

**Họ và tên
DUƠNG VĂN BÁCH**

**TÊN ĐỀ TÀI
XÂY DỰNG HỆ THỐNG THÔNG TIN HỎI ĐÁP
VÀ TÌM KIẾM CHUYÊN GIA TOÁN PHỔ THÔNG**

CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN

ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC

Hà Nội - Năm 2017

ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC

Chuyên ngành: **Hệ thống thông tin**

Mã số: 13150016

Ngày giao đề án: tháng 9 năm 2017

Ngày nộp đề văn: tháng 4 năm 2018

**Tên đề tài: XÂY DỰNG HỆ THỐNG THÔNG TIN HỎI ĐÁP VÀ TÌM KIẾM
CHUYÊN GIA TOÁN PHỔ THÔNG**

Học viên thực hiện: DƯƠNG VĂN BÁCH

Lớp: HTTT12

NGƯỜI HƯỚNG DẪN KHOA HỌC:

1. Hướng dẫn chính:

Họ và tên : Nguyễn Mạnh Hùng

Học hàm học vị : Tiến sĩ

Đơn vị : Khoa CNTT Học viện KTQS

2. Đồng hướng dẫn:

Họ và tên : Đỗ Thị Mai Hương

Học hàm học vị : Thạc sĩ

Đơn vị : Khoa CNTT Học viện KTQS

Hà Nội - Năm 2017

ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC

Tên đề tài: XÂY DỰNG HỆ THỐNG THÔNG TIN HỎI ĐÁP VÀ TÌM KIẾM CHUYÊN GIA TOÁN PHỔ THÔNG

Chuyên ngành: Hệ thống thông tin

Thời gian thực hiện: 7 tháng, từ 9/2017 đến 4/2018

➤ **Cơ sở khoa học và tính thực tiễn của đề tài:**

– **Cơ sở khoa học:**

Cơ sở khoa học của đề tài được dựa trên các công nghệ và thuật toán đã được các tổ chức sử dụng trước đó, trong đó trọng tâm đề cập đến các công nghệ và thuật toán Stack Exchange và Quora đang sử dụng. Đây là các hệ thống trên thực tế được lấy làm khuôn mẫu của đề tài. Dưới đây là tóm tắt sơ lược về các công nghệ đặc thù và thuật toán cần sử dụng cho mục đích *xây dựng hệ thống thông tin hỏi đáp và tìm kiếm chuyên gia toán phổ thông*.

1. Công nghệ

- Máy tìm kiếm(Search Engine): *Elasticsearch*.

Elasticsearch được Stack Exchange sử dụng cho mục đích tìm kiếm câu hỏi. Elasticsearch có ưu điểm là không cần cấu hình phức tạp, nó tự định kiểu cho các dữ liệu mà ta đưa vào. Elasticsearch mặc định được thiết kế cho phép lưu trữ dữ liệu phân tán dẫn đến việc mở rộng hệ thống khá dễ dàng. Nhưng quan trọng nhất, Elasticsearch cho phép tìm kiếm full-text search gần với rất nhanh, dữ liệu được cập nhật vào kho dữ liệu có thể tìm kiếm được sau từ 1,2 giây hoặc ít hơn.

Tuy nhiên nó gặp phải một số nhược điểm như: Không có đảm bảo về xác thực và bảo mật; dữ liệu có thể bị mất khi nó tổ chức lưu trữ phân tán và Elasticsearch không cung cấp cơ chế phục hồi các dữ liệu bị mất này. Elasticsearch cũng không hỗ trợ transaction. Điều này khiến nó không thể được sử dụng như cơ sở dữ liệu chính mà thường chỉ được sử dụng cho mục đích tìm kiếm thuần túy. Mặt khác Elasticsearch không được thiết kế cho tiếng Việt nên khi áp dụng vào tìm kiếm tiếng Việt sẽ gặp phải một số khó khăn do đặc thù tiếng Việt khác với tiếng Anh, ví dụ như: “khoảng trắng” không phải là dấu hiệu phân tách một từ. Do vậy cần sử dụng các thư viện để tách từ tiếng Việt trước khi tìm kiếm trên Elasticsearch để tăng độ chính xác của truy vấn.

