ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Nguyễn Tiến Hải

GIẢI PHÁP TRUY VẤN LỊCH SỬ DUYỆT WEB BẰNG HỌC MÁY

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Khoa học máy tính

HÀ NỘI - 2024

VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Nguyen Tien Hai

A MACHINE LEARNING-BASED SOLUTION FOR QUERYING WEB BROWSING HISTORY

BACHELOR'S THESIS

Major: Computer Science

Supervisor: Dr. Tran Manh Cuong

HANOI - 2024

LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời cảm ơn cán bộ hướng dẫn ThS. Trần Mạnh Cường đã tận tình chỉ đạo, định hướng và động viên cho em trong toàn bộ quá trình thực hiện khóa luận tốt nghiệp này.

Tiếp theo, em cũng xin gửi lời cảm ơn các anh chị, các bạn và các em sinh viên trong phòng thí nghiệm Bộ môn Công nghệ phần mềm đã đồng hành và giúp em rất nhiều trong việc hỗ trợ hoàn thành khóa luận.

Em cũng xin gửi lời cảm ơn tới các thầy cô giáo trong Đại học Công Nghệ - Đại học Quốc gia Hà Nội đã giảng dạy và cung cấp cho em các kiến thức chuyên môn quan trọng và quý báu để em có được nền tảng như ngày hôm nay.

Con xin cảm ơn bố mẹ đã luôn là nguồn chăm sóc, nguồn động lực quý báu cho con vượt qua các khó khăn trên con đường học vấn.

Trong quá trình làm khóa luận tốt nghiệp, do trình độ và kinh nghiệm thực tiễn còn hạn chế, khóa luận không thể tránh khỏi những thiếu sót. Em rất mong được sự chỉ dẫn và góp ý từ các thầy cô nhằm khắc phục những thiếu sót này.

Tôi xin chân thành cảm ơn!

LỜI CAM ĐOAN

Tôi xin cam đoan khóa luận tốt nghiệp với đề tài "Giải pháp truy vấn lịch sử duyệt web bằng học máy" là do tôi tự tay tìm hiểu và thực hiện dưới sự hướng dẫn của ThS. Trần Mạnh Cường. Tất cả các nội dung trình bày trong khóa luận đều là trung thực. Mọi thông tin tham khảo từ các tài liệu ngoài đều đã được trích dẫn nguồn đầy đủ. Tôi xin chịu trách nhiệm về lời cam đoan này.

Hà Nội, ngày 14 tháng 11 năm 2024

Sinh viên

Nguyễn Tiến Hải

TÓM TẮT

Tóm tắt: Trong thời đại công nghệ phát triển nhanh chóng, người dùng internet thường truy cập và lưu trữ một lượng lớn lịch sử trình duyệt, dẫn đến khó khăn trong việc tìm lại các trang web đã truy cập. Việc tìm kiếm thủ công không chỉ mất thời gian mà còn hạn chế về độ chính xác, đặc biệt khi người dùng chỉ nhớ được một số đặc điểm không đầy đủ của trang web như tiêu đề, nội dung, màu sắc, thời gian hoặc thể loại.

Khóa luận tốt nghiệp này tập trung vào việc giải quyết vấn đề trên bằng cách ứng dụng học máy để xây dựng một hệ thống thông minh hỗ trợ tìm kiếm các trang web từ lịch sử trình duyệt dựa trên các mô tả không đầy đủ của người dùng. Hệ thống sẽ khai thác và phân tích các đặc điểm nổi bật của trang web, kết hợp với khả năng xử lý ngôn ngữ tự nhiên để hiểu và truy xuất thông tin chính xác. Ngoài ra, hệ thống còn hướng đến việc tối ưu hóa giao diện và trải nghiệm người dùng, giúp người dùng dễ dàng tương tác và tìm kiếm hiệu quả.

Với sự kết hợp của các công nghệ hiện đại như học máy, xử lý dữ liệu lớn và giao diện thân thiện, hệ thống được kỳ vọng không chỉ cải thiện quá trình tìm kiếm mà còn mang lại trải nghiệm mới mẻ, thông minh hơn trong việc quản lý lịch sử trình duyệt.

Từ khóa: *Tìm kiếm lịch sử trình duyệt, Học máy, Quản lý thông tin, Giao diện người dùng, Truy vấn thông minh*

ABSTRACT

Abstract: In the era of rapid technological advancements, internet users generate and accumulate extensive browser histories, which often leads to challenges in retrieving previously visited websites. Traditional manual search methods are inefficient and frequently lack precision, particularly when users can only recall partial information about the web pages, such as their titles, content, color schemes, timestamps, or categories.

