BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

****

ĐỀ TÀI

**XÂY DỰNG HỆ THỐNG**

**QUẢN LÝ SÂN CẦU LÔNG**

**VÀ BÁN SẢN PHẨM ONLINE**

**KẾT HỢP CHAT BOT**

**SINH VIÊN THỰC HIỆN:**

**Lê Minh Chiến**

**MSSV: B1606935**

**Khóa: 42**

**Cần Thơ, Tháng 01/2021**

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**BỘ MÔN TIN HỌC ỨNG DỤNG**

**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

****

ĐỀ TÀI

**XÂY DỰNG HỆ THỐNG**

**QUẢN LÝ SÂN CẦU LÔNG**

**VÀ BÁN SẢN PHẨM ONLINE**

**KẾT HỢP CHAT BOT**

**GIÁO VIÊN HƯỚNG DẪN: SINH VIÊN THỰC HIỆN:**

**Th.S Nguyễn Cao Hồng Ngọc Lê Minh Chiến**

**MSSV: B1706563**

**Khóa: 43**

**Cần Thơ, Tháng 01/2021**

LỜI CẢM ƠN

Đầu tiên, em xin gửi lời cảm ơn chân thành đến gia đình và bạn bè đã động viên em trong suốt quá trình học tập.

Em xin cảm ơn Bộ môn Công nghệ thông tin, khoa Công nghệ thông tin và Truyền thông, Trường Đại học Cần Thơ đã tạo điều kiện tốt nhất để em thực hiện đề tài này.

Dưới sự hướng dẫn của ThS. Nguyễn Cao Hồng Ngọc, em đã hoàn thành đề tài “Xây dựng web quản lý sân cầu lông kết hợp chatbot”. Em xin bày tỏ lời cảm ơn sâu sắc đến cô vì đã tận tình giúp đỡ, định hướng cho em trong suốt thời gian qua.

Cảm ơn thầy dạy cầu lông của em – thầy Diệp Xuân Tài đã đưa ra đề nghị cũng như ý tưởng để em có ý tưởng thực hiện đề tài luận văn này.

Trong quá trình thực hiện, em đã tích lũy được thêm nhiều kiến thức, giúp bản thân em dần hoàn thiện hơn. Đồng thời, kết hợp với kiến thức nền tảng đã có trong thời gian học tập những năm qua, em đã hoàn thành đề tài này.

Em rất mong nhận được sự góp ý, nhận xét từ quý thầy để đề tài được hoàn thiện nhất.

Cuối lời, em xin chúc quý thầy dồi dào sức khỏe, hạnh phúc và đạt được nhiều thành công trong công việc và cuộc sống.

Trân trọng!

Cần Thơ, ngày 19 tháng 12 năm 2020

Sinh viên thực hiện

Lê Minh Chiến

**NHẬN XÉT CỦA CÁN BỘ HƯỚNG DẪN**

Cần Thơ, ngày … tháng … năm ….  
Cán bộ hướng dẫn

Mục lục

[Mục lục i](#_Toc62855773)

[Danh mục hình ảnh ii](#_Toc62855774)

[Danh mục bảng iii](#_Toc62855775)

[Danh mục các ký hiệu, chữ viết tắt iv](#_Toc62855776)

[Tóm tắt v](#_Toc62855777)

[Abstract vi](#_Toc62855778)

[PHẦN GIỚI THIỆU 1](#_Toc62855779)

[1. Đặt vấn đề 1](#_Toc62855780)

[2. Các nghiên cứu liên quan 2](#_Toc62855781)

[3. Mục tiêu đề tài 3](#_Toc62855782)

[4. Đối tượng và phạm vi nghiên cứu 3](#_Toc62855783)

[5. Phương pháp nghiên cứu 3](#_Toc62855784)

[5.1. Về lý thuyết 3](#_Toc62855785)

[5.2. Về kỹ thuật 4](#_Toc62855786)

[6. Bố cục luận văn 4](#_Toc62855787)

[PHẦN NỘI DUNG 5](#_Toc62855788)

[CHƯƠNG 1: CƠ SỞ LÝ THUYẾT 5](#_Toc62855789)

[1.1. Trí tuệ nhân tạo 5](#_Toc62855790)

[1.2. Chatbot 6](#_Toc62855791)

[1.2.1. Giới thiệu chatbot 6](#_Toc62855792)

[1.2.2. Các khái niệm trong chatbot 7](#_Toc62855793)

[1.2.3. Các phương pháp xây dựng chatbot 7](#_Toc62855794)

[1.2.4. Xử lý ngôn ngữ tự nhiên (NLP) 9](#_Toc62855795)

Danh mục hình ảnh

[Hình 1.1 Trợ lý ảo Siri tích hợp trên Iphone 2](#_Toc62855765)

[Hình 1.2 Chatbot FPT.AI trên nền tảng Facebook 2](#_Toc62855766)

[Hình 1.1 Mối liên hệ giữa AI, Machine learning và Deep learning 5](#_Toc62855767)

[Hình 1.2 Cơ chế hoạt động chung của một chatbot 6](#_Toc62855768)

[Hình 1.3 Minh họa một Artificial Neural Network 8](#_Toc62855769)

[Hình 1.4 Các ứng dụng của NLP 10](#_Toc62855770)

