

TRƯỜNG ĐẠI HỌC VĂN HOÁ HÀ NỘI
KHOA THƯ VIỆN - THÔNG TIN



TÌM KIẾM THÔNG TIN VĂN BẢN TRONG THƯ VIỆN SỐ

KHOÁ LUẬN TỐT NGHIỆP

<i>Giảng viên hướng dẫn</i>	: TS. ĐỖ QUANG VINH
<i>Sinh viên thực hiện</i>	: NGUYỄN THỊ THANH HẢI
<i>Lớp</i>	: THƯ VIỆN 41B

Hà Nội - 2013

MỤC LỤC

MỞ ĐẦU	1
Chương 1: KHÁI QUÁT VỀ THƯ VIỆN SỐ VÀ THÔNG TIN VĂN BẢN TRONG THƯ VIỆN SỐ	10
1.1 Khái niệm, đặc trưng và kiến trúc thư viện số	10
1.1.1 Khái niệm thư viện số	10
1.1.2 Đặc trưng thư viện số	13
1.1.3 Kiến trúc thư viện số	16
1.2 Thông tin văn bản trong thư viện số	22
Chương 2: PHƯƠNG PHÁP TÌM KIẾM THÔNG TIN VĂN BẢN TRONG THƯ VIỆN SỐ	24
2.1 Mô hình tìm kiếm thông tin - IR	24
2.2 Truy vấn Boolean – BQ	28
2.2.1 Truy vấn BQ hội	29
2.2.2 Truy vấn BQ không hội	31
2.3 Truy vấn xếp hạng - RQ	34
2.3.1 So khớp tọa độ	36
2.3.2 Tích trong độ tương tự	36
2.3.3 Mô hình không gian vector	38
Chương 3: THỬ NGHIỆM TÌM KIẾM VÀ ĐÁNH GIÁ HIỆU SUẤT TÌM KIẾM THÔNG TIN VĂN BẢN TRONG THƯ VIỆN SỐ	45
3.1 Thử nghiệm tìm kiếm thông tin văn bản tại một số thư viện số trên thế giới và Việt Nam	45
3.2 Đánh giá hiệu suất tìm kiếm thông tin	52
3.3. Đánh giá phương pháp tìm kiếm thông tin văn bản trong thư viện số ...	55
3.3.1 Ưu điểm	56
3.3.2 Nhược điểm	58
KẾT LUẬN	61
TÀI LIỆU THAM KHẢO	62
PHỤ LỤC	

LỜI CẢM ƠN

Tôi xin bày tỏ lòng kính trọng và biết ơn sâu sắc tới thầy Đỗ Quang Vinh, người đã tận tình hướng dẫn, giúp đỡ, động viên tôi trong suốt thời gian thực hiện khóa luận này.

Tôi cũng xin gửi lời cảm ơn chân thành tới các thầy cô trong Khoa Thư viện – thông tin Trường Đại học Văn hóa Hà Nội đã truyền đạt cho tôi những kiến thức nền tảng về ngành thư viện và tạo điều kiện thuận lợi để tôi hoàn thành khóa luận tốt nghiệp này.

Mặc dù đã có nhiều cố gắng cùng với sự hướng dẫn tận tâm của thầy giáo hướng dẫn, song do hạn chế về thời gian và trình độ của bản thân, vì vậy khóa luận khó tránh khỏi những sai sót. Rất mong nhận được những ý kiến đóng góp chân thành của thầy cô và các bạn để khóa luận tốt nghiệp của tôi được hoàn thiện hơn.

Xin chân thành cảm ơn!

Hà Nội, ngày 03 tháng 06 năm 2013

Sinh viên

Nguyễn Thị Thanh Hải

MỞ ĐẦU

1. Lý do chọn đề tài

Trong kỷ nguyên xã hội thông tin, nền kinh tế tri thức đang phát triển một cách mạnh mẽ. Mạng Internet phát triển rộng khắp toàn cầu khiến cho thông tin gia tăng theo hàm mũ, với sự phong phú, đa dạng và phức tạp của các loại hình thông tin như: Văn bản, hình ảnh, âm thanh, video... Khi đó, thư viện số ra đời nhằm quản trị thông tin và tri thức của nhân loại. Thư viện số lưu trữ một lượng thông tin lớn với hệ thống các cơ sở dữ liệu đa phương tiện và tạo khả năng truy cập thông tin mọi lúc mọi nơi trên phạm vi toàn thế giới.