This thesis explores the application of machine learning to design and implement an intelligent system for querying browser history. The system leverages advanced feature extraction techniques to analyze key attributes of web pages and integrates natural language processing (NLP) to interpret and respond to user queries effectively. Furthermore, the system prioritizes usability by offering an intuitive interface and a seamless user experience, enabling efficient and accurate retrieval of browsing data.

By utilizing cutting-edge technologies such as machine learning, big data analytics, and NLP, this system aims to enhance the effectiveness and efficiency of browser history management. The proposed solution not only addresses the limitations of existing methods but also offers a novel approach to improving the accessibility and usability of personal browsing records.

Keywords: Browser history retrieval, Machine learning, Natural language processing, User interface design, Intelligent search system

Mục lục

Danh sách hình vẽ

Danh sách bảng

Bảng thuật ngữ

STT	Từ viết tắt	Cụm từ đầy đủ	Cụm từ tiếng Việt
1	AI	Artificial Intelligence	Trí tuệ nhân tạo
2	FM	Foundation Model	Mô hình cơ sở
3	GPT	Generative Pre-trained Transformer	Mô hình sinh tạo có trước và sử dụng Transformer
4	LLM	Large Language Model	Mô hình ngôn ngữ lớn
5	LM	Language Model	Mô hình ngôn ngữ
6	NLP	Natural Language Processing	Xử lý ngôn ngữ tự nhiên
7	PD	Predicted Document	Số lượng văn bản pháp luật trong câu trả lời dự đoán
8	RAG	Retrieval Augmented Generation	Tạo tăng cường truy xuất
9	RD	Reference Document	Số lượng văn bản pháp luật trong câu trả lời thực tế

Chương 1

Giới thiệu

1.1 Đặt vấn đề

Ngày nay, trong bối cảnh xã hội ngày càng số hóa mạnh mẽ và mọi khía cạnh của cuộc sống từ làm việc, học tập đến giải trí đều chuyển dịch sang môi trường trực tuyến, trình duyệt web đã trở thành một công cụ không thể thiếu đối với con người. Chính vì điều này dẫn đến sự gia tăng nhanh chóng về khối lượng thông tin được người dùng truy cập, tạo nên một lượng lớn dữ liệu lịch sử duyệt web. Tuy nhiên, nhu cầu tìm kiếm lại các trang web đã truy cập một cách nhanh chóng, chính xác và hiệu quả đang đặt ra nhiều thách thức đối với người dùng. Mặc dù các trình duyệt hiện nay đã cung cấp tính năng lưu trữ lịch sử duyệt web, nhưng vẫn tồn tại nhiều hạn chế. Điều này bắt nguồn từ ba nguyên nhân chính.

Nguyên nhân thứ nhất, khối lượng lịch sử duyệt web ngày càng lớn do tần suất sử dụng trình duyệt cao. Người dùng thường xuyên phải tìm kiếm lại các trang web đã truy cập, nhưng việc ghi nhớ các từ khóa chính xác hoặc thời điểm cụ thể là điều không dễ dàng.

Nguyên nhân thứ hai, các công cụ tìm kiếm lịch sử hiện tại trên trình duyệt chỉ hỗ trợ truy vấn đơn giản dựa trên tiêu đề hoặc URL, chưa đáp ứng được nhu cầu tìm kiếm dựa trên các đặc điểm như nội dung trang web, màu sắc giao diện, hoặc thời gian truy cập tương đối. Điều này dẫn đến việc người dùng mất nhiều thời gian để tìm kiếm thông tin mình cần.

Và cuối cùng, trong một số trường hợp, người dùng cần quản lý và phân loại lịch sử duyệt web theo các nhóm liên quan hoặc tìm kiếm thông tin dựa trên mô tả ngữ nghĩa tự nhiên. Tuy nhiên, các trình duyệt chưa có tính năng hỗ trợ tự động hóa các tác vụ này, khiến người dùng phải thao tác thủ công, tốn nhiều công sức và thời gian.

Hiện tại, việc tìm kiếm lịch sử duyệt web vẫn còn nhiều hạn chế, và nhu cầu có một giải pháp thông minh hơn là rất cần thiết. Giải pháp này không chỉ cần hỗ trợ

truy vấn lịch sử nhanh chóng, chính xác, mà còn phải linh hoạt trong việc phân loại, gợi ý thông tin phù hợp và cải thiện trải nghiệm người dùng. Do đó, giải pháp truy vấn lịch sử duyệt web bằng học máy ra đời nhằm khắc phục những vấn đề hiện tại, nâng cao hiệu quả và sự tiện dụng trong quá trình tìm kiếm lịch sử duyệt web.