Danh mục bảng

Danh mục các ký hiệu, chữ viết tắt

|  |  |
| --- | --- |
| **Ký hiệu, chữ viết tắt** | **Diễn giải** |
| AI | Artificial Intelligence |
| API | Application Program Interface |
| CSDL | Cơ sở dữ liệu |
| DL | Deep Learning |
| HTTP | HTTP HyperText Transfer Protocol |
| JSON | JavaScript Object Notation |
| ML | Machine Learning |
| MVC | Model - View - Controller |
| NLP | Natural Language Processing |
| REST | REpresentational State Transfer |
| SVM | Support Vector Machine |
| URL | Uniform Resource Locator |
| XML | eXtensible Markup Language |

Tóm tắt

Nhận thấy nhu cầu của chủ sân cầu lông là cần một ứng dụng để giúp họ quản lý dữ liệu sân tốt hơn khi phải cạnh tranh với nhau trong thời buổi công nghệ. Việc xây dựng một hệ thống website cho phép họ có thể dễ dàng quản lý dữ liệu kinh doanh của mình cũng như việc kết hợp bán sản phẩm online trên hệ thống là lý do giúp em thực hiện đề tài “Xây dựng hệ thống quản lý sân cầu lông và buôn bán sản phẩm kết hợp với chatbot”.

Giao diện website được xây dựng bằng thư viện ReactJS được phát triển và cung cấp bởi Facebook dựa trên nền tảng nodeJS, một thư viện rất nổi tiếng với việc sử dụng component, render dữ liệu dựa vào state và tạo ứng dụng dưới dạng single page. Back-end của hệ thống được xây dựng bằng NodeJS kết nối đến cơ sử dữ liệu NoSQL MongoDB.

Về phần AI (Artificial intelligence) – trí tuệ nhân tạo đã được áp dụng vào nhiều lĩnh vực trong cuộc sống và nâng cao chất lượng con người. Trong hệ thống này em sẽ áp dụng công nghệ này để tạo chat bot giúp trả lời câu hỏi tự động liên quan đến hệ thống.

Kết quả đề tài luậ văn này sẽ là hệ thống website dành cho hai nhóm người dùng chính, chủ sân cầu lông và người dùng mua các sản phẩm trên hệ thống. Các chủ sân có thể dùng các chức năng chính của hệ thống như quản lý sân, quản lý hoá đơn, xem doanh thu, đăng bán sản phẩm trên hệ thống,… Đối với người dùng thông thường họ có thể dùng chatbot để giúp họ giải đáp thắc mắc, mua sắm trên website và thanh toán online. Về phần quản trị viên có thể huấn luyện mô hình chatbot nhằm mở rộng kiến thức của chatbot.

Abstract

PHẦN GIỚI THIỆU

* 1. Đặt vấn đề

Ngày nay, việc quản lý dữ liệu trong kinh doanh một cách số hoá không chỉ được các chủ doanh nghiệp lớn quan tâm mà còn có những người kinh doanh nhỏ lẻ rất quan tâm và trong đó có các chủ sân cầu lông. Họ cần một ứng dụng giúp mình quản lý các dữ liệu trong lúc kinh doanh như doanh thu, hoá đơn, hàng hoá,… Và bên cạnh đó họ cũng bán các sản phẩm liên quan tới cầu lông như quần, áo, giày, vợt,… để tăng doanh thu. Những đối tượng này quan tâm việc khách hàng sử dụng dịch vụ thuê sân hơn và họ chỉ xem việc kinh doanh các sản phẩm còn lại như một cách tăng doanh thu nên họ rất ít đầu tư tiền bạc để nhờ các công ty phát triển cho mình một phần mềm quản lý hay một trang web để buôn bán các sản phẩm mà thường đăng các sản phẩm của mình trên facebook. Vì vậy một hệ thống giúp họ có thể vừa quản lý dữ liệu của sân cầu lông của mình và cho họ một route riêng trên hệ thống để đăng tải các sản phẩm tiện cho việc kinh doanh là một nhu cầu thiết yếu.

Khái niệm chatbot được nhắc đến đầu tiên vào năm 1950, bởi Alan Turing với ý tưởng “xây dựng một chương trình máy tính giả dạng con người để thực hiện cuộc trò chuyện giữa chương trình và con người thật”. Và những năm gần đây, độ tiện lợi của chatbot ngày càng cao khi nó đã cho thấy được tiện ích của mình và xuất hiện ngày càng nhiều hơn trên các mạng xã hội, website, cho đến các thiết bị di động. Chatot có thể giúp cho con người thực hiện một hành động hoặc cung cấp tri thức cho người dùng. Chatbot có thể giap tiếp nhận mệnh lệnh từ người dùng thông qua tin nhắn hoặc giọng nói. Để xác định ý định (intent) của người dùng, chatbot sẽ dùng hệ thống xử lý ngôn ngữ tự nhiên (NLP) để phân tích và dùng các thuật toán máy học (ML) để phân loại câu (text classification) và đưa ra hành động hoặc câu trả lời phù hợp nhất.

Việc kết hợp một chatbot giúp trả lời khách hàng hiểu rõ hơn về hệ thống sẽ giúp cho họ tiếp cận hệ thống một cách thuận tiện hơn và giúp cho chủ sân đỡ phải trả lời những câu hỏi thường gặp của khách hàng nếu chúng lặp đi lặp lại. Việc lựa chọn website để xây dựng hệ thống cùng nhầm giúp cho khách hàng dễ tiếp cận hơn với hệ thống thay vì phải cài đặt các phầm mềm trên máy tính của họ cũng giúp cho họ dễ tiếp cận với hệ thống hơn. Đây cũng chính là lý do đề tài “Xây dựng hệ thống quản lý sân cầu lông và buôn bán sản phẩm kết hợp chatbot” được thực hiện.

