Post

Tên thuộc tính	Định dạng	Khóa chính	Khóa ngoại	Mô tả
Post_id	ObjectID	X		Mã bài viết
Title	String			Tiêu đề
Image	String			Hình ảnh

Summary	String			Tóm tắt
Slug	String			Link SEO
Content	String			Nội dung
View	Number			Số lượng xem
Is_published	Boolean			Trạng thái xuất bản
Status	Boolean			Trạng thái xóa
Topic_id	ObjectID		X	Mã chủ đề
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa
User_id	ObjectID		X	Mã người đăng

Bảng 11 Topic

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Topic_id	ObjectID	X		Mã chủ đề
Title	String			Tên chủ đề
Slug	String			Link SEO
Image	String			Hình ảnh
Status	Boolean			Trạng thái
Parent_topic_id	ObjectID		X	Mã chủ đề cha
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 12 Contact

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Contact_id	ObjectID	X		Mã liên hệ

Fullname	String			Họ tên liên hệ
Email	String			Email
Phone	String			Số điện thoại
Website	String			Website
Content	String			Nội dung liên hệ
Status	Boolean			Trạng thái xóa
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 13 Role

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Role_id	ObjectID	X		Mã quyền
Name	String			Tên quyền
Status	Boolean			Trạng thái xóa
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa

Bảng 14 Notification

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Notify_id	ObjectID	X		Mã thông báo
Title	String			Tên thông báo
Description	String			Mô tả
Type	String			Dạng thông báo
seenAt	Date			Ngày xem

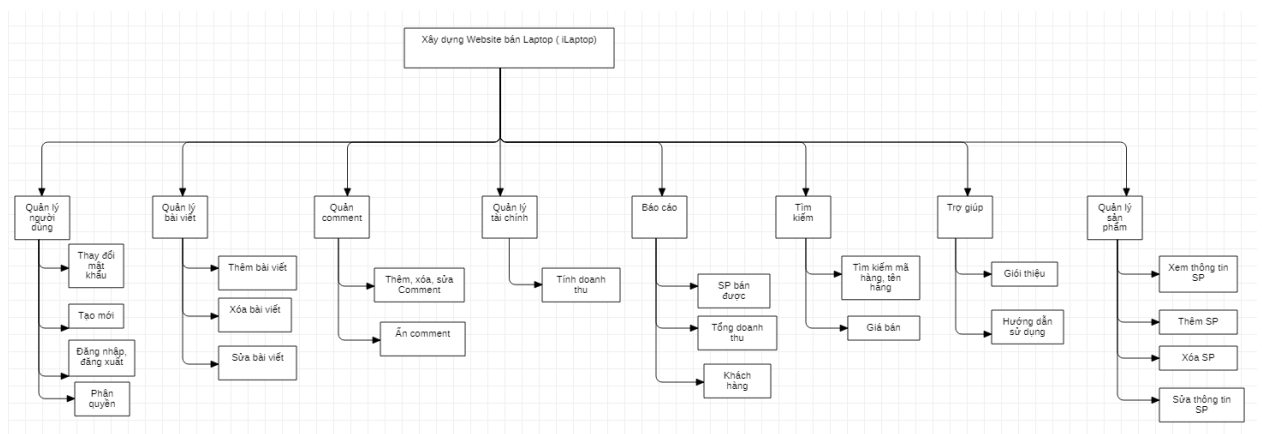
createdAt	Date			Ngày tạo
updatedAt	Date			Ngày sửa
User_id	ObjectID		X	Mã người dùng

Bảng 15 Banner

Tên thuộc tính	Định dạng	Khóa chính	Khóa phụ	Mô tả
Banner_id	ObjectID	X		Mã banner
Title	String			Tên banner
Company_name	String			Tên công ty
Image	String			Hình ảnh
Content	String			Nội dung
Link	String			Link liên kết
Start_day	Date			Ngày bắt đầu
End_day	Date			Ngày kết thúc
Is_confirmed	Boolean			Trạng thái xác nhận
updatedAt	Date			Ngày sửa

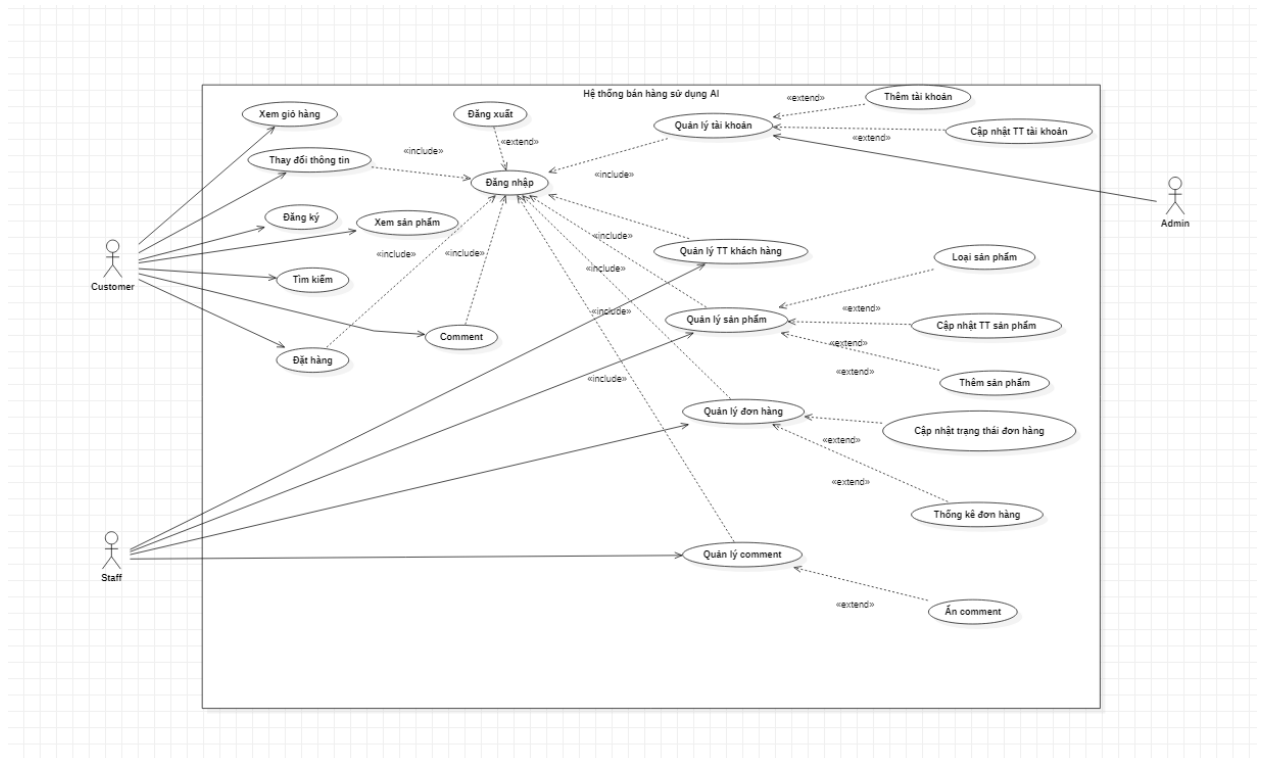
❖ Sơ đồ BFD

Hệ thống bán sách sẽ gồm có các chức năng chính cơ bản như sơ đồ dưới sau:



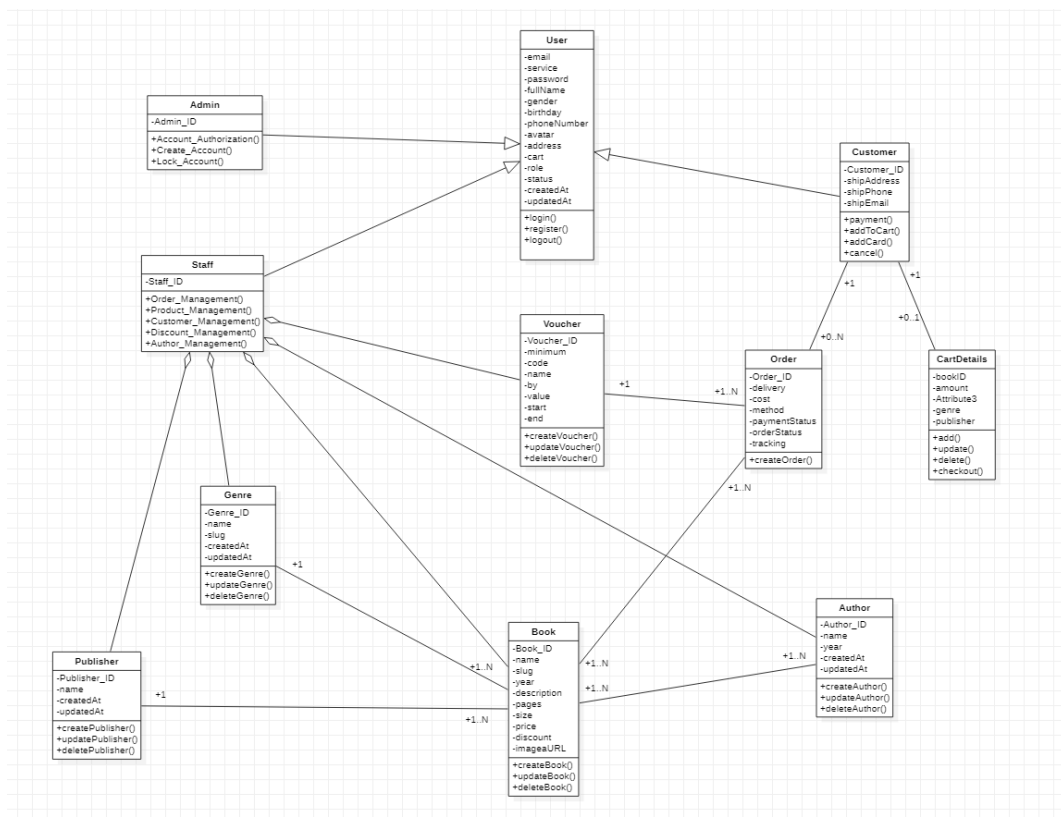
Hình 19 Sơ đồ BFD

❖ Sơ đồ Usecase



Hình 20 Sơ đồ Usecase hệ thống

❖ Sơ đồ Class Diagram



Hình 21 Sơ đồ Classdiagram

3.1.1.4 Thực hiện code, test chức năng cơ bản website

Hệ thống bán hàng sử dụng Nodejs làm nền tảng để chạy ExpressJS cung cấp các API cho frontend sử dụng ReactJS để render nội dung ra giao diện người dùng cụ thể là Browser. Sau khi viết 1 API cụ thể ví dụ lấy toàn bộ danh sách Sách thì sẽ dùng Postman để kiểm tra đầu API này xem có đúng với yêu cầu hay không.

Ngoài ra ở phần giao diện Website để kiểm tra dữ liệu có hợp lệ hay thì cùng với backend đều có các thao tác validation dữ liệu.

Để test chức năng thì có các unit test được viết ra để kiểm thử.

3.1.1.5 Kiểm tra và bảo trì

Sau khi hoàn thành các chức năng của hệ thống thì sau khi hệ thống hoạt động 1 thời gian thì sẽ phải kiểm tra và bảo trì cả về ứng dụng lẫn sourcecode. Sourcecode phải phân tách thành các module với logic business rõ thì mới có thể dễ dàng mở rộng hay sửa chữa mà không ảnh hưởng tới module khác.

3.1.2 Xây dựng triển khai mô hình sử dụng NLP cho website bán sách

Đầu tiên trước khi đi vào mô hình giải quyết bài toán đề xuất thì chúng ta nên biết 2 phần quan trọng sau đây.

3.1.2.1 Các bài toán cơ bản trong NLP mà người mới nên biết

- Mô hình hóa ngôn ngữ (Language modelling)
- Phân loại văn bản (Text classification)
- Trích xuất thông tin (Information extraction)
- Truy xuất thông tin (Information retrieval)
- Tác tử phân mềm hội thoại (Conversational agent)
- Tóm tắt văn bản (Text summarization)
- Hỏi đáp (Question answering)
- Dịch máy (Machine translation)
- Mô hình hóa chủ đề (Topic modelling)

3.1.2.2 Công cụ giải quyết các bài toán NLP

- NLTK
- SpaCy

- Stanford CoreNLP
- Gensim
- TensorFlow & PyTorch
- Hugging Face
- Scikit-learn

3.1.2.3 Các bước áp dụng vào đề tài tìm kiếm đề xuất NLP

Đề tài áp dụng bài toán “Truy xuất thông tin” khá giống Google để áp dụng vào đề tài để làm chức năng đề xuất:

Ghi chú: Trích xuất thông tin (tiếng anh là Information extraction - IE) tự động trích xuất thông tin có liên quan từ các tài liệu văn bản không có cấu trúc và / hoặc bán cấu trúc. Ví dụ về các loại tài liệu này bao gồm lịch sự kiện từ email hoặc tên của những người được đề cập trong một bài đăng trên mạng xã hội.

Bước 1: Tìm kiếm các yếu tố quyết định đến việc tìm kiếm đề xuất và thu thập dữ liệu

Một quyển sách thường gồm các thông tin như tên sách, mô tả, liên kết, tác giả, nhà xuất bản, ...

Nhưng đề tài này chỉ chọn tên sách và mô tả sách làm tiêu chí đề xuất bằng NLP.

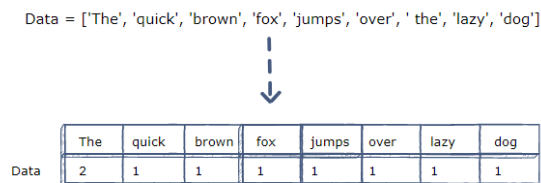
Bước 2: Xử lý văn bản thô bằng xử lý ngôn ngữ tự nhiên NLP.

Các kỹ thuật Xử lý ngôn ngữ tự nhiên NLP rất quan trọng trong với đề tài này khi phải xử lý dữ liệu văn bản. Vì dữ liệu của hệ thống đề xuất không thể được cung cấp cho bất kỳ mô hình máy học nào trừ khi chúng ta phải làm sạch và làm giàu Dataset, đó là lúc NLP phát huy tác dụng của chúng!

Nếu không có khâu xử lý Dataset ví dụ kiểm tra các cột dữ liệu trong file CSV, kiểm tra dữ liệu có để trống hay không, dữ liệu có đúng hay không có khớp với database mà hệ thống sử dụng hay không.

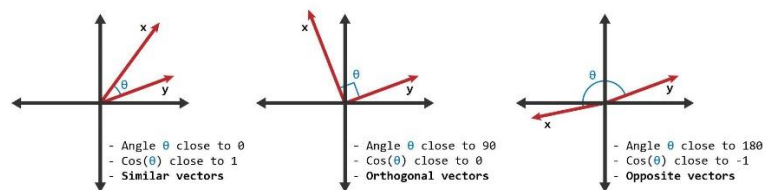
Bước 3: Vector hóa token từ các câu hỏi tìm kiếm và truy xuất thông tin.

Đề tài này sử dụng mô hình Bags of Words là mô hình Truy xuất thông tin giống Google nên sẽ sử dụng Scikit-learn 1 số module tạo nhanh các vector. Dựa vào các vector này mà ta sẽ biết được tần số xuất hiện của các từ (Count Vectorizer).

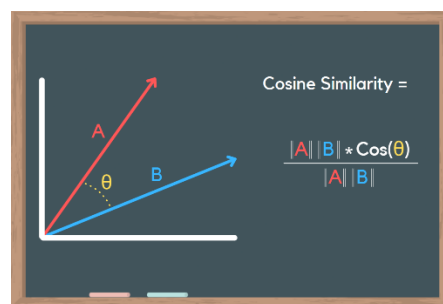


Hình 22 Count Vectorizer

Kết hợp với tính toán dựa trên đồ đồng dạng Cosine, ta sẽ đo được đồ đồng dạng giữa các vector chính là các góc 2 vector. Do mỗi vector tượng trưng cho 1 từ nên đã dễ dàng biết được xem 2 từ đó có mức độ giống nhau bao nhiêu.



Hình 23 Minh họa đồ đồng dạng Cosine



Hình 24 Công thức tính độ đồng dạng Cosine

Chính vì lý do này mà ta sẽ thu được các kết quả tương đồng và cũng sẽ lấy được kết quả đó làm kết quả đề xuất với mức độ tương đồng cao sẽ được xếp hàng trên cùng.

Bước 4: Đào tạo mô hình.

Dựa vào các lý thuyết ở trên thì chúng ta sẽ tiến hành viết code để đào tạo.

Bước 5: Thu nhận kết quả.

