

TIẾN TỚI CÁC DỊCH VỤ THƯ VIỆN SỐ THÔNG MINH VÀ TÙY BIẾN¹

Md Maruf Hasan

Trường công nghệ, Đại học Shinawatra, Thái Lan

Ekawit Nantajeewarawat

Viện công nghệ Sirindhorn, Đại học Thammat, Thái Lan

4. Nghiên cứu thăm dò và đánh giá

Trong phần này, chúng tôi giải thích công trình nghiên cứu thăm dò của mình bằng một sưu tập nhỏ tư liệu thư viện số trong lĩnh vực tin học và một nhóm người sử dụng là những sinh viên hoặc giáo viên khoa tin học.

4.1. Sưu tập thư viện số và những người sử dụng tham gia

Một sưu tập thư viện số nhỏ được chuẩn bị với khoảng 300 bài thuộc lĩnh vực tin học. Các bài mục được tổ chức dựa vào bảng phân loại ACM CCS, và việc định chỉ mục toàn văn được xây dựng bằng phần mềm thư viện số Greenstone. Do đó, các giao diện duyệt lướt và tìm kiếm đều sẵn có cho sưu tập như chúng luôn có trong bất kỳ hệ thống thư viện số điển hình nào. Đây là hệ thống cơ bản của chúng tôi và được lưu trong máy chủ cơ bản không hề nâng cấp thêm.

Cũng vẫn sưu tập ấy được sao chép lại cho một hệ thống khác (sau đây gọi

là hệ thống nâng cấp), ở đó chúng tôi cũng thực thi những mô đun bổ sung (Hình 1, lớp 2- mô đun đăng nhập/diện, giá sách, v.v...) và các dịch vụ giá trị gia tăng (Hình 1, lớp 3 - Diện động, sàng lọc hợp tác, các mô đun hướng về sở thích, v.v...). Một nhóm 10 người sử dụng là những sinh viên và giáo viên khoa tin học với những tiêu điểm nghiên cứu của riêng mình trong lĩnh vực tin học được yêu cầu đăng nhập với hệ thống nâng cấp này. Chúng tôi cũng khuyến họ sử dụng định kỳ hệ thống trong thời gian 3 tháng và bổ sung các tài liệu tương thích cho giá sách của họ ngay khi thấy chúng có ích cho diện nghiên cứu và sở thích của họ.

Hết 3 tháng thử nghiệm, chúng tôi chỉ thấy 4 người sử dụng với hơn 9 tài liệu trong giá sách của họ cùng với những dữ liệu về sử dụng khác. Trong công trình nghiên cứu thăm dò này, chúng tôi mời 4 người sử dụng này thực hiện nhiệm vụ thí điểm sau đây.

¹ Phần đầu của bài dịch đã được đăng trong số 3/2010, Tạp chí “Thông tin và Tư liệu”

4.2. Những nhiệm vụ thử nghiệm chủ yếu

Mỗi người sử dụng được giới thiệu với 3 tập hợp bài tóm tắt, mỗi tập hợp có 6 bài (sau khi đã loại bỏ nhan đề, từ khóa, số phân loại và thông tin khác). Mỗi tập hợp 6 bài tóm tắt bao gồm 3 bài lấy từ các tài liệu trên giá của riêng họ, và 3 bài khác được lựa chọn ngẫu nhiên từ các tài liệu trên giá sách của người sử dụng khác. Không có các tài liệu chồng chéo nhau trong 3 tập hợp bài tóm tắt được giới thiệu cho mỗi người sử dụng. Chúng tôi yêu cầu cả 4 người sử dụng thực hiện 3 nhiệm vụ sau đây với mỗi tập hợp 6 bài tóm tắt trên hệ thống cơ bản và trên hệ thống nâng cấp như giải thích bên dưới. Những người sử dụng được thông báo rằng, cả hai máy chủ đều được trang bị phần mềm ghi trên màn hình Camtasia sao cho chúng tôi có thể nắm bắt được sự tương tác của họ để phân tích thử nghiệm sau này. Vì việc ghi trên màn hình chỉ xâm phạm hoạt động của người sử dụng một cách tối thiểu, nên người sử dụng cần thực hiện nhiệm vụ của họ hết sức tự nhiên, không phải bận tâm về các mục tiêu thử nghiệm mà chỉ chú tâm vào việc tìm kiếm thông tin².

