

WEB SCALE DISCOVERY - GIẢI PHÁP TÌM KIẾM VÀ KHAI THÁC THÔNG TIN CHO CÁC THƯ VIỆN TRONG KỶ NGUYÊN INTERNET

ThS Đỗ Văn Hùng

Trường Đại học KHXH&NV Hà Nội

Tóm tắt: *Tìm kiếm thông tin tập trung - Web scale discovery (WSD) đang là xu thế hiện nay của các thư viện trong môi trường trực tuyến và có nhiều cơ sở dữ liệu phân tán. Bài viết lý giải sự xuất hiện hệ thống tìm kiếm tập trung WSD, đồng thời đưa ra khái niệm và cấu trúc cơ bản của hệ thống. Trên cơ sở đó, đưa ra một số gợi ý cho các thư viện (cụ thể là thư viện đại học) Việt Nam trong việc phát triển và áp dụng hệ thống này.*

Từ khóa: *Tìm kiếm thông tin tập trung; Web scale discovery; thư viện; khai thác thông tin; giải pháp tìm kiếm thông tin.*

Web scale discovery – information searching and exploring solution for libraries in the Internet era

Summary: *Web scale discovery (WSD) is the emerging activity of libraries operating in the online environment and consisting of different fragmented databases. The article explains the emergence of the Web scale discovery WSD, provides the basic definition and structure of the system as well as suggestions of how to develop and apply this system into Vietnamese libraries (especially university libraries).*

Keywords: *Web scale discovery; library; information extraction; information searching solution.*

1. Bối cảnh và yêu cầu cần có hệ thống tìm kiếm tập trung

Các thư viện đang chuyển hoạt động của mình sang môi trường trực tuyến. Đó là môi trường mà thông tin đang dần chuyển sang định dạng số, kể cả tài liệu in ấn cũng được chỉ mục trong cơ sở dữ liệu (CSDL) thư mục, bên cạnh đó là thói quen khai thác và sử dụng tài liệu trực tuyến của bạn đọc. Đưa thông tin lên Internet, tiến gần với bạn đọc hơn thông qua thế giới mạng đang là cách tiếp cận của các thư viện hiện nay. Một trong những vấn đề đặt ra cho các thư viện là làm sao giúp bạn đọc có thể tìm kiếm và khai thác thông tin mà mình cần một cách dễ dàng và hiệu quả nhất.

Các thư viện hiện nay có nhiều CSDL khác nhau và được tổ chức phân tán. Nguồn thông tin trong thư viện hiện nay không còn là chỉ trong nội bộ thư viện như các kho sách, các CSDL trong máy chủ của thư viện, mà các thư viện đang hướng tới mua bản quyền các CSDL toàn văn trực tuyến từ các nhà cung cấp

