



UTEHY RESEARCH: HỆ THỐNG CHIA SẺ THÔNG TIN PHỤC VỤ NGHIÊN CỨU KHOA HỌC

Hồ Khánh Lâm, Nguyễn Minh Quý, Nguyễn Văn Quyết
Trường Đại học Sư phạm Kỹ thuật Hưng Yên

Ngày nhận: 25/1/2016

Ngày xét duyệt: 03/3/2016

Tóm tắt:

Trong những năm gần đây, phong trào nghiên cứu khoa học (NCKH) của Trường Đại học Sư phạm Kỹ thuật (ĐHSPKT) Hưng Yên ngày càng đẩy mạnh. Mỗi năm có hàng trăm bài báo của cán bộ Nhà trường đăng tải trên các hội thảo, hội nghị, tạp chí khoa học trong và ngoài nước. Đây là nguồn tài liệu tham khảo rất quan trọng trong NCKH. Tuy nhiên, các tài liệu này chỉ lưu trữ trên máy tính cá nhân, khó khăn trong việc khai thác thông tin. Trong bài báo này chúng tôi đề xuất hệ thống UTEHY Research hỗ trợ: (1) quản lý các bài báo khoa học; (2) chia sẻ và khai thác tài liệu NCKH; (3) thống kê các số liệu dựa trên thông tin các bài báo; (4) cung cấp tin tức, sự kiện về NCKH trong và ngoài trường. Kết quả thực nghiệm tại <http://research.utehy.edu.vn> đã chỉ ra hệ thống của chúng tôi có thể hỗ trợ hiệu quả cho công tác nghiên cứu khoa học.

Từ khoá: NCKH, UTEHY Research.

1. Đặt vấn đề

Hiện nay trên thế giới đã có một số tổ chức lớn cung cấp các hệ thống chia sẻ thông tin phục vụ NCKH như: Google Research [1], Microsoft Research[2], DBLB [3], v.v... Các thông tin được đăng tải trên hệ thống cho phép dễ dàng tìm kiếm, tuy nhiên các nguồn tài nguyên hầu hết đều tính phí. Ở các trường Đại học, Cao đẳng Việt Nam, việc triển khai hệ thống chia sẻ thông tin dành riêng cho NCKH hầu hết chỉ mang tính chất nội bộ của đơn vị, cần sử dụng tài khoản được cấp mới có thể sử dụng tài nguyên. Trong bài báo này, chúng tôi đề cập đến việc xây dựng một hệ thống chia sẻ thông tin phục vụ nghiên cứu khoa học tại Trường ĐHSPKT Hưng Yên.

Trong những năm gần đây, phong trào nghiên cứu khoa học và công nghệ của Nhà trường ngày càng đẩy mạnh. Trường ĐHSPKT Hưng Yên cần xây dựng một hệ thống chia sẻ các thông tin phục vụ công tác NCKH nhằm thúc đẩy hoạt động này ngày một phát triển hơn. Tính cấp thiết của hệ thống đó được thể hiện qua các khía cạnh sau:

Thứ nhất, hàng năm có hàng trăm bài báo của cán bộ nhà trường được đăng tải trên các hội thảo, hội nghị, tạp chí khoa học trong và ngoài nước. Theo đó, nhu cầu quản lý thông tin các bài báo của cán bộ Phòng Quản lý Khoa học Công nghệ và Hợp tác Quốc tế ngày càng trở nên cần thiết để có được các số liệu thống kê theo Khoa, theo cán bộ.

Thứ hai, trong nghiên cứu khoa học, nguồn tài liệu tham khảo là hết sức quan trọng. Tuy nhiên, trên thực tế, các bài báo khoa học không phải là nguồn tài liệu luôn miễn phí và dễ tìm kiếm. Mỗi đơn vị nghiên cứu có thể có những thư viện riêng

phục vụ cho cán bộ của mình. Hiện tại, số lượng bài báo của cán bộ giảng viên nhà trường nhiều, việc xây dựng một thư viện tài nguyên trực tuyến về NCKH càng trở nên cần thiết để cán bộ giảng viên có thể tra cứu, tham khảo các bài báo đã có trước đó.