Nhiệm vụ duyệt lướt cơ bản: Về nhiệm vụ duyệt lướt cơ bản, người sử

dụng được yêu cầu dùng giao diện duyệt lướt của hệ thống cơ bản để định vị 6 tư liệu mục tiêu chỉ bằng cách duyệt lướt. Chúng tôi sử dụng máy ghi màn hình (screen recorder) để nắm bắt những tương tác của người sử dụng trong khi họ thực hiện nhiệm vụ để phân tích tiếp. Số lượng trung bình những tương tác mà mỗi người sử dụng thực hiện để định vị 3 tư liệu của riêng họ (giả sử là những tư liệu quen thuộc) và 3 tư liệu khác được đếm và liệt kê trong Bảng 1. Cần lưu ý rằng, bảng phân loại ACM CCS gồm 4 cấp (với một số tham chiếu chéo); và toàn bộ sưu tập thư viện số chỉ gồm khoảng 300 tài liệu. Trong phần lớn trường hợp, người sử dụng tìm được những tư liệu quen thuộc trong 4 tương tác hoặc ít hơn. Tuy nhiên, đối với các tư liệu xa lạ, số tương tác đã cao hơn nhiều vì chúng tôi coi việc dò lại là một tương tác bổ sung.

Nhiệm vụ tìm kiếm cơ bản: Về nhiệm vụ tìm kiếm cơ bản, người sử dụng được nhận một tập hợp 6 bài tóm tắt khác (3 bài tóm tắt mới từ giá sách riêng của họ và 3 bài tóm tắt mới nữa được lấy từ những giá sách của người khác). Lần này, chúng tôi yêu cầu người sử dụng dùng phép tìm kiếm theo từ khóa trên hệ thống cơ bản để định vị những tư liệu ấy qua

² Một phiên làm việc sôi nổi được tiến hành với 1 nhiệm vụ duyệt lướt và 1 nhiệm vụ tìm kiếm, dùng một tập hợp 2 bài tóm tắt cho tất cả 4 người sử dụng.

cuộc tìm dựa trên từ khóa. Kết quả tìm kiếm được trình bày cứ 5 tư liệu một trang (sử dụng cách sắp xếp tương thích mặc định của Greenstone), và do đó, dàn trang được coi như một tương tác. Tương tự, việc chỉnh lý hoặc hiệu đính các từ khóa cũng được coi là một tương tác mới. Cũng như trước đây, chúng tôi ghi những tương tác của người sử dụng một cách thụ động để phân tích sau này. Số lượng trung bình các tương tác cho mỗi nhóm (của mình so với của những người khác) được liệt kê trong Bảng 1. Không có mẫu chính quy về số lượng trung bình các tương tác được yêu cầu để định vị tư liệu, bất kể là tài liệu quen thuộc (tư liệu lấy từ giá sách riêng của người sử dụng) hoặc không (tư liệu lựa chọn từ giá sách của người khác).