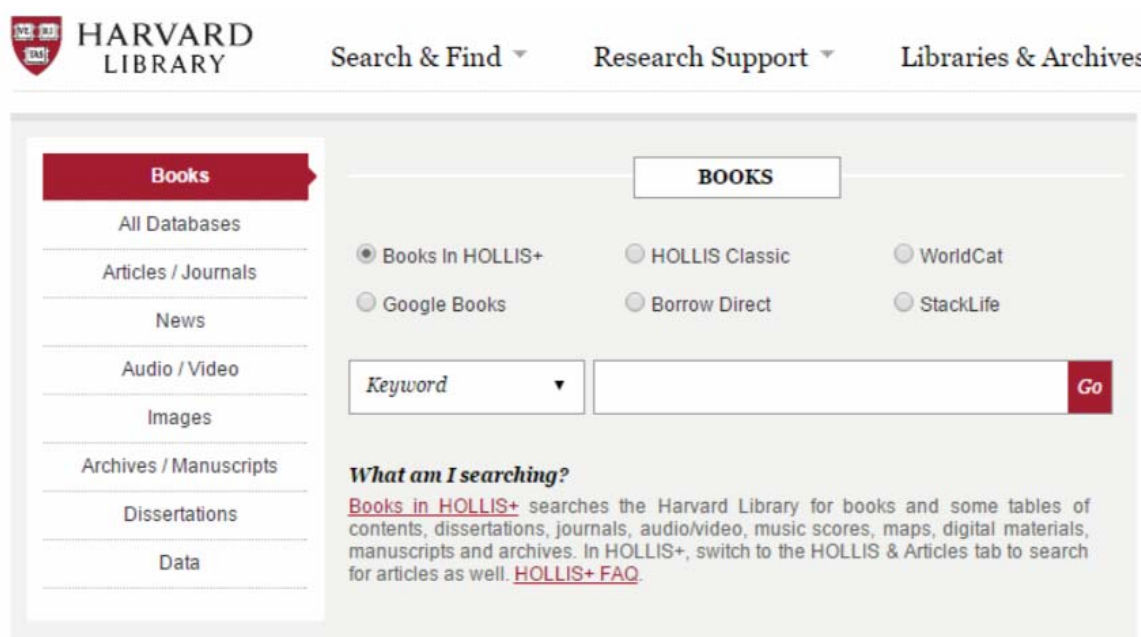
thông tin chuyên nghiệp, cũng như cố gắng khai thác các nguồn thông tin có chất lượng và miễn phí trên Internet. Mục tiêu là đưa ra nhiều lựa chọn, đa dạng hóa nguồn cung cấp thông tin cho bạn đọc. Để giúp bạn đọc khai thác thông tin, cách làm sơ khai của các thư viện là tạo liên kết đến các nguồn thông tin này trên website của thư viện, bạn đọc phải vào từng liên kết để tìm kiếm thông tin. Ngay cả đối với nguồn thông tin do chính thư viện tạo ra, bạn đọc cũng phải tìm kiếm nhiều lần ở các CSDL khác nhau như: CSDL thư mục và CSDL toàn văn. Bạn đọc phải đăng nhập nhiều lần cho các CSDL này. Lý do là một số thư viện quản lý dữ liệu thư mục ở phần mềm quản trị thư viện điện tử tích hợp. Trong khi đó tài liệu số được lưu trữ ở các phần mềm như: Dspace, Greenstone hay Content Pro. Các phần mềm này chưa có sự liên kết với nhau.

Dù muốn hay không, việc tồn tại các CSDL phân tán là điều không thể tránh khỏi ở

Nghiên cứu - Trao đổi

các thư viện. Bài toán đặt ra là, bạn đọc không cần biết thư viện có bao nhiêu CSDL và từ nguồn nào, họ chỉ cần vào một giao diện web duy nhất, với một công cụ tìm kiếm duy nhất có thể tìm kiếm và khai thác được tất cả các CSDL mà thư viện đang có, kể cả các CSDL ngoài thư viện. Đây chính là nhu cầu tìm kiếm thông tin tập trung mà các thư viện đang phải đối mặt trong quá trình xây dựng không gian học tập trực tuyến. Hệ thống tìm kiếm tập trung - Web scale discovery - WSD ra đời

nhằm đáp ứng nhu cầu này. Mong muốn của bạn đọc là thư viện có một hệ thống tìm kiếm như Google: đơn giản và hiệu quả. Với một thanh công cụ tìm kiếm đơn giản nhưng có khả năng khám phá được nhiều nguồn thông tin khác nhau và kết quả trả về được trình bày thân thiện và theo một định dạng thống nhất. Đặc trưng về giao diện của hệ thống WSD đó là tối giản hóa việc tìm kiếm bằng một ô tìm kiếm duy nhất (single search box) như chúng ta thường thấy ở công cụ tìm kiếm google.



Hình 1. Giao diện tìm kiếm của thư viện đại học Harvard sử dụng hệ thống Primo của Ex Libris

2. Hệ thống tìm kiếm tập trung WSD

2.1. Khái niệm

Hệ thống tìm kiếm tập trung WSD ra đời khoảng giữa những năm 2000 và đã nhanh chóng được nhiều thư viện áp dụng, đặc biệt là các thư viện đại học [10]. Triết lý của hệ thống này là lấy người dùng làm trung tâm của hệ thống. Tương tác người dùng đầu cuối “Front-End” của WSD nhằm tới việc cung cấp thống nhất cách trình bày kết quả theo một định dạng chung, trên một giao diện web duy nhất cho dù các nguồn dữ liệu là khác nhau hay các định dạng khác

nhau. Đây chính là điểm ấn tượng của hệ thống WSD. Hệ thống này cho bạn đọc cảm giác đang tìm kiếm và khai thác thông tin ở chính trong thư viện của mình. Thực tế, có thể gọi WSD là hệ thống (Web-scale discovery systems) hoặc là dịch vụ (Web scale discovery services), bởi WSD bao gồm cả các giải pháp tìm kiếm và các dịch vụ liên quan đến khai thác thông tin đi kèm.