- Cơ sở dữ liệu chính(Primary Database): *MsSQL, MySQL*.

MsSQL là cơ sở dữ liệu chính trong kiến trúc của Stack Exchange. MsSQL có đầy đủ các yêu cầu để trở thành một cơ sở dữ liệu chính của một hệ thống bao gồm tính nhất quán và toàn vẹn, tính bảo mật, hỗ trợ transaction.

Tuy nhiên khi triển khai thực tế MySQL là một phần mềm tính phí nên cần tìm một giải pháp tương tự có thể thay thế. MySQL là một phần mềm mã nguồn mở và có cũng có đầy đủ các yêu cầu để trở thành một cơ sở dữ liệu chính. MySQL hoàn toàn có thể thay thế được MsSQL. Quora đang sử dụng MySQL làm cơ sở dữ liệu chính cho hệ thống của họ.

- Bộ cân bằng tải TCP/HTTP(TCP/HTTP Load Balancer): *HAProxy*.

HAProxy được sử dụng cả trong hệ thống của Quora và Stack Echange. Nó là một giải pháp đảm bảo tính sẵn sàng cao, cân bằng tải và proxy cho các hệ thống có sử dụng trao đổi TCP và HTTP. Ưu điểm của HAProxy là miễn phí, nhanh, đáng tin cậy, đặc biệt phù hợp với các website có lượng truy cập cao.

- Máy chủ web(Web server): *NGINX*

NGINX là một máy chủ HTTP hiệu năng cao, một reverse proxy mã nguồn mở, cũng như một máy chủ proxy IMAP / POP3. NGINX được biết đến với tính năng cao, tính ổn định, cung cấp nhiều tính năng, dễ cấu hình và tiêu thụ ít tài nguyên. NGINX có thể mở rộng theo mọi chiều: từ VPS nhỏ đến các cụm máy chủ lớn. Một ưu điểm quan trọng nhất: NGINX là một trong số ít các máy chủ được viết để giải quyết vấn đề C10K, nghĩa là xử lý được 10000 kết nối đồng thời tới hệ thống.

- Máy chủ dữ liệu truy cập nhanh(Cache Server) và kết nối Pub/Sub: *Redis*.

Redis là một cơ sở dữ liệu dạng key-value. Redis cho phép lưu trữ đồng thời dữ liệu trên RAM cho việc truy cập nhanh như memcached và lưu trữ trên đĩa cứng. Ngoài ra một tính năng được sử dụng phổ biến khác của Redis là PubSub. Stack Exchange sử dụng tính năng này cập nhật thông tin mới nhất về một chủ đề đến những người dùng đã đăng ký chủ đề này.

- Máy hiển thị công thức toán học trên trình duyệt(Display engine for mathematics): *Mathjax*.

Mathjax là một thư viện phía client được Stack Exchange sử dụng trên trang *Math.stackexchange.com* để hiển thị công thức toán học. Mathjax hỗ trợ các kiểu định dạng công thức phổ biến như LaTeX, MathML, đồng thời tương thích với tất cả các trình duyệt đời mới, bao gồm IE 6+, Firefox 3+, Chrome 0.2+, Safari 2+, Opera 9.6+, và phần nhiều trình duyệt di động.

- Tách từ tiếng Việt: *vnTokenizer*

Như đã đề cập phía trên, Elasticsearch không được thiết kế cho tiếng Việt nên khi áp dụng vào tìm kiếm tiếng Việt sẽ gặp phải một số khó khăn do đặc thù tiếng Việt khác với tiếng Anh, ví dụ như: “khoảng trắng” không phải là dấu hiệu phân tách một từ. Do vậy cần sử dụng các thư viện để tách từ tiếng Việt trước khi tìm kiếm trên Elasticsearch để tăng độ chính xác của truy vấn.

vnTokenizer của tác giả Lê Hồng Phương là một tokenizer tự động cho phép phân đoạn văn bản tiếng Việt thành các đơn vị từ vựng. Nó được phát triển dựa trên ngôn ngữ lập trình Java và cho kết quả phân đoạn tốt về độ chính xác và tỉ lệ thu hồi trong khoảng từ 96% -98%. Giải pháp tách từ này có thể triển khai trên Elasticsearch như một plugin.