Thông tin văn bản là một dạng dữ liệu đa phương tiện, chiếm đa số trong các cơ quan, tổ chức đặc biệt là trong thư viện nói chung và thư viện số nói riêng. Ngày nay, với khối lượng thông tin văn bản khổng lồ, việc tìm kiếm một cách nhanh chóng, chính xác và phù hợp với nhu cầu người sử dụng lại càng trở nên khó khăn hơn. Vì vậy, tìm kiếm thông tin văn bản có vai trò quan trọng trong mọi lĩnh vực hoạt động của con người. Nó trở thành một nhu cầu thiết yếu không thể thiếu. Với những thông tin hữu ích được lưu trữ, cập nhật trong thư viện số đã thúc đẩy mọi hoạt động như nghiên cứu - giáo dục - kinh tế - chính trị - văn hóa... ngày một phát triển hơn.

Hơn nữa, thư viện số là một mô hình thư viện mới và hiện đại để khai thác thông tin một cách có hiệu quả đang là vấn đề cần quan tâm của các nhà nghiên cứu trong hoạt động thư viện, công nghệ thông tin và các ngành khác.

Hầu hết người sử dụng thư viện không có kiến thức cụ thể về các lược đồ, hoặc cấu trúc truy vấn ngôn ngữ để tiếp cận thông tin được lưu trữ trong các cơ sở dữ liệu của thư viện số. Với mục đích giúp người sử dụng có thể dễ dàng truy cập, tìm kiếm và khai thác thông tin một cách có hiệu quả nhất từ

nguồn tài nguyên được cập nhật và lưu trữ trong thư viện số, tôi đã chọn đề tài *“Tìm kiếm thông tin văn bản trong thư viện số”*.

2. Tình hình nghiên cứu

Liên quan đến thư viện số, đã có nhiều đề tài nghiên cứu dưới các góc độ khác nhau như: “Xây dựng một số công cụ hỗ trợ tra cứu và tổng hợp thông tin trong các thư viện số” – đề tài nghiên cứu khoa học do Đỗ Phúc và Nguyễn Minh Hiệp làm chủ nhiệm; “Digital Library” – luận văn tốt nghiệp của nhóm sinh viên Trường Đại học nông lâm thành phố Hồ Chí Minh... Về vấn đề tìm kiếm thông tin văn bản cũng có đề tài đề cập đến như: “Kỹ thuật tìm kiếm văn bản trên cơ sở nội trong các cơ sở dữ liệu đa phương tiện”... Tuy nhiên, đề tài này chưa đi cụ thể vào việc tìm kiếm thông tin văn bản trong thư viện số. Có thể nói, *“Tìm kiếm thông tin văn bản trong thư viện số”* là đề tài nghiên cứu đầu tiên của sinh viên khoa Thư viện – thông tin Trường Đại học Văn hóa Hà Nội về hai phương pháp tìm kiếm thông tin cơ bản trong thư viện số và có thực hành cụ thể giúp người sử dụng thư viện có thể tìm kiếm thông tin hiệu quả hơn.

3. Đối tượng nghiên cứu

- Phương pháp tìm kiếm thông tin văn bản trong thư viện số

4. Mục đích nghiên cứu

- Đánh giá các phương pháp tìm kiếm thông tin văn bản trong thư viện số

5. Phạm vi nghiên cứu

- Tìm kiếm thông tin văn bản trong thư viện số.