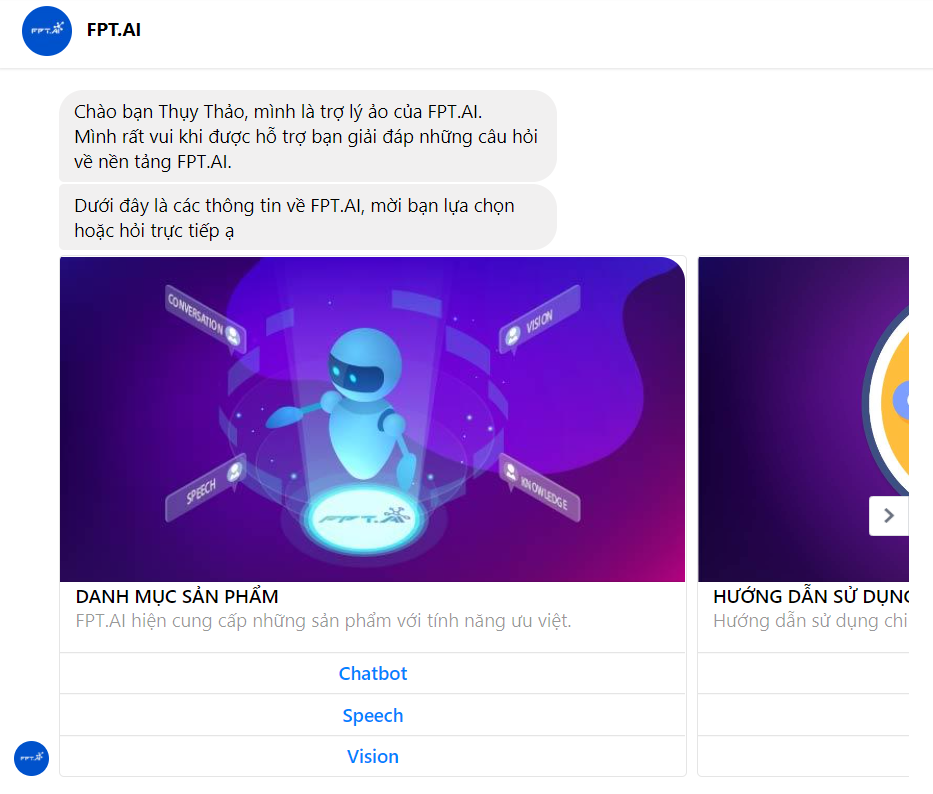
* 1. Các nghiên cứu liên quan

Trợ lý ảo là một dạng chatbot. Siri của Apple, Cortana của Windows hay Google Assistant của Google là những trợ lý ảo rất nổi tiếng được tích hợp vào thiết bị di động hay máy tính để hỗ trợ người dùng. Các trợ lý ảo có thể nhận lệnh từ giọng nói và thay người dùng thao tác như đọc thông báo, đặt lịch hẹn, báo thức, thực hiện cuộc gọi, gửi tin nhắn, …



* + - * 1. Trợ lý ảo Siri tích hợp trên Iphone

Hiện nay, trên nền tảng mạng xã hội Facebook cho phép xây dựng các chatbot. Chatbot dạng này thường đưa ra các lựa chọn cho người dùng chọn để thao tác.



* + - * 1. Chatbot FPT.AI trên nền tảng Facebook
  1. Mục tiêu đề tài

Mục tiêu của đề tài là xây dựng được một website hệ thống quản lý sân cầu lông gồm các năng: quản lý sân, quản lý sản phẩm, quản lý hoá đơn, xem các báo cáo doanh thu, mua hàng và thanh toán online cho khách hàng,.. kết hợp với một chatbot có thể trả lời các câu hỏi liên quan tới hệ thống giúp cho khách hàng tiện sử dụng trong quá trình sử dụng hệ thống. Và trang quản lý cho các chủ sân giúp họ dễ dàng quản lý dữ liệu của bản thân. Bên cạnh đó là một trang admin quản lý toàn bộ cơ sở dữ liệu của hệ thống tiện cho việc quản lý dữ liệu của người dùng cũng như huấn luyện cho chatbot.

* 1. Đối tượng và phạm vi nghiên cứu

Đề tài tập trung nghiên cứu các vấn đề:

* Tìm hiểu các chatbot, hệ thống gợi ý hiện có
* Tìm hiểu xử lý ngôn ngữ tự nhiên, thuật toán để huấn luyện mô hình, phân loại ý định người dùng trong chatbot và thiết kế quy trình đưa ra hành động thích hợp.
* Xây dựng thuật toán cho hệ thống gợi ý.
* Xây dựng website dành cho người dùng.
* Xây dựng website dành cho người quản trị.
* Tích hợp chatbot, hệ thống gợi ý vào website.
  1. Phương pháp nghiên cứu
     1. Về lý thuyết
* Tìm hiểu ngôn ngữ Python, các package, thư viện cần thiết cho phần core AI.
* Tìm hiểu các thuật toán sử dụng trong chatbot và hệ thống gợi ý.
* Tìm hiểu framework Flask cho việc xây dựng server cho service AI.
* Tìm hiểu thư viện React và các thư viện hỗ trợ khác cho việc xây dựng frond end của web.
* Tìm hiểu NodeJS và các thư viện hỗ trợ khác xây dựng server backend cho web.
* Tìm hiểu MongoDB và Mongoose cho việc lưu trữ và truy vấn dữ liệu.
* Tìm hiểu kỹ thuật lập trình một website và tích hợp với core AI.
  + 1. Về kỹ thuật
* Sử dụng ngôn ngữ Python cho việc lập trình phần core AI.
* Sử dụng MongoDB và Mongoose để quản lý cơ sở dữ liệu.
* Sử dụng React để xây dựng giao diện và NodeJS để xây dựng server.
  1. Bố cục luận văn