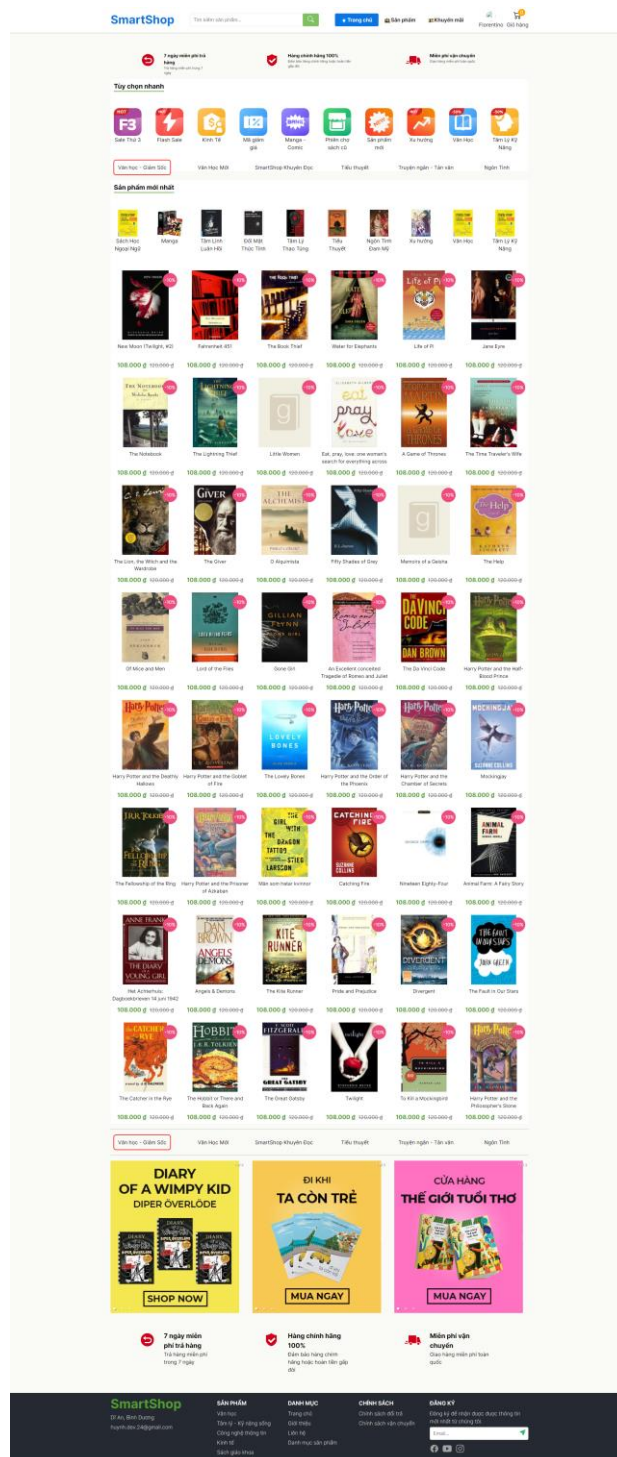
Sau khi đào tạo ta sẽ thu được 1 danh sách các sách được đề xuất và ta có thể sử dụng chúng để cho frontend render hiển thị trên giao diện đề xuất.

3.2 Các kết quả nghiên cứu thực nghiệm

Sau khi đã tiến hành các nghiên cứu lý thuyết thì sau đến bước áp dụng các kết quả đạt được trên xử lý ngôn ngữ tự nhiên NLP áp dụng và hệ thống bán sách giúp khách hàng tìm kiếm đề xuất cho mình những cuốn sách ưng ý.

3.2.1 Các kết quả thực nghiệm áp dụng NLP vào bán hàng

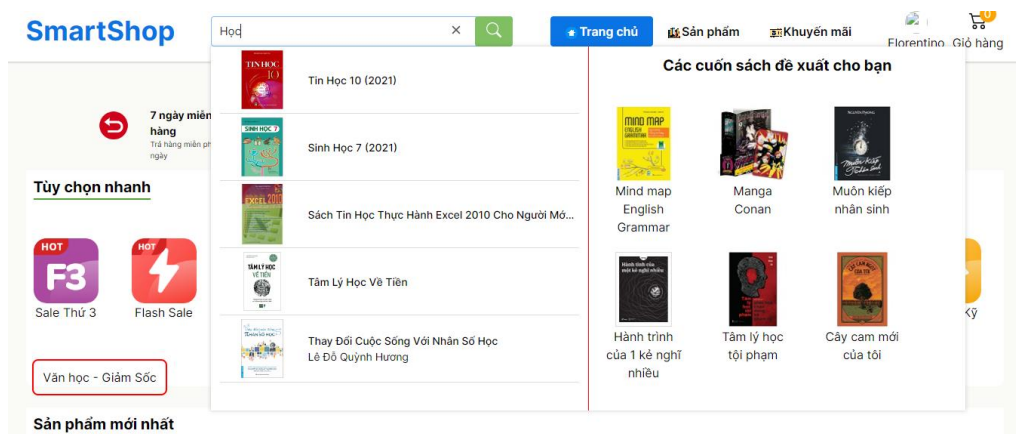
Trang chủ website



Hình 25 Trang chủ website

3.2.1.1 Tìm kiếm

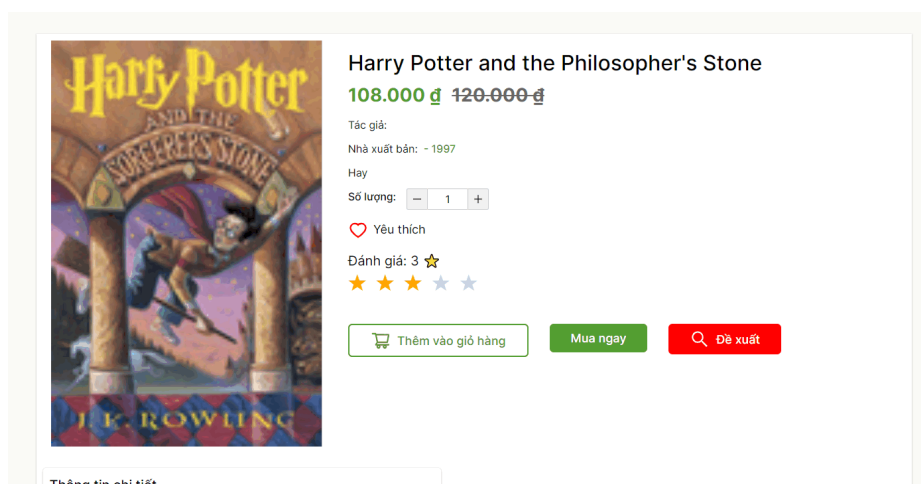
Cho phép người dùng tìm kiếm sách, đề xuất sách theo lịch sử đã mua.



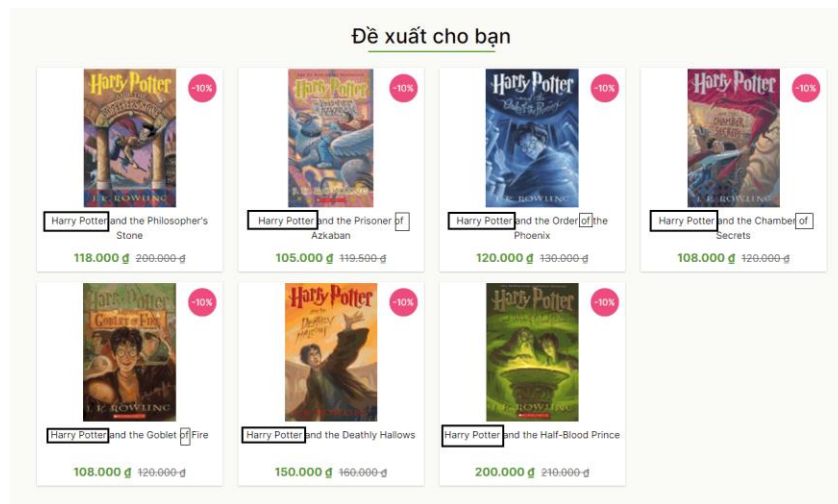
Hình 26 Tìm kiếm sử dụng đề xuất lịch sử mua hàng

3.2.1.2 Đề xuất sách

Đề xuất các cuốn sách theo NLP sẽ được trình bày trong giai đoạn số 2. Khi người dùng chọn 1 sản phẩm bất kỳ và nhấn vào nút đề xuất các sản phẩm tương tự thì trình xử lý NLP ở dưới backend sẽ xử lý và cho kết quả đề xuất ở phía dưới, NLP sẽ tách nhỏ truy vấn theo tên hoặc mô tả thành các token để xử lý tính toán độ tương đồng Consine.