Thứ ba, nhiều cán bộ giảng viên đã có kinh nghiệm nghiên cứu, có nhiều tài liệu tham khảo quý giá đã và đang muốn chia sẻ các nguồn tài nguyên đang có cho cán bộ giảng viên khác. Nguồn tài nguyên này có thể lên tới hàng triệu tài liệu, bài báo khoa học, nếu không tổ chức lưu trữ tốt, không có hệ thống tìm kiếm đủ mạnh thì khó có thể khai thác được. Chính vì vậy, một hệ thống chia sẻ, hỗ trợ người dùng đăng tải, tra cứu và khai thác tài liệu tham khảo trong NCKH là rất cần thiết.

Ngoài ra, hàng năm theo định kỳ nhà trường cũng có những thông báo về tổ chức các sự kiện liên quan đến NCKH như việc đăng nộp bài báo cho tạp chí của Nhà trường, các đơn vị tổ chức trong và ngoài nước. Như vậy, cũng cần có kênh thông tin chia sẻ các thông tin này cho cán bộ giảng viên của nhà trường. Kênh thông tin này cần được tích hợp trong một hệ thống mà nhiều cán bộ quan tâm đến NCKH sử dụng.

Chúng tôi đã đề xuất xây dựng một hệ thống website trên nền tảng công nghệ .NET, sử dụng các công nghệ mới của Microsoft như: ASP.NET 4.5, LINQ. Hệ thống được cài đặt hỗ trợ tìm kiếm Full-Text trong cơ sở dữ liệu SQL Server – một giải pháp hỗ trợ tìm kiếm mạnh dựa trên thuật toán tìm kiếm mờ (Fuzzy Search). Hệ thống được xây dựng trên cơ sở các thông tin thực tế của các bài báo khoa học, cung cấp nhiều tính năng sau hỗ trợ: cán bộ Phòng

Quản lý Khoa học Công nghệ và Hợp tác Quốc tế quản lý tốt các bài báo khoa học của cán bộ giảng viên trong toàn trường; giúp các cán bộ giảng viên có thể chia sẻ và khai thác tài liệu nghiên cứu khoa học theo các lĩnh vực; thống kê báo cáo các số liệu về bài báo khoa học theo nhiều tiêu chí; cung cấp các thông tin, sự kiện liên quan đến NCKH trong Nhà trường và các đơn vị ngoài trường.

Phần còn lại của bài báo được tổ chức như sau: chúng tôi sẽ giới thiệu một số kiến thức nền tảng liên quan đến vấn đề xây dựng đề tài trong phần 2. Trong phần 3 chúng tôi trình bày nội dung giải pháp xây dựng hệ thống chia sẻ thông tin phục vụ NCKH. Phần 4 là các kết quả đạt được và chúng tôi đưa ra một vài bản luận. Cuối cùng là chúng tôi đưa ra kết luận trong phần 5.

2. Kiến thức nền tảng

2.1. Cấu trúc của một bài báo khoa học

Về nguyên tắc, một bài báo khoa học cần đảm bảo bố cục IMRAD (viết tắt lần lượt của: Introduction, Methods, Results, And, Discussion). Bố cục này được giới khoa học trong nước và quốc tế chấp nhận rộng rãi vì nó phù hợp với dạng thức đơn giản nhất và lô-gic nhất của việc công bố kết quả nghiên cứu khoa học. Tùy vào lĩnh vực nghiên cứu, vấn đề nghiên cứu và phong cách tác giả, đôi khi bố cục trên có thể được đơn giản hóa đi, cụ thể hóa thêm, hoặc có những thay đổi về trật tự, hoặc được thể hiện một cách ngầm ẩn nhưng về cơ bản, một bài báo khoa học cần thể hiện được các yếu tố trên [10]. Theo đó, các thuộc tính cơ bản của bài báo bao gồm: tiêu đề (title), tác giả (authors), tóm tắt (abstract), từ khóa (keyword), nội dung nghiên cứu (method), tài liệu tham khảo (references).