Bố cục quyển báo cáo luận văn gồm 3 phần, mỗi phần có những nội dung như sau:

* Phần giới thiệu: Gồm các nội dung như sau:

+ Đặt vấn đề

+ Các nghiên cứu liên quan

+ Mục tiêu đề tài

+ Đối tượng và phạm vi nghiên cứu

+ Phương pháp nghiên cứu

* Phần nội dung: Gồm các chương 1, 2, 3 với các nội dung như sau:

+ Chương 1: Cơ sở lý thuyết

+ Chương 2: Nội dung, phương pháp nghiên cứu

+ Chương 3: Đánh giá kiểm thử

* Phần kết luận: Gồm các nội dung như sau:

+ Kết luận

+ Hướng phát triển.

* 1. Kế hoạch làm việc

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tuần 1-4 | Tuần 5-8 | Tuần 9-10 | Tuần 11-15 | Tuần 16-17 | Tuần 18-20 |
| Thiết kế các chức năng, cơ sở dữ liệu của hệ thống và viết back-end. | Viết các chúc năng của hệ thống ở phía front end. | Tìm hiểu cơ sở lý thuyết về chatbot | Viết chatbot và tích hợp với hệ thống | Kiểm thử, sửa chửa lỗi, hoàn thiện chức năng | Hoàn thiện báo cáo |

PHẦN Nội DUNG

Chương 1: Cơ sở Lý Thuyết

* 1. Phân tích hệ thống thông tin
     1. Web service

Web Service đã mở ra một hướng mới cho việc phát triển các ứng dụng Internet. Web Service có thể tạm hiểu là dịch vụ web. Web Service kết hơp sử dụng nhiều công nghệ khác nhau cho phép các ứng dụng khác nhau được viết bằng nhiều ngôn ngữ khác nhau, vận hành trên các nền tảng khác nhau những vẫn có thể trao đổi thông tin qua lại thông qua môi trường Internet. Web Service là một khái niệm rộng hơn so với khái niệm web thông thường, nó cung cấp thông tin thô, và khó hiểu với đa số người dùng, chính vì vậy nó được sử dụng bởi các ứng dụng. Các dữ liệu này thường được trả về dưới dạng XML hoặc phổ biến hơn là JSON.

* + 1. RESTful API

RESTfull API là một tiêu chuẩn dùng trong việc thiết kế API cho các ứng dụng web (thiết kế Web Service) để tiện cho việc quản lý resource. Nó chú trọng vào tài nguyên hệ thống (tệp văn bản, âm thanh, video, hoặc dữ liệu động, …), bao gồm các trạng thái tài nguyên được định dạng và được truyền tải qua HTTP.

Diễn giải các thành phần:

* API (Application Programming Interface) là một tập hợp các quy tắc và cơ chế mà theo đó, một ứng dụng hay một thành phần sẽ tương tác với một ứng dụng hay một thành phần khác. API có thể trả về dữ liệu mà bạn cần cho ứng dụng của mình ở những kiểu dữ liệu phổ biến như JSON hay XML.
* REST (Representational State Transfer) là một dạng chuyển đổi cấu trúc dữ liệu, một kiến trúc để viết API. Nó sử dụng phương thức HTTP đơn giản để tạo giao tiếp giữa các máy. Vì vậy, thay vì sử dụng một URL cho việc xử lý một số thông tin người dùng, REST gửi một yêu cầu HTTP như GET, POST, DELETE,…vv đến một URL để xử lý dữ liệu.
* RESTful API là một tiêu chuẩn dùng trong việc thiết kế các API cho các ứng dụng web để quản lý resource. RESTful API ngày nay được sử dụng phổ biến cho các ứng dụng (web, mobile,…) khác giao tiếp với nhau.
* Chức năng quan trọng nhất của REST là quy định các sử dụng HTTP method (như GET, POST, PUT, DELETE,..) và cách định dạng các URL cho ứng dụng web để quản lý resource. RESTful không quy định logic code ứng dụng và không giới hạn bởi ngôn ngữ lập trình ứng dụng, bất kỳ ngon ngữ hoặc framework nào cũng có thể sử dụng để thiết RESTfull API.

Các thức hoạt động



* + - * 1. Cách thức hoạt động của REST

REST hoạt động chủ yếu dựa vào giao thức HTTP. Các hoạt động cơ bản nêu trên sẽ sử dụng những phương thức HTTP riêng.