1.2 Phân tích thực trạng

Trong bối cảnh số hóa đang ngày càng phát triển, trình duyệt web không chỉ đóng vai trò là công cụ truy cập Internet mà còn là phương tiện chính để người dùng tương tác với các hệ thống số. Mặc dù có sự phổ biến rộng rãi và khả năng hỗ trợ người dùng cơ bản, các trình duyệt hiện nay vẫn tồn tại nhiều hạn chế khi xử lý khối lượng lớn dữ liệu lịch sử duyệt web và các yêu cầu tìm kiếm phức tạp. Những hạn chế này đã được nhận diện qua các phân tích thực trạng cụ thể như sau:

1.2.1 Khối lượng dữ liệu lớn và sự thiếu tổ chức

Trung bình mỗi người dùng Internet dành hơn **6 giờ mỗi ngày** trên các thiết bị số [**?**]. Tại Việt Nam, số lượng người dùng Internet đã đạt **77,93 triệu người**, tương ứng với **79,1%** dân số [**?**]. Mỗi người dùng truy cập hàng trăm trang web mỗi tuần với các mục đích khác nhau, dẫn đến việc tích lũy một lượng lớn dữ liệu lịch sử duyệt web. Tuy nhiên, các trình duyệt hiện nay không cung cấp cơ chế đủ mạnh để tự động tổ chức và quản lý lượng dữ liệu khổng lồ này. Điều này khiến người dùng gặp khó khăn trong việc tìm lại các trang web quan trọng, đặc biệt khi cần truy vấn lịch sử đã lưu từ trước đó.

1.2.2 Hạn chế trong công cụ tìm kiếm và quản lý lịch sử

Mặc dù các trình duyệt hiện nay đã tích hợp các công cụ tìm kiếm lịch sử cơ bản, nhưng các tính năng này chủ yếu chỉ hỗ trợ tìm kiếm theo các yếu tố đơn giản như tiêu đề, URL và số lần truy cập. Điều này tạo ra khó khăn khi người dùng không nhớ chính xác tiêu đề hoặc địa chỉ URL của trang web đã truy cập. Hệ thống tìm kiếm hiện tại thiếu khả năng xử lý các tìm kiếm phức tạp hơn, chẳng hạn như tìm kiếm dựa trên nội dung trang web, màu sắc giao diện hoặc thời gian truy cập tương

đối. Điều này khiến người dùng gặp khó khăn trong việc truy xuất lại các trang đã xem, đặc biệt khi không có những thông tin cơ bản như tiêu đề hoặc URL.

Các công cụ tìm kiếm lịch sử trong trình duyệt không cung cấp các tính năng tìm kiếm nâng cao, chẳng hạn như khả năng nhận diện nội dung chính của các trang web đã duyệt hoặc gợi ý các trang web có màu sắc hoặc chủ đề tương tự. Do đó, việc tìm kiếm thông tin trong một lượng lớn lịch sử duyệt web trở nên khó khăn và tốn thời gian, đặc biệt đối với những người dùng có nhu cầu tìm kiếm phức tạp hoặc khi muốn tìm lại các trang web có chủ đề liên quan.

- **Microsoft Edge** và **Mozilla Firefox**: Các trình duyệt này chỉ hỗ trợ tìm kiếm lịch sử dựa trên các yếu tố cơ bản như tên trang, URL, và số lần truy cập. Tuy nhiên, các tính năng tìm kiếm này chưa hỗ trợ tìm kiếm nâng cao dựa trên nội dung trang, màu sắc hoặc thời gian.
- **Google Chrome**: Chrome nổi bật hơn với tính năng **"Nhóm lịch sử"** (trước đây gọi là *"Hành trình"*), được giới thiệu vào năm 2022. Tính năng này tự động nhóm các trang web liên quan dựa trên chủ đề hoặc hoạt động duyệt web của người dùng, giúp việc tìm lại thông tin trở nên dễ dàng hơn. Đặc biệt, tính năng này không chỉ hoạt động với *Google Search* mà còn hỗ trợ cả các công cụ tìm kiếm khác, nâng cao khả năng quản lý lịch sử duyệt web cho người dùng [**?**].