- Công nghệ truyền thông điệp tức thời (Instant messaging and Video Call): *WebRTC*

Facebook đang sử dụng WebRTC để xây dựng tính năng gọi thoại của họ trên Messenger. WebRTC là tập hợp các tiêu chuẩn và giao thức cho phép các trình duyệt Web thực hiện trực tiếp các tính năng truyền thông đa phương tiện thời gian thực như gọi điện, tin nhắn hình, truyền dữ liệu bằng các API JavaScript mà không cần phải cài thêm plugins hay phần mềm hỗ trợ nào từ bên ngoài.

2. Thuật toán

- Thuật toán xếp hạng của Quora

Quora là một trang web cho phép người dùng đăng ký và sau đó đăng câu hỏi trên trang web. Người dùng khác nhau có thể đăng câu trả lời và nó được xếp hạng theo thời gian.

Theo Quora, công ty sử dụng các thông số nhất định để xếp hạng câu trả lời được đăng bởi người dùng cho các câu hỏi trên trang web. Các tham số được liệt kê như sau:

- Số lượng upvotes càng cao, xếp hạng càng cao
- Số lượng downvotes cao hơn, xếp hạng thấp hơn.
- Vote của người dùng viết câu trả lời tốt hơn mang nhiều giá trị hơn những người khác.
- Câu trả lời được viết bởi người dùng đã có những câu trả lời tốt trong quá khứ được xếp hạng cao hơn.
- Người sử dụng spam hoặc can thiệp trái phép hệ thống, phiếu bầu của họ sẽ không được xem xét.
- Cấp bậc không bị ảnh hưởng nếu người dùng là quản trị viên, đánh giá viên.

- Thuật toán xếp hạng của Stack Overflow – một website trong hệ thống Stack Exchange

Stack Overflow là một trang web được người dùng đăng câu hỏi liên quan đến bất kỳ ngôn ngữ lập trình và bất kỳ người dùng nào đều có thể đăng câu trả lời cho câu hỏi. Câu trả lời là sau đó được xếp hạng bằng cách sử dụng một thuật toán. Stack Overflow có một thuật toán để xếp hạng các câu trả lời được đăng cho bất kỳ câu hỏi liên quan đến lập trình trên trang web. Người ta chủ yếu sử dụng hai thông số, đó là upvotes và downvotes, trừ đi các downvotes từ upvotes để tính toán các cấp bậc. Ngoài ra, Stack Overflow còn sử dụng một số yếu tố khác trong việc xác định thứ hạng của người dùng:

- Điểm của người dùng đăng câu hỏi
- Lượt xem cho câu hỏi
- Câu trả lời cho câu hỏi
- Điểm cho câu trả lời được đăng
- Ngày và giờ gửi câu hỏi
- Ngày và giờ xem lần cuối câu hỏi
- Danh tiếng của người dùng đặt câu hỏi
- Danh tiếng của người dùng trả lời câu hỏi

– **Tính thực tiễn:**

Trên thế giới cũng có rất nhiều hệ thống, website, công cụ để giúp sức cho con người giải quyết vấn đề vừa đề cập ở phần trước. Một số ví dụ có thể kể đến như:

- Trên thế giới:
 - Các công cụ tìm kiếm: Google, Yahoo, Bing...
 - Mạng xã hội: Facebook, Twitter, Youtube...
 - Các trang web hỏi đáp: Yahoo Ask & Answer, Math.stackexchange.com, Reddit.com, MathExchange .net...
 - Các trang web học tập: Math.com, Schoolyourself.org, Khanacademy.com, Onlinemathlearning.com, Mathplanet.com...
 - Các diễn đàn: Mathforum.org, Artofproblemsolving.com, Guruquest.net...
- Ở Việt Nam:
 - Các công cụ tìm kiếm: Cốc Cốc...
 - Mạng xã hội: Zalo, Beat.vn, Go.vn, Zing.me...
 - Các trang web hỏi đáp: Hoidapnhanh.vn, Olm.vn, Quickrep.vn, Daynhauhoc.com, Hoc247.net...
 - Các trang web học tập: Luyenthi123.com, Vted.vn, Cunghoc.vn, Luyenthi123.com, Hocmai.vn, Moon.vn, Zuni.vn...
 - Các diễn đàn: Diendantoanhoc.net, Diendan.hocmai.vn, Diendantoanthcs.forumvi.net, K2pi.net.vn...