* GET (SELECT): Trả về một tài nguyên hoặc danh sách tài nguyên
* POST (CREATE): Tạo mới một tài nguyên.
* PUT (UPDATE): Cập nhật thông tin cho tài nguyên.
* DELETE (DELETE): Xoá một tài nguyên

Xác thực dữ liệu trả về:

RESTfull API không sử dụng session và cookie, nó sử dụng một access\_token với mỗi request. Dữ liệu trả về thường có cấu trúc sau:

{

‘id’: ‘B1706563,

‘name’: ‘Lê Minh Chiến’,

‘class’: ‘DI17V7A2’

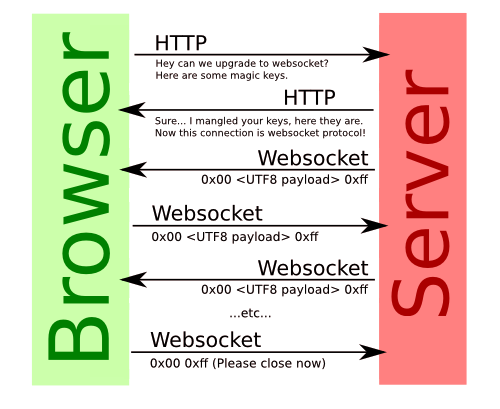
}

* + 1. Socket và Websocket

**Socket** là điểm cuối trong liên kết truyền thông tin hai chiều (two-way-communication) biểu diễn kết nối giữa Client-Server. Các lớp Socket được rang buộc với một cổng port (thể hiện là một con số cụ thể) để các tầng TCP (TCP Layer) có thể định danh ứng dụng mà dữ liệu sẽ được gửi tới.

Chức năng của socket là kết nối giữa client và server thông qua TCP/IP và UDP để truyền và nhận dữ liệu qua Internet. Giao diện lập trình ứng dụng mạng này chỉ có thể hoạt động khi đã có thông tin về thông số IP và số liệu cổng của 2 ứng dụng cần trao đổi dữ liệu cho nhau.

**Websocket** là giao thức hỗ trợ giao tiếp hai chiều giữa client và server dể tạo một kết nối trao đổi dữ liệu. Giao thức này không sử dụng HTTP mà thực hiện nó qua TCP. Mặc dù thiết kế để chuyên sử dụng cho các ứng dụng web, lập trình viên vẫn có thể đưa chúng vào bất kỳ loại ứng dụng nào.



* + - * 1. Mô tả cách thức hoạt động của Websocket
    1. Các nền tảng phía Client
       1. ReactJS

ReactJS là một thư viện của JavaScript, dùng để xây dựng giao diện cho người dùng được phát hành bởi facebook. React được phát triển để giúp lập trình viên tạo ra một single-page application một cách dễ dàng. React chia các thành phần của một trang web thành các component nhỏ dễ viết hơn nhờ việc sử dụng JSX, mở rộng cú pháp tuỳ chọn của JavaScript cho phép người dùng kết hợp HTML và JavaScript, người dùng chỉ đơn giản xử lý logic với JavaScript trong các component thay vì phải quan tâm đến toàn bộ ứng dụng. Việc chia ứng dụng web thành nhiều component cũng giúp người dùng có thể tái sử dụng một các dễ dàng nếu phát triển các chức năng tương tự. Hơn thế, ReactJS mang lại hiệu suất tốt hơn với Virtual DOM (Virtual Document Object Model – Mô hình đối tượng tài liệu), cho phép người dùng xây dựng các virtual DOMs và host chúng lại trong bộ nhớ. Mỗi khi component cần render lại virtual DOM sẽ được so sánh với DOM thật để tìm ra sự khác biệt từ đó thay thế chúng trong DOM thật.

* + - 1. Boostrap

Boostrap là một framework automatic CSS, cung cấp cho người dùng một loạt các class CSS được viết sẵn một cách có quy tắc, người dùng sẽ chỉ đơn giản quan tâm đến việc sử dụng chúng như thế nào và tuỳ chỉnh để phù hợp với trang web của mình thay vì phải viết lại từng style cho trang web. Một giải pháp tiện lợi giúp cho người dùng đỡ mất thời gian hơn trong việc thiết kế giao diện website.

* + 1. Các nền tảng phía Server
       1. MongoDB

**Định nghĩa:**

MongoDB là một cơ sở dữ liệu mã nguồn mở và là cơ sử dữ liệu NoSQL hàng đầu, được hàng triệu người sử dụng được viết bởi ngôn ngữ C++.

NoSQL là 1 dạng CSDL mã nguồn mở không sử dụng Transact-SQL để truy vấn thông tin. NoSQL được viết tắt bởi: Non-Relationship SQL, hay có nơi thường gọi là Not-Only SQL. CSDL này được phát triển trên JavaScript Framework với kiểu dữ liệu JSON. (Cú pháp của JSON là “key: value”) NoSQL ra đời như một mảnh vá cho những khiếm khuyết và thiết xoát cũng như hạn chế của mô hình dữ liệu quan hệ. RDBMS về tốc độ, tính năng, khả năng mở rộng, memory cache,…

**Các thuật ngữ trong MongoDB:**

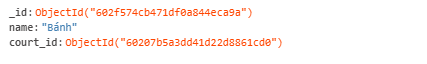
**\_id –** Là một trường bắt buộc được tự động sinh đại diện cho giá trị duy nhất trong MongoDB. Có thể được dùng để làm khoá chính cho document nhờ vào tính duy nhất của nó. Khi một document mới được thêm vào collection MongoDB sẽ tự động cấp cho nó một \_id để định danh document đó.

**Collection –** Là nhóm nhiều document trong MongoDB. Collection có thể được hiể là một bảng tương ứng trong cơ sở dữ liệu RDMS (Relationship Database Management System). Collection nằm trong một cơ sở dữ liệu duy nhất. Các collection không phải định nghĩa các cột, các hàng hay kiểu dữ liệu trước.