2.2. Kỹ thuật tìm kiếm toàn văn (Full-Text Search)

Full-Text Search là tìm kiếm theo chỉ mục (Index), để sử dụng tìm kiếm Full-Text ta cần cài đặt SQL 2008 hoặc phiên bản mới hơn. Ưu điểm của tìm kiếm Full-Text là kết quả tìm kiếm chính xác hơn so với tìm kiếm “like” thông thường, đặc biệt thời gian tìm kiếm tương đối nhanh, phù hợp với những website có hệ thống cơ sở dữ liệu lớn. Ta có thể sử dụng các kiểu tìm kiếm như: FREETEXT, CONTAINS, FREETABLE, CONTAINSTABLE. Hệ thống của chúng tôi được xây dựng sử dụng hai kiểu tìm kiếm: *FREETEXT* và *CONTAINS* trong Full-Text Search. Áp dụng trong việc tìm kiếm tiêu đề bài báo, tên tác giả.

FREETEXT: Đây là kiểu tìm kiếm theo kiểu fuzzy (tìm kiếm mờ). Full-Text có ưu điểm không phân biệt tiếng Việt có dấu hay không dấu. Ví dụ khi ta tìm kiếm từ khóa “bai bao” thì “bài báo” cũng

được tìm thấy.

CONTAINS: Đây là kiểu tìm kiếm sử dụng các phép toán logic “and” hoặc “or”. Khi sử dụng kiểu “and”, ví dụ ta tìm với từ khóa “bài báo” thì tất cả dữ liệu có nội dung là “bài” và “báo” (có thể rời rạc) sẽ được tìm thấy. Khi ta tìm kiếm theo kiểu “or” tất cả dữ liệu có “bài” hoặc “báo” kiểu này gần giống với *FREETEXT*.

3. HỆ THỐNG UTEHY RESEARCH

3.1. Xác định các yêu cầu của hệ thống

Trải qua quá trình khảo sát, phân tích chúng tôi xác định được các yêu cầu của hệ thống dựa trên hai phân hệ: khai thác tài nguyên và quản lý tài nguyên.

a) Phân hệ khai thác tài nguyên

Là phân hệ được thiết kế dành cho cán bộ giảng viên sử dụng hệ thống tra cứu bài báo, tác giả, tài liệu và cập nhật các tin tức sự kiện NCKH liên quan.

Chức năng tra cứu thông tin: hệ thống cho phép người dùng tra cứu thông tin một cách nhanh chóng bằng từ khóa dựa vào tiêu đề bài báo hoặc tên tác giả. Tra cứu hỗ trợ gợi ý nhanh, tra cứu gần đúng không phân biệt Tiếng Việt có dấu, không dấu.

Chức năng xem thông tin: hệ thống cho phép người dùng xem các thông tin cơ bản của một bài báo như được trình bày trong mục 2.1; xem danh mục bài báo theo lĩnh vực; xem đánh sách bài báo của một tác giả. Ngoài ra, hệ thống cho phép người dùng xem các tin tức sự kiện liên quan đến NCKH.

Chức năng đăng tải bài báo: hệ thống cho phép người dùng có tài khoản được phép đăng, quản lý bài báo đã xuất bản. Và có thể tải về các bài báo trên hệ thống.

Chức năng tạo thư viện cá nhân: hệ thống cho phép người dùng có tài khoản được tạo thư viện các bài báo đã xem.

Chức năng trình bày nguồn tài nguyên: hệ thống trình bày danh sách các bài tạp chí của Nhà trường, danh sách các bài báo của cán bộ trong trường.

b) Phân hệ quản lý tài nguyên

Phân hệ này được thiết kế các chức năng phục vụ việc quản lý các bài báo khoa học, quản lý người dùng, xét duyệt việc đăng tải các tài liệu của người dùng.

Chức năng quản lý bài báo: hệ thống thiết kế các chức năng cho phép đăng tải bài báo, cập nhật thông tin bài báo.

Chức năng quản lý các danh mục: hệ thống thiết kế chức năng cho phép quản lý các danh mục như: lĩnh vực nghiên cứu, tạp chí, bài báo, tác giả.

Chức năng phê duyệt tài liệu người dùng:

hệ thống thiết kế chức năng cho phép quản lý các bài báo, tài liệu khoa học mà người dùng đăng tải.