Tuy nhiên các ví dụ kể trên có những nhược điểm không thể giải quyết hoàn toàn được vấn đề hoặc bản thân nó được ra đời không nhằm mục đích giải quyết vấn đề mà ta đang đặt ra. Dưới đây là phân tích chi tiết về những ưu nhược điểm mà các hệ thống, website, công cụ kể trên.

1. Công cụ tìm kiếm

- *Ví dụ:*
 - Google, Yahoo, Bing, Cốc cốc..
- *Ưu điểm:*

- Giúp người dùng tìm kiếm được hàng triệu, hàng tỉ thông tin liên quan đến thắc mắc, vấn đề của người dùng thông qua từ khóa tìm kiếm.
- *Nhược điểm:*
 - Không cung cấp các thông tin về người có kiến thức trong lĩnh vực có liên quan đến từ khóa.
 - Không cung cấp cách thức khác để có thể người cần tìm kiếm trao đổi trực tiếp với người có kiến thức.

2. Mạng xã hội

- *Ví dụ:*
 - Facebook, Twitter, Youtube, Zalo, Beat.vn, Zing.me..
- *Ưu điểm:*
 - Kết nối con người với con người, người ta hoàn toàn có thể kết nối trực tiếp, trao đổi với nhau;
- *Nhược điểm:*
 - Tìm kiếm nội dung khó khăn.
 - Không có cơ chế để phân biệt được ai là chuyên gia.

3. Trang web hỏi đáp

- *Ví dụ:*
 - Yahoo Ask & Answer, Math.stackexchange.com, Reddit.com, MathExchange .net, Hoidapnhanh.vn, Olm.vn, Quickrep.vn, Daynhauhoc.com, Hoc247.net...
- *Ưu điểm:*
 - Giúp người dùng tìm được thông tin giải quyết vấn đề họ gặp phải nhờ sự trợ giúp của cộng đồng(bao gồm người người có chuyên môn hoặc không có chuyên môn).
- *Nhược điểm:*
 - Không cấp cách thức trao đổi trực tiếp tức thời
 - Người cần giải đáp hoàn toàn bị động, không có cách thức nào để họ có thể chắc chắn rằng câu hỏi, vấn đề sẽ được giải đáp.

4. Trang web học tập

- *Ví dụ:*
 - Math.com, Schoologyself.org, Khanacademy.com, Onlinemathlearning.com, Mathplanet.com, Luyenthi123.com, Vted.vn, Cunghoc.vn, Luyenthi123.com, Hocmai.vn, Moon.vn, Zuni.vn...
- *Ưu điểm:*
 - Cung cấp các khóa học dành cho học sinh và môi trường để học sinh có thể trao đổi, hỏi đáp, trao đổi kiến thức lẫn nhau.
- *Nhược điểm:*

- Không cấp cách thức trao đổi trực tiếp tức thời với chuyên gia.
- Chưa đầu tư nhiều cho giải quyết vấn đề hỏi đáp, thường chỉ tập chung vào việc kinh doanh các khóa học.

5. Các diễn đàn

- *Ví dụ:*

- Mathforum.org, Artofproblemsolving.com, Guruquest.net, Diendantoanhoc.net, Diendan.hocmai.vn, K2pi.net.vn, Diendantoanthcs.forumvi.net...

- *Ưu điểm:*

- Giúp người dùng tìm được thông tin giải quyết vấn đề họ gặp phải nhờ sự trợ giúp của cộng đồng(bao gồm người người có chuyên môn hoặc không có chuyên môn).

- *Nhược điểm:*

- Tìm kiếm nội dung khó khăn.
- Không có cơ chế để phân biệt được ai là chuyên gia.
- Không cấp cách thức trao đổi trực tiếp tức thời với chuyên gia.
- Giao diện khó sử dụng và không phù hợp với việc tìm kiếm thông tin.