**Database –** Nơi chức các Collection, giống với cơ sở dữ liệu quan hệ, chúng chức các bảng. Mỗi Database có một tập tin riêng lưu trữ trên bộ nhớ vật lý. Một máy chủ MongoDB có thể chứa nhiều Database.

**Document –** Một bản ghi thuộc một Collection thì được gọi là một Document. Các Document lần lượt bao gồm các trường tên và giá trị.

**Field –** Là một cặp năm – value trong một document. Một document có thể có không hoặc nhiều trường. Các trường giống các cột ở cơ sở dữ liệu quan hệ.



* + - * 1. Một document trong MongoDB
      1. Express

Express là một framework nhỏ nhưng linh hoạt, được xây dựng trên nền tảng NodeJS. Nó cung cấp nhiều tính năng mạnh mẽ để phát triển web và ứng dụng di động. Express có nhều package hỗ trợ lập trình. Một chức năng của express chính là thiết lập các lớp trung gian để trả về các request, đjnh nghĩa 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.

* + - 1. NodeJS

NodeJS là một mã nguồn mở được xây dựng trên các nền tảng JavaScript V8 Engine. NodeJS sử dụng rộng bởi hàng ngàn lập trình viên trên toàn thế giới. Nó có thể chạy trên nhiều nền tảng điều hành như Windows, Linux, MacOS. NodeJS cung cấp các thư viện phong phú ở dạng JavaScript Module khác nhau giúp đơn giản hoá việc lập trình và giảm thời gian ở mức thấp nhất. Đặc tính nổi trội của NodeJS là tính bất đồng bộ. Điều này giúp các request riêng lẽ được xử lý đồng thời và gần như là ngay lập tức.

* + 1. Các nền tảng phía Service AI
       1. Ngôn ngữ Python

Python là một ngôn ngữ lập trình bậc cao cho các mục đích lập trình đa năng do Guido van Rossum tạo ra và lần đầu ra mắt năm 1991. Python được thiết kế với ưu điểm mạnh là dễ độc, dễ học và dễ nhớ. Python là ngôn ngữ có hình thức, cấu trúc rõ rang, thuận tiện cho người mới học lập trình. Python còn được tích hợp các thư viện và công cụ mạnh mẽ trong việc xây dựng các service AI, nơi thiên về toán học và giải thuật hơn. Nhờ hai điều này mà mà Python được ưa chuộng khi phát triển các service AI.

* + - 1. Framework Flask

Flask là một web frameworks thuộc loại micro-framework, được xây dựng bằng ngôn ngữ lập trình Python Flask cho phép xây dựng các ứng dụng web từ đơn giản tới phức tạp. Flask có thể xây dựng các api nhỏ, ứng dụng web như các trang web, blog, trang wiki hoặc một website dựa theo thời gian hay thậm chí là một trang web thương mại điện tử. Flask là một micro-framework. Điều này có nghĩa Flask là một môi trường độc lập, ít sử dụng các thư viện bên ngoài. Do vậy, flask có ưu điểm là nhẹ, có rất ít lỗi cũng như dễ dàng phát hiện và xử lý các lỗi bảo mật.

* + - 1. Thư viện Scikit-learn

Như đã được đề cập, Python được tích hợp các thư viện mạnh để phát triển service về AI và một trong số chúng là Scikit-learn. Thư viện cung cấp một tập các công cụ xử lý các bài toán về machine-learning và statistical modeling gồm: classification, regression, clustering,… Thư viện được cấp phép bản quyền chuẩn FreeBSD và chạu được trên nhiều nền tảng. Scikit-learn được sử dụng như một tài liệu để học tập.

* + - 1. Pymongo

Để kết nối đến CSDL trong python, cần một driver để có thể truy cập vào CSDL. Pymongo là một driver cho phép kết nối đến MongoDB.

* 1. Thương mại điện tử

Thương mại điện tử được hiểu là việc thực hiện hoạt động kinh doanh qua các phương tiện điện tử. Nó dựa trên việc xử lý và truyền dữ liệu điện tử dưới dạng văn bản (text), âm thanh và hình ảnh. Thương mại điện tử gồm nhiều hành vi trong đó hoạt động mà bán hàng hoá và dịch vụ qua phương tiện điện tử, giao nhận các nội dung kỹ thuật số trên mạng chuyển tiền điện tử. Thương mại điện tử được thực hiện với các thương mại hàng hoá và thương mại dịch vụ, các hoạt động truyền thống và các hoạt động mới (ví dụ như siêu thị ảo).

* 1. Chatbot
     1. Các khái niệm về chatbot

Các khái niệm chính được sử dụng trong chatbot gồm:

* Intent: Là ý định của người dùng trong câu nói.

Ví dụ: “Tôi muốn tra cứu sản phẩm Áo cầu lông Yonex” từ người dùng, thì ý định của họ ở đây là mong muốn biết các thông tin về sản phẩm Áo cầu lông Yonex như tên đầy đủ, mô tả sản phẩm, giá, …

* Entity: Là các thực thể bổ nghĩa cho một ý định

Ví dụ: “Tôi muốn tra cứu sản phẩm Áo cầu lông Yonex” từ người dùng, thực thể ở đây là “Áo cầu lông Yonex”, là tên của một sản phẩm.