Tuy nhiên, ngay cả với tính năng nhóm lịch sử này, các trình duyệt vẫn còn thiếu một số khả năng quan trọng:

- **Thiếu khả năng phân loại thông minh:** Các trình duyệt hiện nay chưa hỗ trợ phân loại tự động lịch sử duyệt web thành các danh mục như "học tập," "giải trí," "mua sắm," hay "công việc." Điều này khiến người dùng phải tự quản lý thủ công, gây mất thời gian và giảm hiệu quả.
- **Gợi ý thông minh còn hạn chế:** Các hệ thống quản lý lịch sử chưa thể đưa ra các gợi ý về trang web liên quan dựa trên hành vi duyệt web trước đó. Ví dụ, một người đang nghiên cứu một chủ đề cụ thể có thể không nhận được gợi ý phù hợp từ những trang web đã truy cập trước đây có nội dung tương đồng.

- **Hiển thị lịch sử chưa trực quan:** Các trình duyệt chưa cung cấp giao diện hiển thị lịch sử dưới dạng dòng thời gian hoặc nhóm chủ đề rõ ràng. Điều này làm cho người dùng khó theo dõi và quản lý các hoạt động duyệt web một cách hiệu quả.

Những hạn chế trên cho thấy sự cần thiết của việc cải tiến hệ thống quản lý lịch sử duyệt web, nhằm mang lại trải nghiệm tốt hơn và tăng cường khả năng truy xuất thông tin cho người dùng.

1.3 Mục tiêu khóa luận

Với sự gia tăng nhanh chóng của lượng dữ liệu trên Internet, việc quản lý và tìm kiếm lịch sử duyệt web trở thành một vấn đề ngày càng quan trọng đối với người dùng. Khóa luận này tập trung vào việc xây dựng một hệ thống quản lý lịch sử duyệt web thông minh, giúp người dùng dễ dàng tìm kiếm lại các trang web đã truy cập, phân loại các trang web vào các danh mục phù hợp và cải thiện trải nghiệm duyệt web tổng thể.

Hệ thống sẽ được thiết kế để giải quyết một số vấn đề nổi bật trong việc quản lý lịch sử duyệt web hiện tại, bao gồm:

- **Tìm kiếm nâng cao**: Hệ thống sẽ cung cấp tính năng tìm kiếm lịch sử duyệt web dựa trên các yếu tố không chỉ dừng lại ở tiêu đề và URL mà còn mở rộng ra các yếu tố như nội dung chính của trang web, màu sắc giao diện và thời gian truy cập. Điều này giúp người dùng có thể dễ dàng tìm lại các trang web ngay cả khi không nhớ rõ tiêu đề hoặc URL.

- **Phân loại lịch sử duyệt web tự động**: Một trong những tính năng quan trọng của hệ thống là khả năng phân loại lịch sử duyệt web vào các danh mục như "giải trí," "học tập," "mua sắm," "công việc," và nhiều hơn nữa. Việc phân loại tự động này sẽ giúp người dùng quản lý và theo dõi các hoạt động duyệt web hiệu quả hơn, tiết kiệm thời gian và giảm thiểu sự rối loạn trong lịch sử duyệt web.

- **Gợi ý thông minh**: Hệ thống sẽ cung cấp các gợi ý thông minh dựa trên hành vi duyệt web của người dùng. Điều này có nghĩa là người dùng sẽ nhận

được các gợi ý về các trang web liên quan, giúp họ tiếp cận thông tin mới một cách nhanh chóng và thuận tiện.

- **Nhóm lịch sử duyệt web**: Hệ thống sẽ cho phép người dùng xem lịch sử duyệt web được nhóm lại theo chủ đề hoặc hoạt động tương đồng. Tính năng này, lấy cảm hứng từ các công cụ như *Google Chrome's Group History*, sẽ giúp người dùng dễ dàng theo dõi chuỗi các trang web liên quan và phục vụ nhu cầu truy vấn thông tin hiệu quả hơn.

Mục tiêu của khóa luận này là không chỉ nâng cao hiệu quả trong việc tìm kiếm và quản lý lịch sử duyệt web mà còn tạo ra một công cụ mạnh mẽ giúp người dùng duyệt web một cách thông minh và an toàn hơn. Hệ thống này sẽ mang lại một bước tiến quan trọng trong việc tối ưu hóa trải nghiệm duyệt web và giúp người dùng tiết kiệm thời gian, tăng cường hiệu suất làm việc và dễ dàng truy xuất lại thông tin quan trọng từ lịch sử duyệt web của mình.