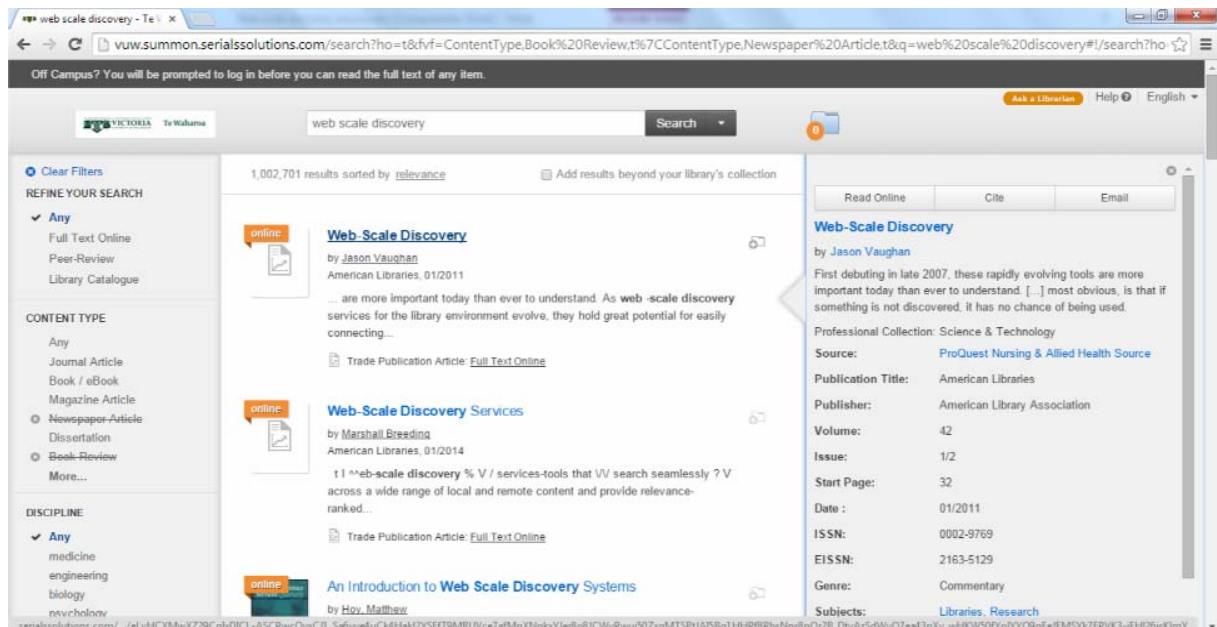
Hệ thống WSD dành cho các thư viện được định nghĩa là máy tìm kiếm (search engines), cho phép người dùng tìm kiếm không chỉ đơn

Nghiên cứu - Trao đổi

thuần là nhan đề của cuốn sách, tạp chí hay video, mà giúp người dùng khai thác thông tin trong từng nội dung của bài tạp chí, từng chương trong một cuốn sách và nhiều hơn thế nữa [10]. Về lý thuyết thì các máy tìm kiếm này mang đến một công cụ tìm kiếm đầy sức mạnh như Google cho tất cả các nội dung tài liệu mà thư viện đang nắm giữ hoặc có mua bản quyền từ bên ngoài [11]. Trong thực tế thì các máy tìm kiếm của WSD đã làm khá tốt chức năng này. Dĩ nhiên không có một dịch vụ WSD hoàn hảo, nó phụ thuộc rất nhiều vào người sử dụng hệ thống đó cũng như nguồn lực thông tin mà các thư viện có thể cung cấp. Nhưng ít nhất, một trong những điểm mạnh của hệ thống WSD là giải quyết được những vấn đề khó khăn trong tìm kiếm các bài báo của tạp chí. Với hàng ngàn đầu tạp chí, hàng triệu bài báo đề cập đến tất cả các lĩnh vực khoa học, việc tìm kiếm, sắp xếp và chuyển tải đến người dùng là vấn đề không đơn giản. WSD làm cho công việc này dễ dàng hơn. Điểm khác biệt giữa các máy tìm kiếm trên Internet phổ biến hiện nay (ví dụ như: Google,

Yahoo hay Bing) với hệ thống tìm kiếm tập trung là WSD đánh chỉ mục, tìm kiếm và cung cấp những nguồn thông tin đã qua sàng lọc và thẩm định (bởi thư viện) trong khi các máy tìm kiếm Internet đánh chỉ mục toàn bộ nguồn thông tin mà nó có thể với tới trên Internet.