Nhiệm vụ tìm kiếm nâng cao:

Về nhiệm vụ tìm kiếm nâng cao, người sử dụng được nhận một tập hợp 6 bài tóm tắt khác (3 bài tóm tắt mới từ giá sách riêng của họ và 3 bài tóm tắt mới nữa được lấy từ những giá sách của người khác). Lần này, chúng tôi yêu cầu người sử dụng dùng phép tìm kiếm theo từ khóa trên hệ thống nâng cao, nơi chúng tôi đã có diện động và quá trình sử dụng của họ được ghi lại và nội dung thư viện số được nâng cao thêm, có sử dụng phép tách cụm từ, suy ra đề tài theo hướng bản thể

học, v.v... Tuy nhiên, những kết quả tìm kiếm trong hệ thống nâng cao được chọn ra đối ứng với diện động của người sử dụng (mà lần lượt phản ánh mẫu sử dụng của chúng; khác với kết quả tìm kiếm cơ bản). Chúng tôi đã yêu cầu người sử dụng tìm kiếm tất cả 6 tư liệu, dựa vào từ khóa. Kết quả tìm kiếm được trình bày cứ 5 tư liệu một trang (hệt như là tìm kiếm cơ bản); và những tương tác của người sử dụng được ghi lại như trước đây. Số lượng trung bình các tương tác cho mỗi nhóm (của mình so với của những người khác) được liệt kê trong Bảng 1 để so sánh. Trong hầu hết trường hợp, người sử dụng định vị các tư liệu riêng của họ với một số lượng tương tác trung bình thấp, trong khi số lượng tương tác trung bình được yêu cầu cho các tư liệu xa lạ lại cao hơn. Điều này là do kết quả tìm kiếm được sắp xếp theo sự ưu tiên của người sử dụng và văn cảnh, và do đó, cần phải có sự dàn trang và chỉnh lý lại các thuật ngữ tìm kiếm. Khả năng không sử dụng các từ khóa (kết hợp từ khóa) đúng đối với các tư liệu xa lạ cũng không thể loại trừ hết. Tuy nhiên, trong khung cảnh thực nghiệm thăm dò với quy mô nhỏ, chúng tôi tiếp tục thăm dò hiện tượng này bằng cách phỏng vấn những người tham gia.

Nhìn ra thế giới

Bảng 1. Số lượng tương tác trung bình trong 3 nhiệm vụ khác nhau. Giá trị cho biết số lượng tương tác trung bình đối với một nhóm nhỏ gồm 3 tư liệu

người sử dụng #	Duyệt lướt cơ bản		Tìm kiếm cơ bản		Tìm kiếm nâng cao	
	tài liệu của mình / 3	tài liệu của người khác / 3	tài liệu của mình / 3	tài liệu của người khác / 3	tài liệu của mình / 3	tài liệu của người khác / 3
#1	3.3	6.3	3.7	4.0	2.0	5.3
#2	4.0	6.7	4.0	4.0	3.3	4.0
#3	3.7	5.0	4.0	5.3	2.3	5.0
#4	3.0	6.7	5.0	4.7	2.7	4.3

4.3. Những đánh giá thực nghiệm khác

Trong phiên phỏng vấn, chúng tôi đã cho từng người sử dụng biết các phương pháp và mục tiêu thử nghiệm của mình. Chúng tôi cũng trình bày diện động của người sử dụng (các đề tài hiện có trọng số dùng ngưỡng tách biệt), và diện đăng ký ban đầu của họ kê bên nhau và yêu cầu họ đánh giá diện nào phản ánh sự quan tâm của họ nhiều hơn. Cũng có nhiều người đồng ý và nhiều người phản đối. Tuy nhiên, khi chúng tôi tiếp tục giải thích diện động đã liên quan đến hoạt động của họ (như các từ khóa mà họ sử dụng và các tài liệu mà họ chọn cho giá sách) như thế nào và khung cảnh của chúng (như là hướng về sở thích), thì tất cả 4 người sử dụng đều đồng ý rằng, diện động có liên quan với việc sử dụng và hoạt động của họ.