1.4 Cấu trúc khóa luận

Phần còn lại của khóa luận có cấu trúc như sau:

Chương 2 cung cấp nền tảng lý thuyết, bao gồm các khái niệm cốt lõi về hệ thống tìm kiếm, thuật toán xử lý văn bản, phương pháp đo lường sự tương đồng giữa các văn bản, và công nghệ chính được áp dụng trong nghiên cứu.

Chương 3 tập trung mô tả quá trình thiết kế và xây dựng công cụ, nhấn mạnh vào kiến trúc tổng quan của hệ thống và cách thức xử lý dữ liệu lịch sử duyệt web.

Chương 4 trình bày chi tiết về xử lý truy vấn và đánh giá mức độ tương đồng. Các nội dung chính bao gồm kiểm tra phạm vi câu hỏi, xử lý thông tin đầu vào từ người dùng, và các phương pháp tính điểm để đưa ra kết quả tìm kiếm chính xác.

Chương 5 trình bày các thử nghiệm được thực hiện để đánh giá hệ thống, bao gồm quy trình kiểm thử, chỉ số đo lường hiệu quả, và phân tích chi tiết kết quả nhằm làm rõ hiệu suất và tiềm năng ứng dụng thực tiễn.

Cuối cùng, **Chương 6** tóm tắt những đóng góp đạt được trong khóa luận, đồng thời đưa ra các định hướng phát triển và mở rộng hệ thống trong tương lai.

Chương 2

Cơ sở lý thuyết

Trong chương này, chúng ta sẽ thảo luận về hai khía cạnh quan trọng liên quan đến hệ thống được xây dựng: tiện ích mở rộng trình duyệt (browser extension) và các khái niệm cơ bản về học máy và xử lý ngôn ngữ tự nhiên. Nội dung bao gồm định nghĩa, điểm mạnh, cấu trúc phát triển của các tiện ích mở rộng, cũng như các mô hình học máy và các ứng dụng trong phân tích lịch sử duyệt web.

2.1 Tiện ích mở rộng trình duyệt

Tiện ích mở rộng trình duyệt là các ứng dụng nhỏ được phát triển để mở rộng tính năng của trình duyệt web, cung cấp các chức năng bổ sung hoặc tùy chỉnh giao diện và hành vi của trình duyệt. Các tiện ích mở rộng này hoạt động dựa trên giao diện lập trình ứng dụng (API) do các trình duyệt cung cấp, cho phép truy cập và tương tác với các thành phần của trình duyệt hoặc nội dung của trang web.

2.1.1 Điểm mạnh của tiện ích mở rộng

Tiện ích mở rộng có nhiều ưu điểm vượt trội, đặc biệt trong việc nâng cao trải nghiệm người dùng và tối ưu hóa hiệu suất trình duyệt:

- **Khả năng tùy chỉnh:** Người dùng có thể tùy chỉnh giao diện và hành vi trình duyệt theo nhu cầu cá nhân.
- **Tích hợp dễ dàng:** Các tiện ích mở rộng hoạt động liền mạch trên trình duyệt mà không yêu cầu cài đặt phần mềm bổ sung.
- **Tiện lợi:** Cung cấp các chức năng bổ sung mà trình duyệt mặc định không có, như quản lý lịch sử, ghi chú, hoặc đồng bộ hóa dữ liệu.
- **Bảo mật:** Tiện ích mở rộng hiện đại được thiết kế với các chính sách bảo mật nghiêm ngặt, giúp bảo vệ người dùng khỏi các rủi ro tiềm ẩn.

2.1.2 Cấu trúc phát triển của tiện ích mở rộng

Một tiện ích mở rộng trình duyệt thường bao gồm các thành phần chính sau:

- **Tệp cấu hình (`manifest.json`):** Chứa thông tin cơ bản về tiện ích, bao gồm quyền truy cập, tệp JavaScript, và các thành phần khác.
- **Giao diện người dùng (UI):** Bao gồm các tệp HTML, CSS, và JavaScript để hiển thị và tương tác với người dùng.
- **Tập lệnh nền (Background Script):** Xử lý các sự kiện quan trọng như truy vấn dữ liệu, lắng nghe thay đổi, và giao tiếp với backend.
- **Tập lệnh nội dung (Content Script):** Tương tác trực tiếp với nội dung của trang web, như trích xuất dữ liệu hoặc thay đổi giao diện hiển thị.
- **Quyền và chính sách bảo mật:** Xác định quyền truy cập và đảm bảo dữ liệu của người dùng được bảo vệ.
- **Phân phối và cập nhật:** Sau khi phát triển, tiện ích được đóng gói và đăng tải lên các cửa hàng trực tuyến như *Chrome Web Store* hoặc *Mozilla Add-ons*.