Hình 27 Đề xuất sách bằng NLP



Hình 28 Kết quả đề xuất sách

3.2.2 Các chức năng của website bán hàng

3.2.2.1 Đăng ký

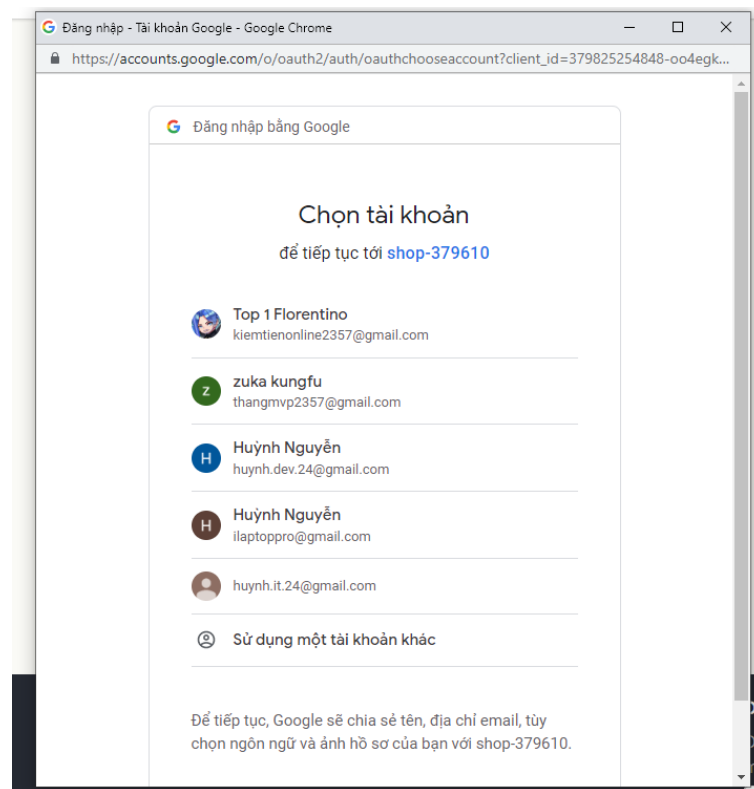
Cho phép khách hàng đăng ký thành viên vào hệ thống.

Hình 29 Đăng ký

3.2.2.2 Đăng nhập

Cho phép khách hàng thành viên, nhân viên chăm sóc khách hàng, quản trị viên đăng nhập vào hệ thống (có tích hợp Google Login).

Hình 30 Đăng nhập



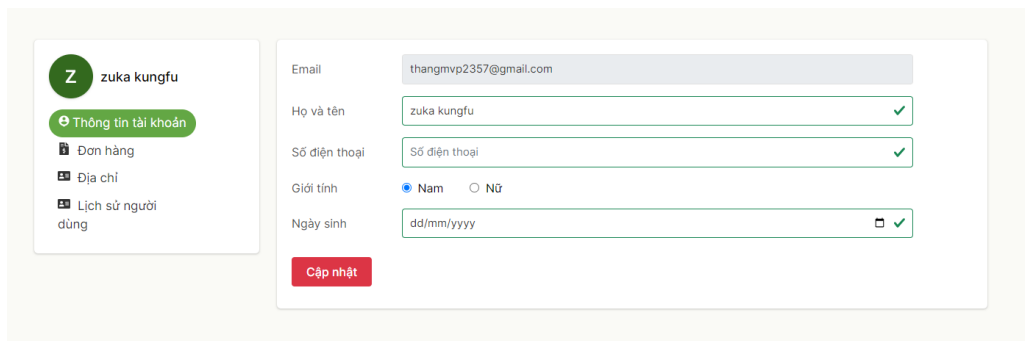
Hình 31 Đăng nhập Google

3.2.2.3 Đăng xuất

Cho phép quản trị viên, nhân viên chăm sóc khách hàng, khách hàng thành viên đăng xuất khỏi hệ thống.

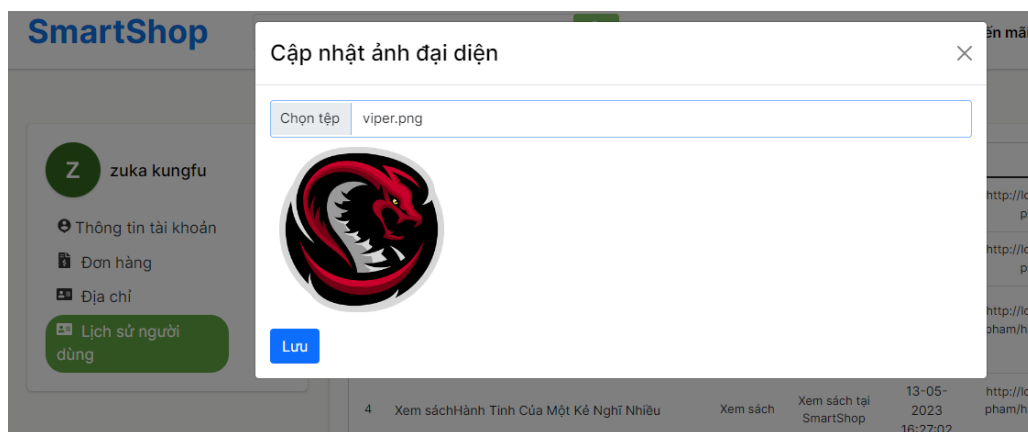
3.2.2.4 Quản lý tài khoản

Cho phép khách hàng thành viên xem thông tin tài khoản, cập nhật thông tin cá nhân gồm: Họ tên, địa chỉ, giới tính, profile cá nhân.



The screenshot shows a user profile update interface. On the left, a sidebar for user 'zuka kungfu' lists options: 'Thông tin tài khoản' (selected), 'Đơn hàng', 'Địa chỉ', and 'Lịch sử người dùng'. The main form contains fields for 'Email' (thangmvp2357@gmail.com), 'Họ và tên' (zuka kungfu), 'Số điện thoại' (SỐ ĐIỆN THOẠI), 'Giới tính' (Male selected), and 'Ngày sinh' (dd/mm/yyyy). A red 'Cập nhật' button is at the bottom.

Hình 32 Hồ sơ cá nhân

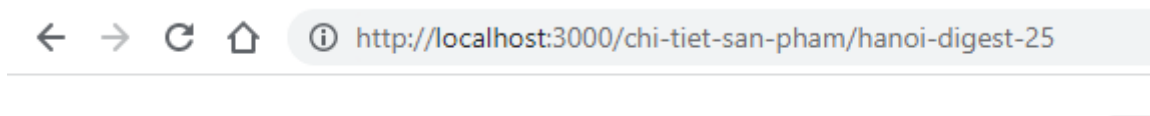


The screenshot shows a modal titled 'Cập nhật ảnh đại diện'. It includes a file selection area with 'Chọn tệp' and 'viper.png', a preview of a red dragon logo, and a blue 'Lưu' button. The background shows the SmartShop sidebar and a list of items.

Hình 33 Thay đổi ảnh đại diện

3.2.2.5 SEO URL Sản phẩm & Blog

Tối ưu hóa công cụ tìm kiếm với SEO Url cho đến nội dung.



Hình 34 SEO URL

3.2.2.6 Quản lý lịch sử xem hàng

Cho phép khách hàng thành viên xem lịch sử xem hàng.