Có thể thấy rằng các công cụ này đang thực sự tạo ra những khó khăn cho người cần tìm kiếm giải đáp cho vấn đề của mình. Thực tế cũng cho thấy khi con người sẽ chọn cách đơn giản để giải quyết vấn đề họ gặp phải và thường là bỏ qua. Rõ ràng công nghệ phát triển rất nhanh nhưng lại chưa được áp dụng triệt để vào giải quyết những vấn đề bức thiết của nhu cầu học tập, những công cụ hiện tại chưa thực sự đủ tốt, chưa thúc đẩy được nhu cầu học tập, tiếp thu kiến thức thậm chí còn là rào cản cho việc học tập, tiếp thu kiến thức con người

Nhu cầu cấp thiết là cần xây dựng hệ thống thông tin hỏi đáp và tìm kiếm chuyên gia có giao diện tương tác thông qua website, đảm bảo các yêu cầu sau:

- Yêu cầu phi chức năng:

- Tính sẵn sàng cao
- Tính bảo mật cao
 - Bảo mật thông tin cá nhân
 - Bảo mật thông tin thanh toán
 - Bảo mật dữ liệu các phiên hỏi đáp thu phí

- Yêu cầu chức năng:

- Chức năng hiển thị các câu hỏi mới
- Chức năng tìm kiếm câu hỏi bằng từ khóa
- Chức năng tìm kiếm câu hỏi bằng bộ lọc

- Chức năng đặt câu hỏi miễn phí
- Chức năng trả lời câu hỏi
- Chức năng đánh giá câu hỏi
- Chức năng đánh giá câu trả lời
- Chức năng tìm kiếm chuyên gia bằng từ khóa
- Chức năng tìm kiếm chuyên gia bằng bộ lọc
- Chức năng hỏi đáp(chat) tính phí với chuyên gia
- Chức năng đánh giá chất lượng phiên hỏi đáp với chuyên gia
- Chức năng thông báo tin tức đến người dùng

➤ **Mục tiêu của đề tài:**

Xây dựng hệ thống thông tin hỏi đáp và tìm kiếm chuyên gia toán phổ thông theo mô hình chia sẻ về kiến thức, phục vụ cho nhu cầu học tập, chia sẻ kiến thức toán học phổ thông cho học sinh ở các bậc học phổ thông và các đối tượng khác.

➤ **Phương pháp nghiên cứu:**

– **Về lý thuyết:**

- Thu thập dữ liệu xây dựng bộ dữ liệu danh mục toán phổ thông, xây dựng bộ dữ liệu câu hỏi toán phổ thông, bộ dữ liệu về chuyên gia toán.
- Nghiên cứu các thuật toán tìm kiếm văn bản trong ứng dụng thực tế và việc ứng dụng search engine như Elasticsearch trong việc tìm kiếm văn bản tiếng Việt.
- Nghiên cứu các thuật toán xếp hạng trong một số hệ thống hỏi đáp hiện nay như: Quora, StackExchange , Reddit.
- Nghiên cứu phương thức hoạt động và phương pháp triển khai hệ thống truyền thông Instant messaging và Video Call trong thực tế.
- Nghiên cứu phương pháp triển khai hệ thống trên nền điện toán đám mây và các giải pháp bảo mật hệ thống trên nền điện toán đám mây.

– **Về thực nghiệm:**

- Xây dựng bộ dữ liệu danh mục toán phổ thông gồm 12 danh mục tương ứng với 12 chương trình học của toán phổ thông.
- Xây dựng bộ dữ liệu 5000 câu hỏi toán học.
- Xây dựng bộ dữ liệu 2000 chuyên gia toán.
- Xây dựng thuật toán tìm kiếm chuyên gia với đầu vào là từ khóa hoặc cụm từ khóa của một chương trình học.
- Xây dựng thuật toán tìm kiếm câu hỏi với đầu vào là từ khóa hoặc cụm từ khóa.
- Xây dựng thuật toán xếp hạng được tùy biến phục vụ cho chức năng hiển thị danh sách chuyên gia.