Theo Hiệp hội Thư viện Hoa Kỳ (American Library Association) thì dịch vụ WSD là một công cụ đầy tiềm năng để biến đổi bản chất của hệ thống thư viện. Các dịch vụ này có khả năng tìm kiếm thông tin nhanh chóng từ nhiều nguồn khác nhau (trong và ngoài thư viện, nội bộ hay từ xa), tạo ra sự liền mạch và thống nhất trong việc cung cấp thông tin, thông tin được tìm kiếm trên một phạm vi rộng lớn, các kết quả tìm kiếm được xếp hạng và có gợi ý các tài liệu liên quan, kết quả được trình bày trong một giao diện trực quan đúng như mong đợi của người tìm kiếm thông tin. Mỗi hệ thống sẽ bao gồm ba thành tố cơ bản: nội dung thông tin, giao diện trình bày và các chức năng liên quan đến tìm kiếm, thu thập và phân phối thông tin.



Hình 2. Giao diện kết quả tìm kiếm của Đại học Victoria sử dụng hệ thống Summon của ProQuest

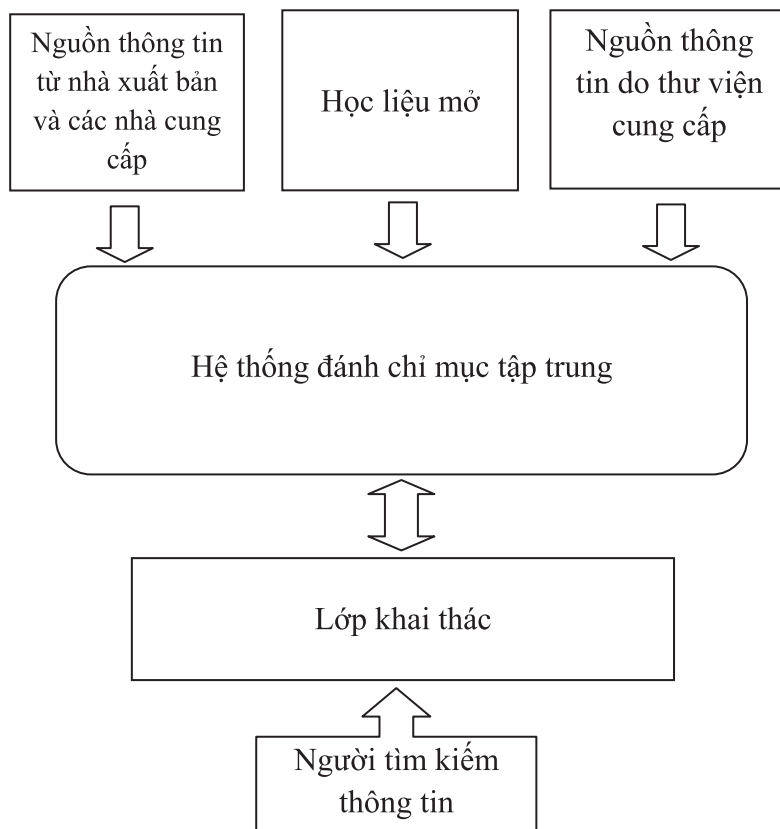
Nghiên cứu - Trao đổi

Một trong những ưu điểm của giải pháp tìm kiếm tập trung WSD là không chỉ cung cấp giải pháp về công nghệ tìm kiếm, mà các nhà cung cấp này còn giúp các thư viện đàm phán với các nhà cung cấp dữ liệu trực tuyến để bạn đọc của các thư viện có thể tìm kiếm và khai thác CSDL của các nhà cung cấp dữ liệu, các nhà xuất bản hay các tạp chí danh tiếng. Tức là tiến tới những thỏa thuận về hợp tác và chia sẻ dữ liệu. Đây chính là điều tạo nên sự khác biệt của tìm kiếm tập trung. Có thể nói, thỏa thuận giữa nhà cung cấp dịch vụ tìm kiếm thông tin và các nhà cung cấp thông tin đóng vai trò quan trọng trong giải pháp tìm kiếm tập trung. Chính vì vậy, tên của các giải pháp này thường có thêm từ “services” - dịch vụ (Web Scale Discovery Services) để chỉ rằng đây không chỉ là giải pháp về công nghệ mà bao hàm cung cấp các dịch vụ kèm theo [3]. Điều này cũng lý giải

tại sao hệ thống WSD thường có giá rất đắt và phải mua bản quyền theo khoảng thời gian nhất định.

2.2. Cấu trúc của WSD

Một hệ thống WSD bao gồm các thành phần sau: nguồn thông tin (nguồn thông tin nội bộ do thư viện quản lý và nguồn từ xa, học liệu mở và nguồn thông tin thư viện mua từ các nhà xuất bản hay nhà cung cấp thông tin chuyên nghiệp), hệ thống đánh chỉ mục tập trung; lớp khai thác và người dùng/người tìm kiếm thông tin. Các thành phần và mối liên hệ trong một hệ thống WSD được mô tả như Hình 3. Trong đó, xét về mặt công nghệ, có hai thành phần quan trọng là lớp khai thác và hệ thống đánh chỉ mục tập trung. Hệ thống chỉ mục tập trung được coi là nội dung, trong khi đó lớp khai thác sẽ là giao diện để người dùng tương tác với nội dung.