2.1.3 Ứng dụng trong hệ thống

Trong khóa luận này, tiện ích mở rộng được sử dụng để:

- Thu thập dữ liệu lịch sử duyệt web từ trình duyệt và gửi đến backend để phân tích.
- Hiển thị kết quả phân tích, bao gồm các nhóm lịch sử (history lines) và gợi ý thông minh.
- Tạo giao diện người dùng thân thiện để hỗ trợ việc tìm kiếm và quản lý lịch sử.

2.2 Học máy và xử lý ngôn ngữ tự nhiên

2.2.1 Học máy (Machine Learning)

Học máy là một nhánh của trí tuệ nhân tạo (AI), cho phép máy tính học hỏi và đưa ra quyết định dựa trên dữ liệu. Các ứng dụng học máy trong hệ thống này bao gồm:

- Phân tích và trích xuất các đặc điểm từ lịch sử duyệt web, như tiêu đề, nội dung chính, màu sắc, và thời gian truy cập.
- Xây dựng hệ thống gợi ý thông minh để đề xuất các lịch sử liên quan dựa trên hành vi người dùng.
- Tính toán và đánh giá mức độ tương đồng giữa các bản ghi lịch sử để nhóm chúng thành các "line lịch sử."

2.2.2 Xử lý ngôn ngữ tự nhiên (NLP)

Xử lý ngôn ngữ tự nhiên (Natural Language Processing) là lĩnh vực nghiên cứu cách máy tính tương tác với ngôn ngữ con người. Trong hệ thống này, NLP được áp dụng để:

- Trích xuất thông tin từ các câu hỏi hoặc mô tả của người dùng.
- Phân tích nội dung chính của các trang web để tạo embedding cho các tính năng tìm kiếm.
- Gợi ý các nhóm lịch sử dựa trên nội dung văn bản và ngữ cảnh.

2.2.3 Các công cụ và mô hình sử dụng

- **Transformers:** Sử dụng các mô hình ngôn ngữ như BERT hoặc RoBERTa để tính toán embeddings cho nội dung và tiêu đề trang web.
- **Faiss:** Công cụ tìm kiếm dựa trên vector, giúp xử lý và so khớp các embeddings để tìm kiếm lịch sử liên quan.

- **BeautifulSoup và OpenCV:** Phân tích nội dung và trích xuất màu sắc chủ đạo từ các trang web.

2.2.4 Ứng dụng trong hệ thống

Học máy và xử lý ngôn ngữ tự nhiên đóng vai trò trung tâm trong hệ thống, hỗ trợ:

- Phân loại lịch sử duyệt web theo các danh mục như học tập, giải trí, công việc, hoặc mua sắm.
- Xây dựng hệ thống gợi ý thông minh, giúp người dùng tìm lại các lịch sử liên quan dựa trên các mô tả ngắn gọn.
- Tăng cường trải nghiệm người dùng bằng các tính năng tìm kiếm nâng cao và phân tích dữ liệu tự động.

Chương 3

Thiết kế và phát triển công cụ

Chương này sẽ trình bày về giải pháp mà khóa luận tốt nghiệp đã phát triển, bao gồm các yêu cầu chức năng, yêu cầu phi chức năng và kiến trúc giải pháp cho hệ thống. Ngoài ra, chương cũng sẽ giới thiệu một số phương pháp kiểm thử để đảm bảo hệ thống hoạt động đúng chức năng và phạm vi, cùng với đó là trình bày các hướng phát triển cho phần mềm trong tương lai.

3.1 Cấu trúc tổng quan

3.2 Phân tích dữ liệu lịch sử duyệt web

Chương 4

Xử lý Truy vấn và Đánh giá Mức độ Tương đồng

4.1 Kiểm tra Phạm vi Câu hỏi

4.2 Xử lý và Đánh giá các Yếu tố Tìm kiếm

4.3 Tính toán Điểm Mức độ Tương đồng

4.4 Công thức Tổng hợp Điểm

Chương 5

Thực nghiệm và Đánh giá Hiệu quả Hệ thống

5.1 Quy trình Thực nghiệm

5.2 Các Đo lường và Chỉ Số Đánh giá

5.3 Kết quả Thực nghiệm và Phân tích

Chương 6

Kết luận