* Stop words: Là những từ xuất hiện nhiều trong văn bản nhưng không có ý nghĩa quan trọng trong việc phân lớp.
* Negative words: Là các từ mang sắc thái tiêu cực
* Positive words: Là các từ mang sắc thái tích cực
  + 1. Các phương pháp xấy dựng chatbot

Chatbot được xây dựng trên các phương pháp sau:

* **Pattern matchers:** Chaatbot xây dựng dựa trên phương pháp pattern matchers sẽ tìm kiếm câu hỏi của người dùng có khớp với các câu hỏi đã được huấn luyện hay không. Phương pháp này dựa vào cấu trúc tiêu chuẩn của ngôn ngữ lập trình “Artificial Intelligence Markup Language” (AIML).

<category>

<pattern>Chatbot là gì </pattern>

<template>

Chatbot là một chương trình máy tính được tạo ra để thực hiện cuộc trò chuyện một cách tự động với người dùng

</template>

</category>

<category>

<pattern>Thế nào là chatbot </pattern>

<template>

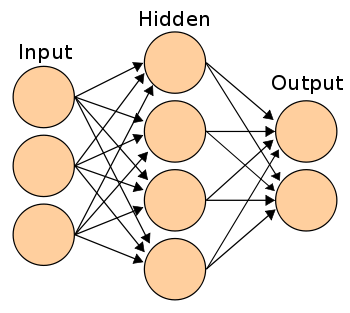
<srai>Chatbot là gì</srai>

</template>

</category>

Nếu người dùng hỏi “Chatbot là gì” hoặc “Thế nào là chatbot” thì sẽ nhận được câu trả lời “Chatbot là một chương trình máy tính được tạo ra dể thực hiện cuộc trò chuyện một cách tự động với người dùng”.

* **Sử dụng thuật toán:** Trong trường hợp có rất nhiều câu hỏi riêng biệt vơi ý định khác nhau đến từ khách hàng, mà chatbot chỉ có thể phản hồi khi tất cả câu hỏi này đã được thiết lập, huấn luyện cho chatbot chỉ có thể phản hồi khi tất cả câu hỏi này đã được thiết lập, huấn luyện cho chatbot với các kịch bản đối thoại đều được lưu trữ trong cơ sở dữ liệu. Do đó khi người dùng hỏi những câu hỏi phức tạp hơn thì cần một phương pháo khác để nhận dạng câu. Vì vậy các nhà phát triển đã đưa các thuật toán vào chatbot để phân loại câu, từ đó tăng độ chính xác cho các phản hồi đưa đến cho người dùng. Các thuật toán có thể sử dụng như SVM, KNN, Navie Bayes,…
* **Artificial Neural Network (ANN – mạng nơ ron nhân tạo):** Phương pháp này là tính toán tỉ lệ chính xác dữ liệu đầu ra từ dữ liệu đầu vào bằng cách tính toán các trọng số cho mỗi kết nối (connection) từ các lần lặp lại trong khi huấn luyện dữ liệu cho chatbot. Mỗi bước huấn luyện đầu ra được xuất ra với độ chính xác cao. Mỗi câu sẽ được phân thành từng từ và mỗi từ chính là dữ liệu đầu vào input đưa vào ANN. Các từ sẽ được kết nối với nhau khi huấn luyện dữ liệu Chatbot, các kết nối này sẽ được đo bằng trọng số cho hàng ngàn lần lặp lại ở các thời điểm khác nhau. Mỗi lần lặp lại trọng số có thể tăng và độ chính xác sẽ cao.



* + - * 1. Minh hoạ một mạng nơ ron nhân tạo
    1. Xử lý ngôn ngữ tự nhiên (NLP)

Xử lý ngôn ngữ tự nhiên (NL) là một nhánh của trí tuệ nhân tạo tập trung vào các ứng dụng trên ngôn ngữ con người. Trong trí tuệ nhân tjao thì xử lý ngôn ngữ tự nhiên là một trong những phần khó nhất vì nó liên quan đến việ phải hiểu ý nghĩa ngôn ngữ - công cụ hoàn hảo nhất của tư duy và giao tiếp.

* + - 1. Các bước xử lý trong NLP
* **Phân tích hình thái câu:** Trong bước này từng từ sẽ được phân tích và ký tự không phải chữ (như các dấu câu) sẽ được tách ra khỏi các từ. Trong tiếng Anh và nhiều ngôn ngữ khác, các từ được phân tách dùng để bằng dấu cách. Tuy nhiên trong tiếng Việt, dấu cách lại được dùng để phân tách tiếng (âm tiết) chứ không phải từ. Cùng với các ngôn ngữ như tiếng Trung, tiếng Hàn, tiếng Nhật, phân tacsh từ trong tiếng Việt là một công việc không hề đơn giản.
* **Phân tíc cú pháp:** Dãy các từ sẽ được biến đổi thành các cấu trúc thể hiện sự liên kết giữa các từ này. Sẽ có những dãy từ bị loại bỏ do vi phạm các luật văn phạm.
* **Phân tích ngữ nghĩa:** Thêm ngữ nghĩa vào các cấu trúc có thể phụ thuộc vào những câu đứng trước, đồng thời nó cũng có thể ảnh hưởng đến câu phía sau.
* **Tích hợp văn bản:** Ngữ nghĩa của một câu riêng biệt có thể phụ thuộc vào bộ những câu đứng trước, đồng thời nó cũng có thể ảnh hưởng đến các câu phía sau.
* **Phân tích thực nghĩa:** Cấu trúc thể hiện điều được phát ngôn sẽ được thông dịch lại để xác định nó thật sự có nghĩa là gì.