STT	Thông tin giao hàng	Ngày đặt hàng	Tổng tiền	Tình trạng	Phương thức	Hành động
1	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: TP. Di An, Bình Dương, Xã Thảo Chư Phìn, Huyện Xi Ma Cai, Lào Cai	17-05-2023 22:14:51	108.000 đ	Chờ cửa hàng xác nhận	Trả tiền mặt khi nhận hàng	
2	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: TP. Di An, Bình Dương, Xã Thảo Chư Phìn, Huyện Xi Ma Cai, Lào Cai	17-05-2023 22:13:21	108.000 đ	Chờ cửa hàng xác nhận	Trả tiền mặt khi nhận hàng	
3	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 0365204529 Địa chỉ: ktx khu B, Phường Đồng Hòa, Thành phố Di An, Bình Dương	13-05-2023 16:35:53	54.000 đ	Giao hàng thành công	Vì momo Thanh toán thành công	
4	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: Nhà trọ, Phường 3, Quận Bình Thạnh, Hồ Chí Minh	10-04-2023 14:13:16	261.260 đ	Đã đóng gói, Chuẩn bị giao	Trả tiền mặt khi nhận hàng	
5	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 0365204529 Địa chỉ: Đường Rừng, Xã Phú Sơn, Huyện Tân Phú, Đồng Nai	19-03-2023 18:20:01	270.000 đ	Chờ cửa hàng xác nhận	Vì momo Thanh toán thành công	

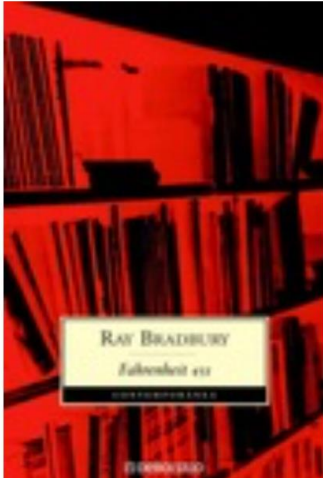
Hình 35 Xem lịch sử mua hàng

STT	Hoạt động	Kiểu	Ghi chú	Ngày	Liên kết
1	Thêm sách vào giỏ hàng	Thêm giỏ hàng	Lịch sử thêm vào giỏ hàng	17-05-2023 22:06:53	http://localhost:3000/chi-tiet-san-pham/hanoi-digest-25
2	Xem sáchThe Hanoi Digest 5	Xem sách	Xem sách tại SmartShop	17-05-2023 21:25:12	http://localhost:3000/chi-tiet-san-pham/hanoi-digest-25
3	Đặt sách	Đặt sách muốn thanh toán	Lịch sử đặt sách	13-05-2023 16:27:05	http://localhost:3000/chi-tiet-san-pham/hanh-tinh-cua-mot-ke-nghi-nhie
4	Xem sáchHành Tinh Của Một Kẻ Nghĩ Nhiều	Xem sách	Xem sách tại SmartShop	13-05-2023 16:27:02	http://localhost:3000/chi-tiet-san-pham/hanh-tinh-cua-mot-ke-nghi-nhie
5	Xem sách(Bestseller hải hước SỐ MỘT nước Anh của David Williams) NHÓC TÍ TỶ PHỤ – Tony Ross minh họa – Seal dịch – Nhà Nam – NXB Hội nhà văn (Bìa mềm)	Xem sách	Xem sách tại SmartShop	13-05-2023 10:29:22	http://localhost:3000/chi-tiet-san-pham/bestseller-hai-huoc-so-mot-nuo

Hình 36 Xem lịch sử thao tác trên website

3.2.2.7 Quản lý giỏ hàng

Cho phép khách hàng, khách hàng thành xem giỏ hàng, thêm sản phẩm vào giỏ, cập nhật số lượng sản phẩm trong giỏ, xóa sản phẩm khỏi giỏ hàng.



Fahrenheit 451

108.000 đ ~~120.000 đ~~

Tác giả:

Nhà xuất bản: - 1953

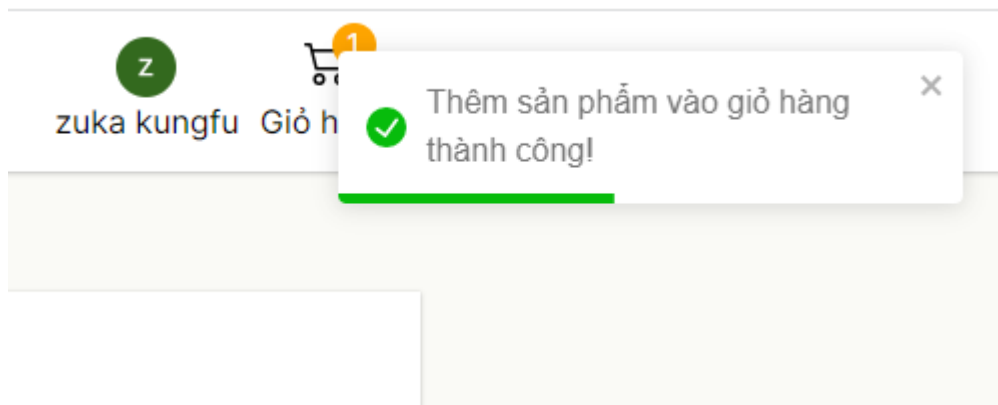
Hay

Số lượng:

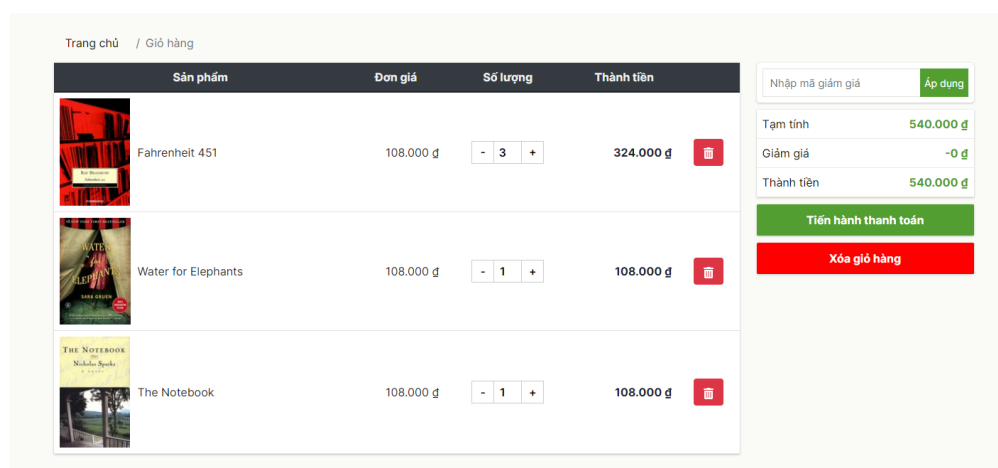
Yêu thích

Đánh giá: 3

Thêm vào giỏ hàng
 Mua ngay
 Đề xuất



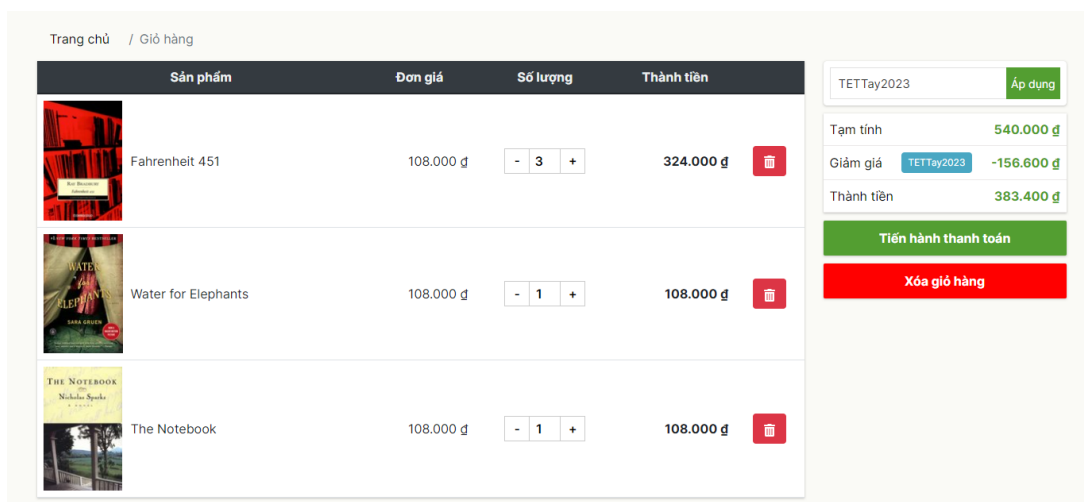
Hình 37 Thông báo thêm vào giỏ hàng thành công



Hình 38 Xem chi tiết giỏ hàng

3.2.2.8 Đặt hàng

Cho phép khách hàng tiến hành đặt hàng các sản phẩm trong giỏ hàng.