Trong giai đoạn cuối cùng của việc đánh giá thăm dò, chúng tôi đã sử

dụng môđun khuyến nghị để tách ra 5 tư liệu hàng đầu cho mỗi người sử dụng. Chúng tôi yêu cầu người sử dụng đọc nhan đề và các bài tóm tắt và phê duyệt bao nhiêu trong số các tư liệu này họ sẽ đọc tiếp nếu được hệ thống khuyến nghị trong lần đăng nhập tiếp theo. Những câu trả lời của 4 người sử dụng tập hợp được cho thấy 14 trong số 20 tư liệu khuyến nghị đã được chọn xứng đáng để đọc tiếp.

5. Kết luận

Trong bài này, chúng tôi trình bày một kiến trúc thư viện số 3 lớp, nhằm phục vụ cho các dịch vụ thư viện số thông minh và tùy biến. Chúng tôi toan tính tích hợp diện người sử dụng, bản thể học lĩnh vực, mẫu sử dụng và phân tích nội dung thư viện số với nhau trong việc hình thành các dịch vụ thông minh và tùy biến cho các người sử dụng thư viện số. Công trình thử nghiệm và phân tích thăm dò này có sử dụng một nguyên mẫu nhỏ cho thấy rằng, cách tiếp cận và các thuật

toán của chúng tôi nhằm tích hợp một số diện của kịch bản tìm tin trong thư viện số; và do đó, chúng tôi dự định tiếp tục nâng cao các thuật toán và tích hợp chúng một cách hoàn hảo cho một sưu tập và các người sử dụng quy mô lớn.

Tìm tin trong khung cảnh thư viện số rất độc đáo và đa diện. Do đó, mong rằng các nhà nghiên cứu về thư viện số sẽ ngày một thích nghi với những bước phát triển mới đây của

việc nghiên cứu có liên quan đến sự tương tác giữa người và máy tính và những công trình nghiên cứu về người sử dụng trong khung cảnh thư viện số. Do đó, chúng tôi dự định tích hợp những tương tác hỗn hợp như tích hợp giữa duyệt lướt và tìm kiếm và tích hợp các tương tác đến lướt và hết lướt [18], [19] để tạo điều kiện thuận lợi cho việc thu nhận những kinh nghiệm tìm tin nâng cao trong môi trường thư viện số.

Tài liệu tham khảo

1. Chowdhury G.G., Chowdhury S. : *Introduction to Digital Libraries*. Facet Publishing, London (2003)
2. Feng I., Jensfeld M.A., Hoppenbrouwers J. *Beyond Information Searching and Browsing: Acquiring Knowledge from Digital Libraries*. (Retrieved March 25) (2007)
3. <http://citesser.ist.psu.edu/421460.html>
4. Marchionini G. *Information Seeking in Electronic Environments*. Cambridge Series on Human- Computer Interactions. Cambridge University Press, Cambridge (1997)
5. Straccia U.: *Collaborative Working in the Digital Library Environment*. Cyclades (Retrieved March 12) (2007) <http://dlibcenter.iei.pi.cnr.it/>
6. Hurley B.J., Price-Wilkin J., Proffitt M., Besser H. *The Making of America II. Test-bed Project: A Digital Library Service Model*. The Digital Library Federation. Washington DC (1999)
7. Brusilovsky P. : *Adaptive Hypermedia. User Modeling and User -Adapted Interaction*. 11(1-2), 87-110 (2001)
8. Crestani F.: *Application of Spreading Application Techniques in Information Retrieval*. *Artificial Intelligence Review* 11(6), 453-482 (1997)
9. *Greenstone Digital Library Software. Project*. Retrieved 2/2/2007 from <http://www.greenstone.org/>
10. ACM-CCS Add-on Ontology. University Minho Web Site (Accessed March 12, 2006). http://dSPACE-dev.dsi.uminho.pt:8080/en/research_about.jsp
11. Witten I.H., Paynter G.W., Frank E. Gutwin C., Nevill-Manning and C.G.: *KEA: Practical Automatic Keyphrase Extraction*. In: *Fourth ACM Conference on Digital Libraries DL 1999*, pp.254-255. ACM New York (1999)
12. Pitkow J., Schutze H., Cass T., Cooley R., Turnbull D., Edmonds A., Adar E., Breuel T.: *Personalized Search*. *Communications of ACM* 45(9), 50-55 (2002)
13. Dumais S., Cutrell E., Chen H. : *Optimizing Search by Showing Results in Context*. In: *ACM Conference on Human Factors in Computing Systems (CHI 2001)*, Seattle, WA, pp. 277-284. ACM Press, New York (2001)
14. *Forecasting with Single Exponential Smoothing*, NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved 10/02/2007 from <http://www.itl.nist.gov/div898/handbook>