Tuy nhiên, ranh giới giữa 5 bước xử lý này cũng rất mỏng manh. Chúng có thể được tiến hành cùng một bước, hoặc tiến hành cùng lúc – tuỳ thuộc vào giải thuật và ngữ cảnh cụ thể.

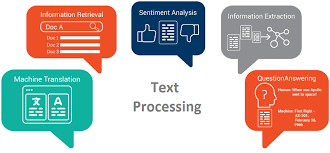
* + - 1. Các ứng dụng cơ bản của NLP và các bài toán trong xử lý tiếng Việt

NLP được ứng dụng vào thực tế như sau:

* Nhận dạng chữ viết
* Nhận dạng tiếng nói
* Tổng hợp tiếng nói
* Dịch tự động
* Tìm kiếm thông tin
* Tóm tắt văn bản
* Khai phá dữ liệu

Trong tiếng việt, NLP được ứng dụng vào các bài toán như:

* Phân tích câu
* Phân tách từ
* Tự động thêm dấu



* + - * 1. Các ứng dụng của NLP
      1. Phương pháp Bag-of-word

**Bag of Words** (BoW – mô hinh túi từ) là một thuật toán hỗ trợ xử lý ngôn ngữ tự nhiên và mục đích của BoW là phân loại câu hay văn bản. Ý tưởng của BoW là phân tích và phân nhóm dựa theo “Bag of Words” (corpus).

Xét hai ví dụ sau:

1. Tôi thích cầu lông. Em tôi cũng vậy.
2. Bạn tôi thì thích bóng đá.

Dưaj vào hai văn bản trên, xây dựng được danh sách như sau:

1. “tôi”, “thích”, “cầu”, “lông”, “em”, “tôi”, “cũng”, “vậy”
2. “bạn”, “tôi”, “thì”, “thích”, “bóng”, “đá”

Dựa vào hai danh sách trên, ta xây dựng được một từ điểm gồm 10 từ như sau:

1. [“tôi”, “thích”, “cầu”, “lông”, “em”, “cũng”, “vậy”, “bạn”, “thì”, “bóng”, “đá”]

Mỗi văn bản sẽ tạo ra một vector đặc trưng có số chiều bằng 10, mỗi phần tử đại diện cho số từ tương ứng xuất hiện trong văn bản đó. Hai văn bản trên sẽ có hai vector đặc trưng là:

1. [2,1,1,1,1,1,1,0,0,0]
2. [1,1,1,0,0,0,0,1,1,1]

Tuy nhiên, vector đại diện này không bao gồm thứ tự qua các từ trong câu gốc.

* + - 1. Phương pháp TF-IDF

**TF-IDF** (Temp Frenquecy – Inverse Document Frenquency) là một kỹ thuật sử dụng trong khai phá dữ liệu văn bản. Trọng số này được sử dụng để đánh giá tầm quan trọng của một từ trong một văn bản. Giá trị cao thể hiện độ quan trọng cao và nó phụ thuộc vào só lần xuất hiện trong văn bản nhưng bù lại bởi tần suất của từ đó trong tập dữ liệu. Một vài biến thể thường được sử dụng trong các hệ thống tìm kiếm như một công cụ để đánh giá và sắp xếp văn bản dựa vào truy vấn của người dùng. Tf-idf cũng được sử dụng để lọc những từ stopwords trong các bài toán như tóm tắt văn bản và phân loại văn bản.

**TF:** Term Frequency (Tần suất xuất hiện của từ) là số lần xuất hiện trong văn bản. Vì các văn bản có thể có độ dài ngắn khác nhau nên một số từ có thể xuất hiện nhiều lần trong một văn bản dài hơn là một văn bản ngắn, Như vậy, term frequency thường được chia cho độ dài văn bản (tổng số từ trong một văn bản).

Trong đó:

* tf (t,d): tần suất xuất hiện của từ t trong văm bản
* f(t, d): số lần xuất hiện của từ t trong văn bản d
* max(f{w, d) : w }): số lần xuất hiện của từ có số lần xuất hiện nhiều nhất trong văn bản d

**IDF:** Inverse Document Frequency (Nghịch đảo tần suất của văn bản), giúp đánh giá tầm quan trọng của một từ. Khi tính toán TF, tất cả các từ được coi như có độ quan trọng bằng nhau. Nhưng một số từ “là”, “hoặc”, “và” thường xuất hiện rất nhiều nhưng độ quan trọng lại không cao. Như vậy cần giảm độ quan trọng của những từ này xuống.

Trong đó:

* idf (t, D): giá trị idf của từ trong tập văn bản
* |D| : tổng số văn bản trong tập D
* |{d D : t d}| : thể hiện số văn bản trong tập D có chứa từ t

Cơ số logarit trong côcng thức này không thay đổi giá trị idf của từ mà chỉ thu hẹp khoảng giá trị của từ đó. Vì thay đổi cơ số sẽ dấn đến việc trị của các từ thay đổi bởi một số nhất định và tỷ lệ giữa chúng thì không thay đổi. Việc sử dụng logarit nhằm giúp giá trị tf-idf của một từ nhỏ hơn, do công thức tính của một từ trong 1 văm bản là tích của tf và idf của từ đó.

Xét với ví dụ trên, sau khi thực hiện vector hoá bằng BoW, ta có kết quả TF-IDF như sau:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tôi | Thích | Cầu | Lông | Em | Cũng | Vậy | Bạn | Thì | Đá | Bóng |
| 0.00 | 0.00 | 0.00 | 0.09 | 0.09 | 0.09 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 | 0.12 | 0.12 | 0.12 |