Hình 39 Tiến hành đặt hàng, có áp dụng mã giảm giá

Trang chủ / Thanh toán

THÔNG TIN NHẬN HÀNG

Họ và tên

zuka kungfu

Email

thangmp2357@gmail.com

Số điện thoại

03423423456

Địa chỉ giao hàng: Đường số 10, Phường Sơn Kỳ, Quận Tân Phú, Hồ Chí Minh

Hồ Chí Minh

Quận Tân Phú

Phường Sơn Kỳ

Đường số 10

Xác nhận địa chỉ

ĐƠN HÀNG CỦA BẠN

SẢN PHẨM	TỔNG
Fahrenheit 451 x3	324.000 đ
Water for Elephants x1	108.000 đ
The Notebook x1	108.000 đ
Tạm tính	540.000 đ
Giảm giá	-156.600 đ
Phí vận chuyển	+0 đ
Tổng	383.400 đ

DỊCH VỤ VẬN CHUYỂN

Không tìm thấy dịch vụ vận chuyển

0

PHƯƠNG THỨC THANH TOÁN

Trả tiền mặt khi nhận hàng

Vi MoMo

Paypal

ĐẶT HÀNG

Hình 40 Nhập thông tin đặt hàng

mo mo

Cổng thanh toán MoMo

Thông tin đơn hàng

Nhà cung cấp

Shop Laptop

Mã đơn hàng

3ad17d17-66dc-4636-81ee-202f3825e6a7

Mô tả

Thanh toán mua hàng tại SmartShop

Số tiền

383.400đ

Đơn hàng sẽ hết hạn sau:

09

58

Phút

Giây

Quay về

Quét mã QR để thanh toán

Sử dụng App MoMo hoặc ứng dụng camera hỗ trợ QR code để quét mã

Gặp khó khăn khi thanh toán? Xem Hướng dẫn

Hình 41 Thanh toán online bằng MoMo

mo mo

Cổng thanh toán MoMo

Thanh toán thành công

Shop Laptop

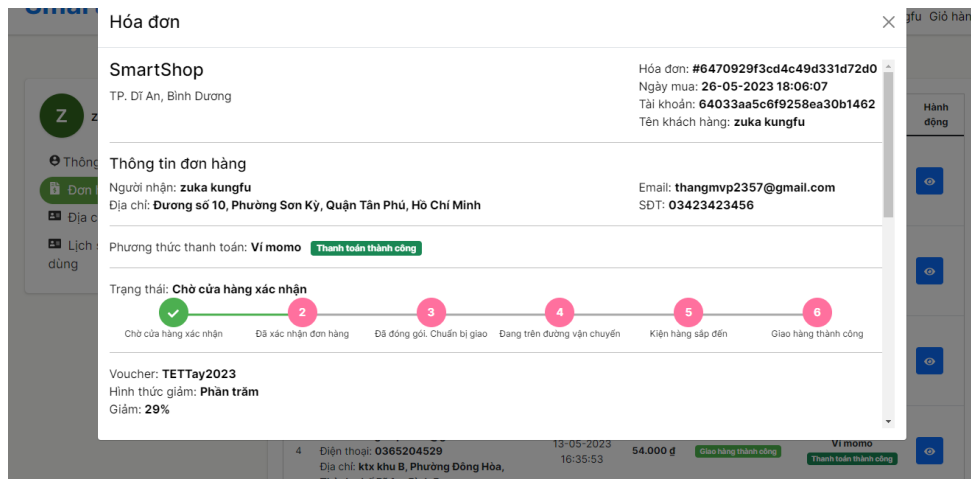
383.400đ

MoMo sẽ tự động đưa bạn về lại trang của Nhà cung cấp

Quay về

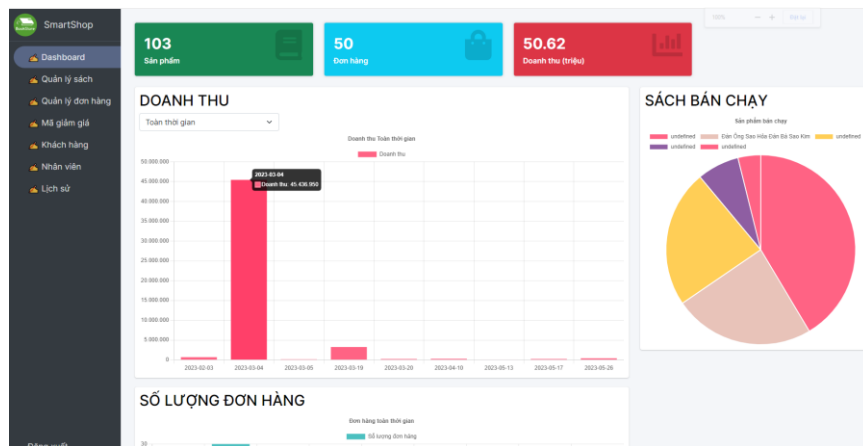
Hình 42 Thông báo thanh toán thành công

47



Hình 43 Khách xem trạng thái đơn hàng của mình

3.2.2.9 Trang quản lý Admin & Nhân viên



Hình 44 Trang quản lý

3.2.2.10 Quản lý đơn đặt hàng

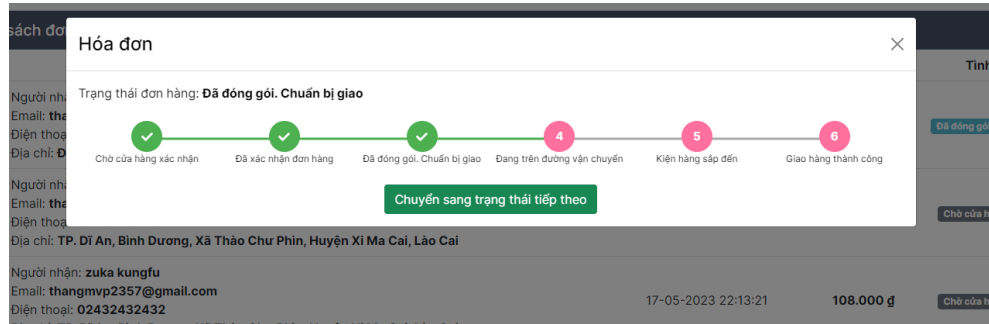
Cho phép khách hàng thành viên xem tình trạng đơn đặt hàng như: Đơn đặt hàng chưa duyệt, đã duyệt, đang giao, đã nhận, đã hủy và thực hiện chức năng hủy đơn hàng khi đơn đặt hàng chưa được duyệt và chức năng xác nhận đã nhận được hàng.

STT	Thông tin giao hàng	Ngày đặt hàng	Tổng tiền	Tình trạng	Hành động
1	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 03423423456 Địa chỉ: Đường số 10, Phường Sơn Kỳ, Quận Tân Phú, Hồ Chí Minh	26-05-2023 18:06:07	383.400 đ	Chờ cửa hàng xác nhận	<input checked="" type="checkbox"/> <input type="checkbox"/>
2	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: TP. Dĩ An, Bình Dương, Xã Tháo Chư Phin, Huyện Xi Ma Cai, Lào Cai	17-05-2023 22:14:51	108.000 đ	Chờ cửa hàng xác nhận	<input checked="" type="checkbox"/> <input type="checkbox"/>
3	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: TP. Dĩ An, Bình Dương, Xã Tháo Chư Phin, Huyện Xi Ma Cai, Lào Cai	17-05-2023 22:13:21	108.000 đ	Chờ cửa hàng xác nhận	<input checked="" type="checkbox"/> <input type="checkbox"/>
4	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 0365204529 Địa chỉ: ktx khu B, Phường Đồng Hòa, Thành phố Dĩ An, Bình Dương	13-05-2023 16:35:53	54.000 đ	Đã thanh toán thành công	<input checked="" type="checkbox"/> <input type="checkbox"/>
5	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 04343243555 Địa chỉ: Nhà trọ, Phường 3, Quận Bình Thạnh, Hồ Chí Minh	10-04-2023 14:13:16	261.260 đ	Đã đóng gói, Chuẩn bị giao	<input checked="" type="checkbox"/> <input type="checkbox"/>
6	Người nhận: 6137,Nguyễn Văn Huỳnh Email: huynh.h.24@gmail.com Điện thoại: 0365204529 Địa chỉ: Mố, Xã Phú Sơn, Huyện Tân Phú, Đồng Nai	20-03-2023 16:06:04	213.790 đ	Chờ cửa hàng xác nhận	<input checked="" type="checkbox"/> <input type="checkbox"/>
7	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 0365204529 Địa chỉ: Đường Rừng, Xã Phú Sơn, Huyện Tân Phú, Đồng Nai	18-03-2023 18:20:01	270.000 đ	Chờ cửa hàng xác nhận	<input checked="" type="checkbox"/> <input type="checkbox"/>