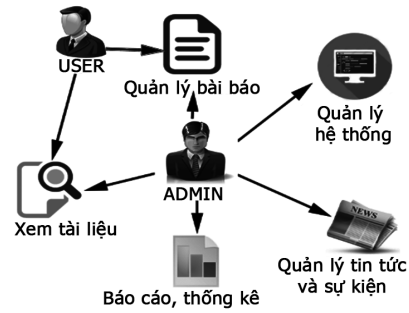
Chức năng quản lý người dùng: hệ thống thiết kế chức năng cho phép tạo mới người dùng và phân quyền sử dụng hệ thống.

Trên đây là những chức năng chính của hệ thống UTEHY Research được chúng tôi đề xuất để xây dựng lên hệ thống chia sẻ thông tin phục vụ NCKH tại trường ĐHSPTK Hưng Yên.

3.2. Thiết kế mô hình hệ thống và dữ liệu

a) Thiết kế mô hình hệ thống

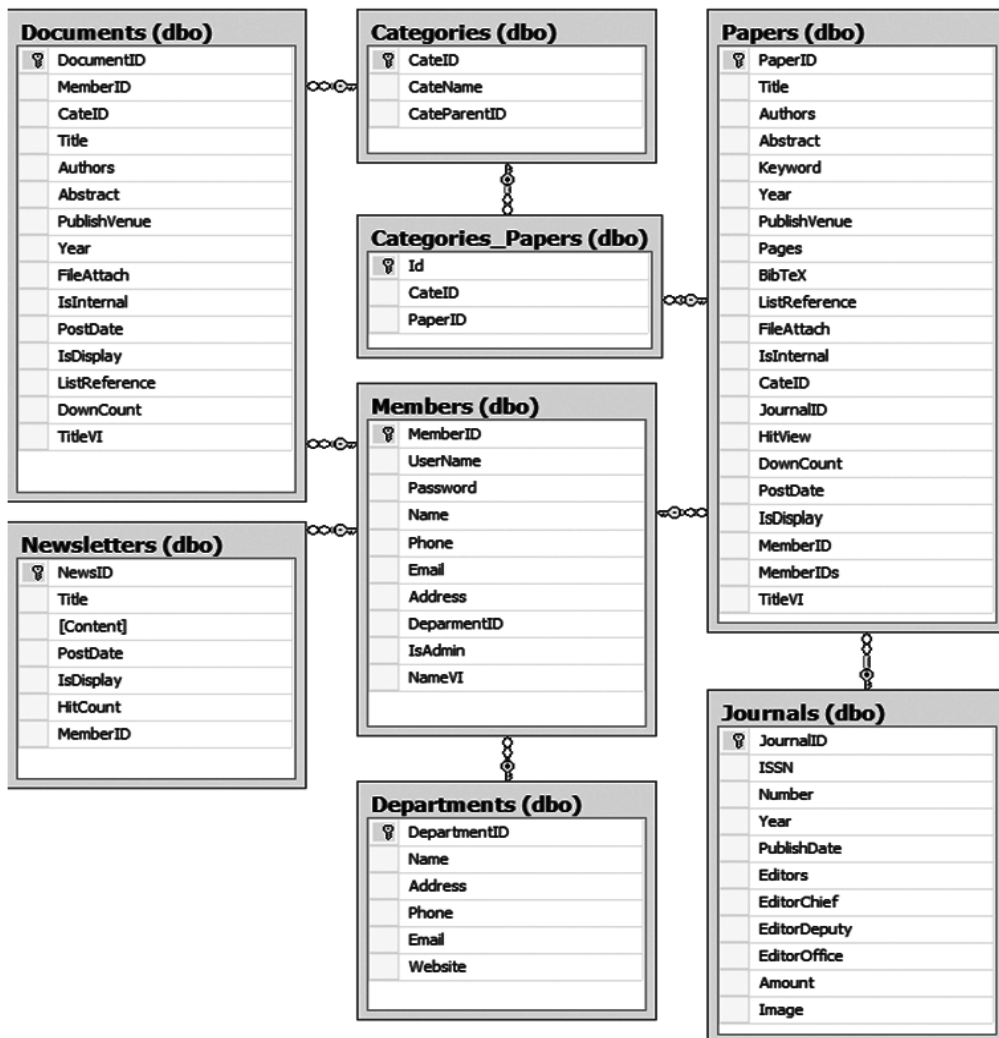
Hình 1 dưới đây trình bày mô hình tổng quan của hệ thống UTEHY Research. Trong đó có hai đối tượng người dùng là bộ phận quản lý (ADMIN) và cán bộ giáo viên (USER). Với cán bộ giảng viên có thể tra cứu: đọc, tải, tạo thư viện bài báo; đối với người quản trị: quản lý toàn bộ thông tin của hệ thống như: đăng bài báo, quản lý tin tức, thống kê báo cáo và một số chức năng khác.