6. Phương pháp nghiên cứu

- Phương pháp tìm kiếm Boolean BQ
- Phương pháp tìm kiếm xếp hạng RQ

TÀI LIỆU THAM KHẢO

TÀI LIỆU TIẾNG VIỆT

1. Nguyễn Minh Hiệp (2004), *“Thế giới thư viện số”*, Bản tin Thư viện – Công nghệ thông tin, Đại học Quốc gia Tp. Hồ Chí Minh, Tp. Hồ Chí Minh.
2. Đinh Trung Hiếu, Vũ Bội Hằng, Nguyễn Cẩm Tú (2004), *Giải pháp tìm kiếm theo lĩnh vực trong máy tìm kiếm*, Báo cáo nghiên cứu khoa học Khoa Công nghệ, Đại học Quốc gia Hà Nội, Hà Nội.
3. Trần Bích Hồng, Cao Minh Kiểm (2009), *Tra cứu thông tin trong hoạt động thư viện – thông tin*, Nxb Đại học Quốc gia Hà Nội, Hà Nội.
4. Kỷ yếu hội thảo quốc tế chủ đề: *Thư viện số*, Đại học Quốc gia TP. Hồ Chí Minh.
5. Lourdes T.D. (2006), *Thư viện số và truy cập mở tài liệu lưu trữ*, Nguyễn Xuân Bình và nnk biên dịch, UNESCO, Hà Nội.
6. Đoàn Sơn (2002), *Phương pháp biểu diễn văn bản sử dụng tập mờ và ứng dụng trong khai phá dữ liệu văn bản* Luận văn thạc sỹ Khoa Công Nghệ, Đại học Quốc gia Hà Nội, Hà Nội.
7. Vũ Văn Sơn (2005), *“Nhập môn thư viện điện tử”*, <http://www.thuvien.net>.
8. Tạp chí thư viện Việt Nam, số 2/2010
9. Vũ Đức Thi (1997), *Cơ sở dữ liệu - Kiến trúc và thực hành*, Nxb Khoa học và kỹ thuật, Hà Nội.
10. Hà Quang Thụy, Phan Xuân Hiếu, Đoàn Sơn, Nguyễn Trí Thành, Nguyễn Thu Trang, Nguyễn Cẩm Tú (2009). *Giáo trình khai phá dữ liệu Web*. Nxb Giáo dục Việt Nam, Hà Nội.
11. Nguyễn Thị Thu Trang (2010), *Kỹ thuật tìm kiếm văn bản trên cơ sở nội dung trong cơ sở dữ liệu đa phương tiện* Luận văn thạc sỹ, Đại học Công nghệ - Đại học Quốc gia Hà Nội, Hà Nội.
12. Đỗ Trung Tuấn (2004), *Cơ sở dữ liệu*, Nxb Đại học Quốc gia Hà Nội.

13. Đỗ Quang Vinh, Quách Tuấn Ngọc (2001), “*Một mô hình dữ liệu hướng đối tượng thời gian đối với tài liệu cấu trúc*”, Tạp chí Bưu chính viễn thông & Công nghệ thông tin, 160(6), tr.29-32.
14. Đỗ Quang Vinh (2006), “*Truy vấn xếp hạng tài liệu văn bản trong thư viện số*”, Kỷ yếu Hội thảo Quốc gia một số vấn đề chọn lọc của công nghệ thông tin và truyền thông lần thứ IX, Đà Lạt.
15. Đỗ Quang Vinh (2009), *Thư viện số chỉ mục và tìm kiếm*, Nxb Đại học Quốc gia Hà Nội, Hà Nội.

TÀI LIỆU TIẾNG ANH

16. Arm W.Y. (2003), *Digital Library*, MIT Press, Cambridge.
17. Barber D. (1996), “*Buiding a Digital Library*”, Library Technology Report, 32(5), pp. 573-738.
18. Barber D. (1998), “*Tools for Managing the Digital Library*”, Library Technology Report, 34(4), pp. 439-552.
19. Barnes S.J. (2004), *Becoming a Digital Library*, Marcel Dekker, New York.
20. Chen H., Houston A.L. (1999), “*Digital Libraries: social issues and technological advances*”, Advanced in Computers, 48, pp. 257-314.
21. Chowdhary G.G (1999), “*Digital Library Research: major issues and trend*”, Journal of Documenttation, 55(4), pp. 409-448.
22. Kowalski G. (1997), *Information Retrieval System*, Kluwer Academic Publisher, Boston.
23. Reddy R., Wladawsky-Berger I. (2001), *Digital Libraries: Universal Access to Human Knowledge*. Report PITAC.
<http://www.trd.gov/pubs/pitac>
24. Robertson S.E., Beaulieu M. (1997), “*Research and Evaluation in Information Retrieval*”, Journal of Documentation, 53(1), pp.51-57.