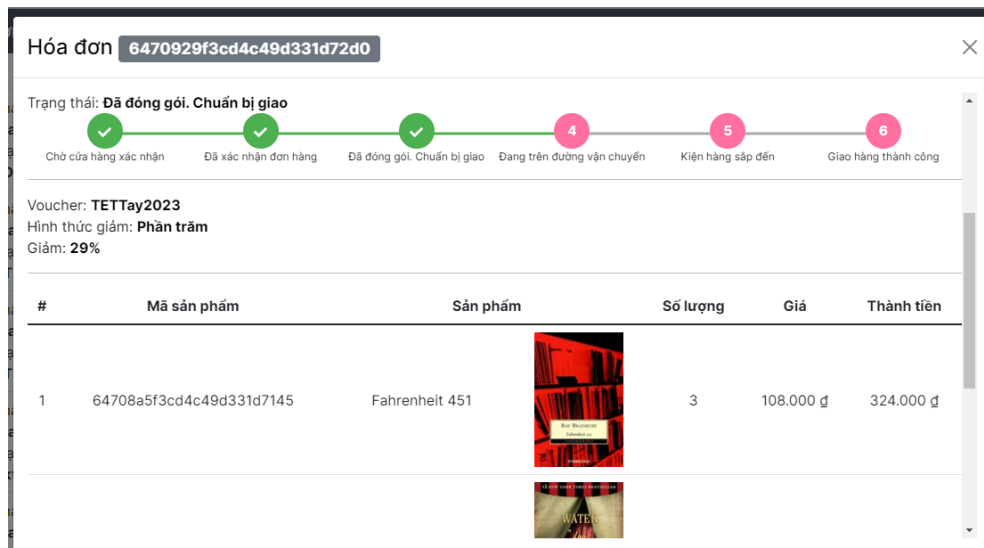
Hình 45 Quản lý đơn hàng

3.2.2.11 Xử lý đơn đặt hàng

Cho phép quản trị viên xem danh sách đơn đặt hàng chưa duyệt, đã duyệt nhưng chưa giao, đã hoàn thành và thực hiện chức năng duyệt đơn đặt hàng, giao hàng, tình trạng đơn hàng xác thực hay đã hủy.



















Hình 46 Xử lý tình trạng đơn hàng






Hình 47 Xem chi tiết đơn hàng quản lý

3.2.2.12 Quản lý Khách hàng

Cho phép Nhân viên chăm sóc, quản lý thông tin khách hàng, quản trị viên thực hiện các chức năng: Xem lịch sử giao hàng

Danh sách khách hàng				
Tìm kiếm bằng tên, email, SDT ...				
STT	Khách hàng	Tài khoản	Trạng thái	Hành động
1	 H Họ tên: Huỳnh Nguyễn Email: huynh.dev.24@gmail.com Điện thoại:	Google		
2	 H Họ tên: Huỳnh Nguyễn Email: ilaptoppro@gmail.com Điện thoại:	Google		
3	 H Họ tên: 6137.Nguyễn Văn Huỳnh Email: huynh.it.24@gmail.com Điện thoại: 0365204529	Google		
4	 Z Họ tên: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại:	Google		
5	 Z Họ tên: Zuka kungfu Email: thangmvp7532@gmail.com Điện thoại: 0345234344	Tài khoản BookStore		

Hình 48 Quản lý khách hàng

Lịch sử giao dịch					
STT	Thông tin giao hàng	Ngày đặt hàng	Tổng tiền	Tình trạng	Hành động
1	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 03423423456 Địa chỉ: Đường số 10, Phường Sơn Kỳ, Quận Tân Phú, Hồ Chí Minh	26-05-2023 18:06:07 5 minutes ago	383.400 đ	Đã đóng gói. Chuẩn bị giao	
2	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: TP. Dĩ An, Bình Dương, Xã Thào Chư Phìn, Huyện Xi Ma Cai, Lào Cai	17-05-2023 22:14:51	108.000 đ	Chờ cửa hàng xác nhận	
3	Người nhận: zuka kungfu Email: thangmvp2357@gmail.com Điện thoại: 02432432432 Địa chỉ: TP. Dĩ An, Bình Dương, Xã Thào Chư Phìn, Huyện Xi Ma Cai, Lào Cai	17-05-2023 22:13:21	108.000 đ	Chờ cửa hàng xác nhận	
Người nhận: zuka kungfu					





















Hình 49 Xem các đơn hàng của mỗi người

3.2.2.13 Quản lý danh mục sản phẩm

Cho phép quản trị viên xem danh sách danh mục sản phẩm và thực hiện chức năng tìm kiếm, thêm danh mục sản phẩm, cập nhật thông tin danh mục sản phẩm, khóa và kích hoạt danh mục sản phẩm.

3.2.2.14 Quản lý sản phẩm

Cho phép quản trị viên xem danh sách sản phẩm và thực hiện chức năng tìm kiếm, thêm sản phẩm, cập nhật thông tin sản phẩm, khóa và kích hoạt sản phẩm.

Danh sách sản phẩm						
<div> <div>Tìm kiếm</div> <div> <input type="button" value="🔍"/> <input type="button" value="📄 Xuất Excel"/> <input type="button" value="📄 Nhập Sách"/> </div> </div>						
STT	Tên sách	Thể loại	Xuất bản	Giá	Khuyến mãi (%)	Hành động
51	Khéo Ăn Nói Sẽ Có Được Thiên Hạ (Tái Bản 2022) - Trắc Nhã		NXB Văn Học - 2022	95.000 đ	2	 
52	Thay Đổi Cuộc Sống Với Nhân Số Học - Lê Đỗ Quỳnh Hương		- 2020	175.000 đ	5	 
53	Sách Giáo Khoa Bộ Lớp 12 - Sách Bài Tập (Bộ 11 Cuốn) (2022) - Nhiều tác giả	Sách giáo khoa	NXB giáo dục Việt Nam - 2017	165.000 đ	5	 
54	(Bestseller hải hước) SỐ MỘT nước Anh của David Walliams) NHÓC TỈ TỶ PHỤ - Tony Ross minh họa - Seal dịch - Nhà Nam - NXB Hội nhà văn (Bìa mềm) - Nhà Nam	truyện cười	NXB Kim đồng - 2020	100.000 đ	1	 
55	Sách trắng Công nghệ Thông tin và Truyền thông Việt Nam 2021 - Bộ Thông tin và Truyền Thông	Công nghệ thông tin	NXB giáo dục Việt Nam - 2021	180.000 đ	4	 
56	Cuốn Sách Dành Cho Những Người Dúng Tiếng Nhật Trong Lĩnh Vực IT: Tiếng Nhật Công Nghệ Thông Tin Trong Ngành Phần Mềm - Thái Hà	Công nghệ thông tin	NXB giáo dục Việt Nam - 2015	216.000 đ	3	 
57	90 Ngày Làm Chủ Cảm Xúc - Chinh Phục Hạnh Phúc Tập 3 (Tái bản 2021) - Tuệ An		- 2021	220.000 đ	4	 
58	Sách Kỹ Năng Sống: 10 Phút Tinh Tầm - 71 Thói Quen Căn Bản Giúp Cuộc Sống Hiện Đại (Tặng Bookmark Happy Life) - S.J. Scott, Barrie Davenport		- 2019	169.000 đ	2	 
59	Thuật Tẩy Não - Nghệ Thuật Thao Tung và Kiểm Soát Tâm Lý Đối Phương - Cao Đức		- 2013	120.000 đ	2	 
60	Hành Tinh của Một Kẻ Nghỉ Nhiều - Nguyễn Đoàn Minh Thư		- 2001	60.000 đ	10	 


Hình 50 Quản lý sản phẩm

[illegible]

Hình 51 Cập nhật thông tin sản phẩm (Sách)

3.2.2.15 Quản lý tài khoản hệ thống

Cho phép quản trị viên xem danh sách tài khoản hệ thống và thực hiện chức năng tìm kiếm, thêm, cập nhật thông tin, khóa và kích hoạt tài khoản Admin.