Hình 1. Mô hình hệ thống UTEHY Research

b) Thiết kế mô hình dữ liệu

Trên cơ sở phân tích kỹ các thông tin liên quan đến bài báo khoa học, chúng tôi đã thiết kế cơ sở dữ liệu của hệ thống trên hệ quản trị SQL Server để lưu trữ được thông tin của: thông tin tạp chí, lĩnh vực, bài báo, tác giả. Ngoài ra còn lưu trữ các thông tin về tin tức, sự kiện, tài liệu tham khảo. Một mô hình dữ liệu quan hệ được trình bày trong Hình 2.



Hình 2. Mô hình dữ liệu của hệ thống UTEHY Research

4. Kết quả nghiên cứu và bàn luận

4.1. Các kết quả đạt được

Chúng tôi đã xây dựng được một hệ thống chia sẻ thông tin phục vụ nghiên cứu khoa học cho trường ĐHSPTK Hưng Yên. Hệ thống được triển khai trên nền tảng công nghệ .NET, áp dụng các công nghệ mới của Microsoft như: ASP.NET 4.5, LINQ. Một điểm mạnh của hệ thống là hỗ trợ tra cứu, tìm kiếm theo nhiều tiêu chí để khai thác được thông tin các bài báo khoa học. Ở đó kỹ thuật tìm kiếm Full-Text trong cơ sở dữ liệu SQL Server – một giải pháp hỗ trợ tìm kiếm mạnh dựa trên thuật toán tìm kiếm mờ (Fuzzy Search) được sử dụng.

Hệ thống đã áp dụng thử nghiệm trên domain: research.utehy.edu.vn với tập dữ liệu là các bài báo khoa học Số 1, Số 2 của Tạp chí Khoa học & Công nghệ Trường ĐHSPTK Hưng Yên và các bài báo của cán bộ giảng viên Khoa Công nghệ Thông tin của Nhà trường. Một hệ thống với hai phân hệ: khai thác tài nguyên và quản lý tài nguyên với các tính năng chính hỗ trợ quản lý tốt các bài báo khoa học của cán bộ giảng viên trong toàn trường; giúp các cán bộ giảng viên có thể chia sẻ và khai thác tài liệu nghiên cứu khoa học theo các lĩnh vực; thống kê báo cáo các số liệu về bài báo khoa học theo nhiều tiêu chí; cung cấp các thông tin, sự kiện liên quan đến NCKH trong Nhà trường và các đơn vị ngoài trường.

4.2. Bàn luận

Như chúng tôi đã trình bày trong phần 1 (Đặt vấn đề), việc xây dựng các hệ thống chia sẻ thông tin về NCKH đã được nhiều tổ chức lớn thực hiện, cung cấp cho các nhà nghiên cứu một kho học liệu không lồ như: Google Scholar, Microsoft Research, ... Tuy nhiên, nhiều bài báo của các hệ thống trên cần tài khoản tính phí. Trong khi đó, số lượng bài báo được đăng của cán bộ mỗi trường đại học như

ĐHSPTK Hưng Yên lên đến hàng trăm bài, đóng góp một khối lượng không nhỏ các tài nguyên khoa học. Một hệ thống như UTEHY Research cho việc lưu trữ và khai thác các tài nguyên cần được triển khai nhân rộng trong các trường đại học, cao đẳng trong toàn quốc. Kết quả đạt được của hệ thống UTEHY Research là bước khởi đầu cho việc quản lý và khai thác tài nguyên phục vụ NCKH.