Nhìn ra thế giới

15. Liao I.E., Liao S.C., Kao K.F., Harn I.F.: A personal Ontology Model for Library Recommendation System. In: Sugimoto S., Hunter J., Rauber A., Morishima A. (eds.) ICADL, 2006, LNCS, vol.4312, pp.173-182. Springer, Heidelberg (2006)

16. Middleton S.E., De Roure D.C., Shadbolt. N.R.: Capturing Knowledge of User Preferences. Ontologies on Recommender Systems. In: First International Conference on Knowledge Capture (K-CAP2001) pp. 100-107 (2001)

17. Sarwar B., Karypis G., Konstan J., Riedl J.: Item-based Collaborative Filtering Recommendation Algorithms. In: 10th Inter-

national World Wide Web Conference (WWW 2010). Hong Kong, pp. 285-295 (2001)

18. Ding Y., Li X.: Time Weight Collaborative Filtering. In: 14th ACM International Conference on Information and Knowledge Management, pp. 485-492 (2005)

19. Olston C., Chi E.H.: ScentTrails: Integrating Browsing and Searching on the Web. ACM Transactions on Computer-Human Interaction 10(3), 177-197 (2003)

20. Perugini S., Ramakrishnan N.: Personalizing Web Sites with Mixed-Initiative Interaction. IEEE IT Professional 5(2), 9-15 (2003).

Vũ Văn Sơn dịch

Tài liệu gốc: "Digital libraries: Universal and Ubiquitous Access to Information", pp. 109-113

CƠ SỞ DỮ LIỆU KẾT QUẢ NGHIÊN CỨU

CSDL KQNC được Cục Thông tin KH&CN Quốc gia triển khai xây dựng từ năm 1990, quản trị tài liệu về báo cáo kết quả của các đề tài, dự án nghiên cứu khoa học cấp nhà nước, cấp bộ, cấp tỉnh/thành phố và cấp cơ sở được thực hiện trên phạm vi toàn quốc (Diện bao quát đề tài: Khoa học xã hội và nhân văn, Khoa học tự nhiên, Khoa học kỹ thuật; Các lĩnh vực khoa học nổi bật: Nông nghiệp, Lâm nghiệp, Thủy sản, Thủy lợi, Y tế, Kinh tế, Giáo dục, Nhà nước và pháp luật, Bảo vệ môi trường, Toán học, Tin học, Kỹ thuật điện, Điện tử, Chế tạo máy,...). Hiện nay, CSDL KQNC đã có hơn 10.000 biểu ghi, số lượng cập nhật: 1.000 tài liệu/năm. Mọi cá nhân và đơn vị có nhu cầu khai thác và chuyển giao:

- * *Tìm tin thư mục*
- * *Tìm tin trên mạng VISTA*
- * *Cung cấp bản sao tài liệu gốc (dạng giấy và điện tử)*
- * *Cung cấp thông tin theo yêu cầu*
- * *Chuyển giao CSDL KQNC thư mục theo yêu cầu*

XIN LIÊN HỆ THEO ĐỊA CHỈ
ĐĂNG KÝ KQNC & ISSN
CỤC THÔNG TIN KHCN QUỐC GIA
24, Lý Thường Kiệt, Hà Nội
ĐT: 04-39349116. Fax: 04-39349127
E-mail: Phque@vista.gov.vn