Danh sách nhân viên						
Nhập tên, mã nhân viên						Thêm nhân viên
STT	Tên nhân viên	Email	SĐT	Ngày tạo	Trạng thái	Hành động
1	 Nguyễn Văn Huỳnh (L)	kiemtienonline2357@gmail.com		04-03-2023 19:28:41	Đang hoạt động	Khóa

Hình 52 Tài khoản hệ thống

3.2.2.16 Quản lý thẻ loại sách, nhà xuất bản

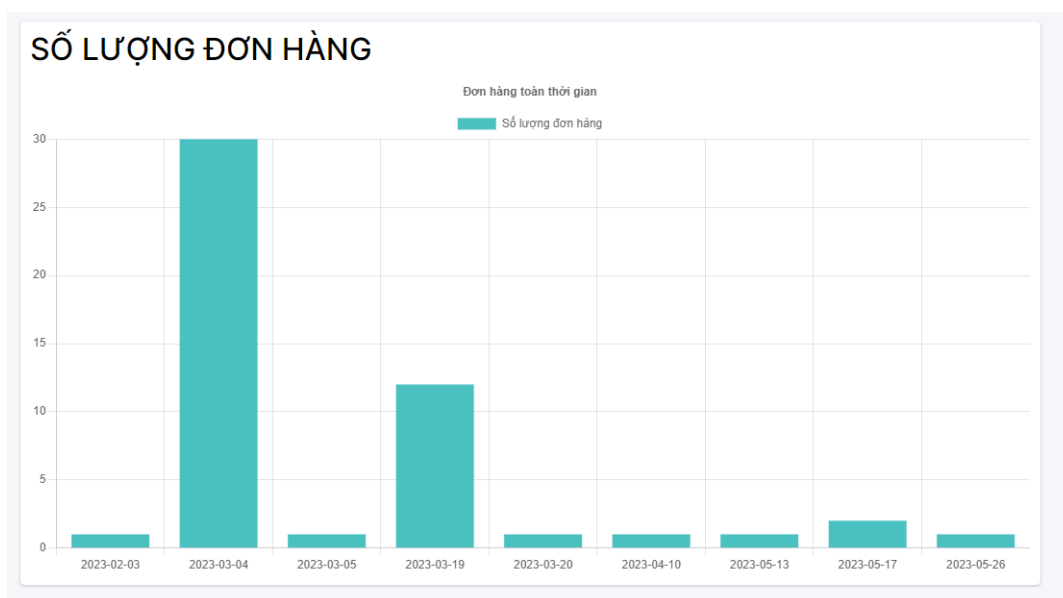
Cho phép quản trị viên xem danh sách thẻ loại sách, nhà xuất bản cũng như nhu cầu và thực hiện chức năng tìm kiếm, thêm hãng, cập nhật thông tin hãng, khóa và kích hoạt hãng.

3.2.2.17 Phân quyền, khóa tài khoản

Cho phép quản trị viên phân quyền cho loại tài khoản.

3.2.2.18 Thống kê

Chức năng cho phép quản trị viên thực hiện thống kê hàng tồn, khách hàng tiềm năng, sản phẩm bán chạy, đơn hàng, nhà cung cấp sách,...



Hình 53 Thống kê

3.2.2.19 Đổi mật khẩu

Cho phép người dùng đổi mật khẩu khi cần thiết.

3.3 Tóm tắt chương 3

Trình bày các bước thực hiện đề tài trên thực nghiệm, các bước xây dựng mô hình xử lý ngôn ngữ tự nhiên NLP và ứng dụng vào website bán sách. Dựa trên các kết quả lý thuyết mà sau này sẽ dựa vào đó để triển khai trên mô hình thực tế là website bán hàng sẽ sử dụng NLP trong đề xuất như thế nào.

Chương này cũng đã thể hiện các kết quả đạt được trên website bán sách các chức năng cơ bản của bán hàng cũng như các kết quả đề xuất bằng NLP bằng các hình chụp trực tiếp trên sản phẩm thực tế.

Ngoài ra chương còn nhắc lại nghiêm vụ của mỗi chức năng trong bán hàng theo tiêu chuẩn mà 1 website bán hàng phải có giao diện dễ dùng để thao tác.

Chương 4. GIẢI PHÁP GIẢI QUYẾT

4.1 Kết luận chung

Sau khi kết thúc quá trình thực hiện đề tài, thì đề tài đã đạt được 1 số kết quả như sau:

→ Đã khảo sát, phân tích, đánh giá các phương pháp xây dựng hệ thống đề xuất mà các hệ thống ứng dụng lớn trên thế giới đã và đang áp dụng và đạt kết quả tốt thậm chí là kết quả đề xuất rất ngờ.

→ Đề tài đã áp dụng 1 phương pháp chuyên xử lý ngôn ngữ tự nhiên vào hệ thống đề xuất không đi theo lối mòn các phương pháp truyền thống.

→ Hệ thống cũng đã đạt được 1 số kết quả đề xuất tốt với lượng data đủ lớn, đáng tin cậy và thuật toán hỗ trợ tốt có độ chính xác khá cao.

4.2 Đề xuất mới

Đề tài ban đầu chỉ đề áp dụng cho hệ thống bán sách, tuy nhiên chúng ta có thể áp dụng mô hình này để áp dụng vào các mặt hàng khác hoặc các thể loại website hay ứng dụng giúp cho mỗi hệ thống đều có thể triển khai được chức năng đề xuất ở quy mô nhỏ và vừa.

TÀI LIỆU THAM KHẢO

Danh sách link website tham khảo hỗ trợ đề cương:

[1] Tài liệu Numpy phần hướng dẫn nhanh (23/03/2023),

https://pytorch.org/tutorials/beginner/basics/quickstart_tutorial.html

[2] Tài liệu Scikit-learn phần API thư viện các hàm, các class (08/04/2023),

<https://scikit-learn.org/stable/modules/classes.html>

[3] Tài liệu Nodejs phần tài liệu sử dụng (08/04/2023),

<https://nodejs.dev/en/api/v19/documentation/>

[4] Tài liệu Reactjs phần tài liệu sử dụng (08/04/2023), <https://reactjs.org/docs/getting-started.html>

[5] Tài liệu MongoDB phần hướng dẫn sử dụng document (08/04/2023),

<https://www.mongodb.com/docs/manual/core/document/>

[6] Bài báo ứng dụng AI trong đề xuất sản phẩm (23/04/2023),

<https://lignsuniversity.com/blog/an-overview-of-artificial-intelligence-in-recommending-user-products-in-the-ecommerce-industry>

[7] Trang Wikipedia mạng Nơ-ron (09/04/2023),

https://en.wikipedia.org/wiki/Neural_network

[8] Trang Wikipedia Deep learning (09/04/2023),

https://en.wikipedia.org/wiki/Deep_learning

[9] Bài báo “What Netflix’s Recommendation Systems Can Teach Us About The Computing Challenges Of The Near Future” của tạp chí Forbes (24/05/2023),

<https://www.forbes.com/sites/forbestechcouncil/2021/02/19/what-netflixs-recommendation-systems-can-teach-us-about-the-computing-challenges-of-the-near-future/?sh=42c0dc7d7489>

[10] Bài viết giới thiệu về Pandas (08/04/2023),

<https://www.learndatasci.com/tutorials/applied-introduction-to-numpy-python-tutorial/>

- [11] Bài viết giới thiệu Numpy (08/04/2023),
<https://www.learndatasci.com/tutorials/applied-introduction-to-numpy-python-tutorial/>
- [12] Bài báo “Modern recommendation system trong các ứng dụng thực tế, bạn đã biết chưa?” (24/05/2023), <https://news.sun-asterisk.com/vi/p/modern-recommendation-system-trong-cac-ung-dung-thuc-te-ban-da-biet-chua-66ed7616b37>
- [13] Bài báo “Xử lý ngôn ngữ tự nhiên: Công nghệ giúp máy tính hiểu và giao tiếp với con người” của trường đại học FPT (22/05/2023),
<https://digital.fpt.com.vn/dxarticles/xu-ly-ngon-ngu-tu-nhien.html>
- [14] Bài báo “Xử lý ngôn ngữ tự nhiên: Bài toán & công cụ bạn nên biết” của Viện VinBigData (20/05/2023), <https://vinbigdata.com/cong-nghe-giong-noi/xu-ly-ngon-ngu-tu-nhien-bai-toan-cong-cu-ban-nen-biet.html>
- [15] Count Vectorizer (18/05/2023),
<https://www.educative.io/answers/countvectorizer-in-python>
- [16] Bài viết “What is Cosine Similarity” (15/05/2023),
<https://towardsdatascience.com/what-is-cosine-similarity-how-to-compare-text-and-images-in-python-d2bb6e411ef0>