Trong tương lai, hệ thống chia sẻ thông tin phục vụ NCKH sẽ mở rộng với chức năng gợi ý bài báo và gợi ý cộng tác nghiên cứu. Thứ nhất, chúng tôi sẽ tích hợp hệ thống gợi ý các bài báo liên quan cho người dùng khi vào xem một bài báo. Kết quả gợi ý đưa ra được lấy từ cơ sở dữ liệu của hệ thống dựa trên một giải thuật trích chọn thực thể liên quan dựa vào tiêu đề, từ khóa. Thứ hai, hệ thống hỗ trợ gợi ý cộng tác nghiên cứu: trên cơ sở dữ liệu các bài báo của hệ thống đã có, áp dụng một giải thuật để đánh giá mức độ tương đồng về hướng nghiên cứu, từ đó đưa ra gợi ý cộng tác giữa các nhà nghiên cứu. Đó là hướng mở rộng rất thiết thực và tạo lên sức mạnh của hệ thống phục vụ NCKH.

5. Kết luận

Trong bài báo này chúng tôi đã trình bày việc đề xuất một hệ thống chia sẻ thông tin phục vụ nghiên cứu khoa học. Đây là một hệ thống với nhiều tính năng hữu ích, hỗ trợ tốt cho cán bộ giáo viên nói riêng và các nhà nghiên cứu nói chung trong việc khai thác các bài báo khoa học. Với các kết quả thử nghiệm ban đầu cho thấy hệ thống đã chạy ổn định trên mạng Internet.

Trên cơ sở nghiên cứu đã đạt được, trong thời gian tới, chúng tôi sẽ phát triển hệ thống UTEHY Research mạnh hơn bằng việc tích hợp hệ thống gợi ý bài báo liên quan và gợi ý cộng tác nghiên cứu như đã bàn luận ở trên.

Tài liệu tham khảo

- [1]. Google. (2014, May), *Google Scholar*, [Online]. <http://scholar.google.com.vn/>
- [2]. Microsoft. (2014, May), *Microsoft Research*, [Online]. <http://research.microsoft.com/>
- [3]. DBLP team. (2014, May), *DBLP Computer Science Bibliography*, [Online]. <http://www.informatik.uni-trier.de/~ley/db/>
- [4]. Jie Tang, Yutao Zhang, and Huaiyu Wan. (2014, May), *Aminer*, [Online]. <http://arnetminer.org/citation>
- [5]. Phạm Hữu Khang, *Lập trình ASP.NET 2.0*, NXB Lao Động Xã Hội, Hà Nội, Việt Nam, 2005.
- [6]. Nguyễn Văn Quyết, *Công nghệ Web và ứng dụng*, Khoa Công nghệ Thông tin - Trường Đại học Sư phạm Kỹ thuật Hưng Yên, Hưng Yên, Việt Nam, 2010.
- [7]. Paolo Pialorsi and Marco Russo, *Programming Microsoft LINQ in Microsoft.NET Framework 4*, California, America: Adam Zaremba, 2010.
- [8]. Matthew MacDonald and Mario Szpuszta, *Pro ASP.NET 3.5 in C# 2008*, New York, America: Katie Stence, 2007.

[9]. Microsoft, (2014, May), *Microsoft MSDN*, [Online]. <http://msdn.microsoft.com/en-us/library/ms142571.aspx>

[10]. Ban biên tập, (2014, May), Đại Học Quốc Gia Hà Nội, [Online]. <http://ulis.vnu.edu.vn/taxonomy/term/28/2766>.

UTEHY RESEARCH: INFORMATION SHARING SYSTEM FOR SCIENCE RESEARCHES

Abstract:

In recent years, the movement of scientific research in Hung Yen University of Technology and Education increasingly promoted. Hundreds of research papers a year were published in the workshops, conferences, journals. These are important resources in scientific research. However, these materials are only stored on personal computers, so it is difficult for information exploiting. In this paper, we propose UTEHY Research System: (1) management of scientific articles; (2) sharing and exploiting document for research; (3) statistics of the data based on information of the papers; (4) providing news and events related to scientific research both within and outside school. The experimental results in the address <http://research.utehy.edu.vn> demonstrate that our system can efficiently support for scientific research.

Keywords: Publications software management, Articles software management.