- Xây dựng thuật toán xếp hạng được tùy biến phục vụ cho chức năng hiển thị danh sách câu hỏi.
- Xây dựng hệ thống truyền thông Instant messaging và Video Call cho tính năng gọi thoại.
- Xây dựng hệ thống đảm bảo các đầy đủ các yếu tố về bảo mật trong môi trường điện toán đám mây hiện nay.

➤ **Nội dung nghiên cứu:**

- Thu thập dữ liệu về chương trình toán phổ thông để xây dựng bộ dữ liệu danh mục – từ khóa.
- Thu thập dữ liệu câu hỏi toán phổ thông để xây dựng bộ dữ liệu 5000 câu hỏi toán phổ thông.
- Giả lập dữ liệu về người dùng chuyên gia để xây dựng bộ dữ liệu 2000 chuyên gia toán phổ thông.
- Thu thập và nghiên cứu các thuật toán xếp hạng của một số hệ thống có liên quan như: Stack Exchange , Quora, Reddit..., để xây dựng thuật toán xếp hạng riêng.
- Thiết kế các truy vấn, triển khai thử với các bộ dữ liệu 5000 câu hỏi toán, 2000 chuyên gia, bộ dữ liệu danh mục – từ khóa toán phổ thông.
- Lựa chọn giải pháp tách từ tiếng Việt, triển khai với các truy vấn tìm kiếm.
- Xây dựng hệ thống truyền thông Instant messaging và Video Call cho tính năng gọi thoại.
- Nghiên cứu và triển khai các giải pháp đảm bảo tính sẵn sàng của hệ thống khi có số lượng người dùng lớn.
- Nghiên cứu các giải pháp triển khai hệ thống trên môi trường điện toán đám mây, nghiên cứu các giải pháp bảo mật và triển khai vào hệ thống.
- Phân tích và thiết kế hệ thống.
- Mã hóa, cài đặt và chạy thử nghiệm hệ thống.
- Cài đặt, triển khai và đo đạc các thông số của hệ thống.

Nội dung luận văn gồm một số vấn đề chính sau:

ĐẶT VẤN ĐỀ

Chương 1: Khảo sát hệ thống

- 1.1. Mục đích và phạm vi của đề tài
- 1.2. Nghiên cứu các hệ thống có liên quan
- 1.3. Mô tả hệ thống mới
 - 1.3.1. Các đối tượng có liên quan
 - 1.3.2. Quy trình xử lý & quy trình nghiệp vụ
 - 1.3.3. Quy tắc nghiệp vụ
- 1.4. Mô tả các chức năng nghiệp vụ của hệ thống
- 1.5. Đặc tả yêu cầu
 - 1.5.1. Các yêu cầu hệ thống
 - 1.5.2. Xác định các chức năng của hệ thống
 - 1.5.3. Lựa chọn giải pháp

Chương 2: Tổng quan dữ liệu và công nghệ

- 2.1. Xây dựng biểu đồ ca sử dụng (Use case)
 - 2.1.1. Xác định các tác nhân và ca sử dụng
 - 2.1.2. Biểu đồ ca sử dụng
 - 2.1.3. Mô tả kịch bản cho các ca sử dụng
- 2.2. Biểu đồ hoạt động(Activity Diagram)

Chương 3: Phân tích hệ thống

- 3.1. Xây dựng biểu đồ ca sử dụng (Use case)
 - 3.1.1. Xác định các tác nhân và ca sử dụng
 - 3.1.2. Biểu đồ ca sử dụng
 - 3.1.3. Mô tả kịch bản cho các ca sử dụng
- 3.2. Biểu đồ hoạt động(Activity Diagram)
- 3.3. Mô hình liên kết thực thể

Chương 4: Thiết kế hệ thống

- 4.1. Thiết kế kiến trúc hệ thống
- 4.2. Thiết kế kiểm soát
- 4.3. Thiết kế cơ sở dữ liệu

Chương 5: Xây dựng hệ thống

- 5.1. Cài đặt môi trường
- 5.2. Triển khai hệ thống
- 5.3. Kết quả triển khai hệ thống

Kết luận

- 6.1. Kết quả đạt được
- 6.2. Hướng phát triển

5. Tài liệu tham khảo:

TIẾNG ANH

- [1] Viraj Anchan , Sarang Deshpande , Deep Doshi³, Akshat Kedia - *Ranking Algorithm*, IJARCCCE, 2015.
- [2] James Rumbaugh, Ivar Jacobson, Grady Booch - *The Unified Modeling Language Reference Manual, Second Edition*, 2004.
- [3] PKE Consulting - *Introduction to WebRTC – 2015*, 2015.
- [4] Salvatore Loreto, Simon Pietro Romano - *Real-Time Communication with WebRTC*, 2014.
- [5] Nick Craver - *Stack Overflow: The Architecture - 2016 Edition*, 2016.
- [6] Phan Quy Trung - *Optimizing the Search Experience on E-commerce Websites*, 2017.
- [7] Arnaud Budkiewicz - *Facebook Silently Released WebRTC Video Calling*, 2015.
- [8] Clinton Gormley, Zachary Tong - *Elasticsearch: The Definitive Guide*, 2015.
- [9] Phil Whelan - *Quora’s Technology Examined*, 2011.
- [10] Le Hong Phuong - *A hybrid approach to word segmentation of Vietnamese texts*, 2008.
- [11] Makble - *The advantages and disadvantages of MySQL*, 2016.
- [12] Paul Shan - *Node.js – reasons to use, pros and cons, best practices!*, 2014.
- [13] Tomislav Capan - *Why The Hell Would I Use Node.js? A Case-by-Case Tutorial*, 2013.
- [14] Ynori Seven - *The Pros and Cons of MongoDB*, 2014.
- [15] Cody Arsenaault - *The Pros and Cons of 8 Popular Databases*, 2017.

6. Dự kiến kế hoạch thực hiện:

TT	Nội dung công việc	Thời gian thực hiện	Địa điểm thực hiện	Ghi chú
1	Thu thập dữ liệu về chương trình toán phổ thông để xây dựng bộ dữ liệu danh mục – từ khóa	1/9 – 20/9/2017		
2	Giả lập dữ liệu về người dùng chuyên gia để xây dựng bộ dữ liệu 2000 chuyên gia toán phổ thông.	21/9 – 25/9/2017		
3	Thu thập dữ liệu câu hỏi toán phổ thông để xây dựng bộ dữ liệu 5000 câu hỏi toán phổ thông	26/9– 15/10/2017		

4	Thu thập và nghiên cứu các thuật toán xếp hạng của một số hệ thống có liên quan như: Stack Exchange , Quora, Reddit..., để xây dựng thuật toán xếp hạng riêng.	16/10 – 5/11/2017		
5	Thiết kế các truy vấn, triển khai thử nghiệm với các bộ dữ liệu 5000 câu hỏi toán, 2000 chuyên gia, bộ dữ liệu danh mục – từ khóa toán phổ thông.	6/11 – 22/11/2017		
6	Lựa chọn giải pháp tách từ tiếng Việt, triển khai với các truy vấn tìm kiếm	23/11 – 30/11/2017		
7	Nghiên cứu các giải pháp triển khai hệ thống trên môi trường điện toán đám mây, nghiên cứu các giải pháp bảo mật và triển khai vào hệ thống.	8/12 – 24/12/2017		
8	Phân tích và thiết kế hệ thống	25/12/2017 – 28/2/2018		
9	Mã hóa, cài đặt và chạy thử nghiệm hệ thống	1/3 – 14/3/2018		
10	Cài đặt, triển khai và đo đạc các thông số của hệ thống	15/3 – 30/3/2018		

7. Các cơ quan, đơn vị cần liên hệ:

- Khoa CNTT – HVKTQS

8. Kinh phí thực hiện đề tài, điều kiện đảm bảo:

- Tự thực hiện.

Ngày 26 tháng 9 năm 2017

CHỦ NHIỆM BỘ MÔN

NGƯỜI LẬP ĐỀ CƯƠNG

Dương Văn Bách

CHỦ NHIỆM KHOA

GIÁO VIÊN HƯỚNG DẪN