Hình 3. Cấu trúc một hệ thống WSD [10]

Nghiên cứu - Trao đổi

Lớp khai thác (Discovery layer - DL)

Lớp khai thác được định nghĩa là *giao diện người dùng và hệ thống tìm kiếm* để khai thác, hiển thị kết quả và tương tác với nội dung thông tin với người dùng. Lớp khai thác có mối liên hệ chặt chẽ với hệ thống đánh chỉ mục tập trung. Về phía người dùng, lớp khai thác chính là giao diện tổng hợp mà ở đó họ được cung cấp các dịch vụ của WSD. DL được cấu trúc như là một cổng thông tin - nơi giao tiếp giữa bạn đọc và thư viện. Thông qua cổng thông tin này người dùng có thể khám phá và khai thác các nguồn tin mà thư viện cung cấp. Lớp khai thác có những đặc tính sau:

- Chức năng tìm kiếm duy nhất (single search) dựa trên nền tảng hệ thống chỉ mục tập trung (central index). Với một ô tìm kiếm đơn giản và duy nhất người dùng có thể tìm kiếm nhiều nguồn tài nguyên thông tin điện tử trong và ngoài thư viện. WSD có công cụ tìm kiếm thông minh và mạnh. Hệ thống hỗ trợ người dùng tìm kiếm theo từ khóa, tác giả, nhan đề hay chủ đề với các toán tử hỗ trợ, hay tìm theo cấu trúc cũng như kiểm tra chính tả và gợi ý đối với từ hoặc cụm từ không chính xác được đưa vào tìm kiếm.

- Thời gian phản hồi kết quả tìm kiếm nhanh. Đây là điều kiện tiên quyết đối với năng lực của các hệ thống WSD. Phản hồi kết quả tức thì là yêu cầu quan trọng của một máy tìm kiếm.

- Xếp hạng mức độ liên quan trong kết quả tìm. Đưa các kết quả có nội dung gần nhất đối với yêu cầu tìm kiếm thông qua việc xếp hạng thông minh các kết quả tìm cũng như gợi ý các tài liệu liên quan đến yêu cầu của người tìm.

- Có các công cụ để sắp xếp, thu hẹp, lọc... dựa trên kết quả đã tìm kiếm được. Người dùng có thể tinh chế yêu cầu tìm tin, thu hẹp kết quả tìm kiếm dựa trên các tiêu chí về thời gian, loại hình, chủ đề... của tài liệu.

- Liên kết trực tiếp đến các nguồn dữ liệu số - đây là thế mạnh của các dịch vụ WSD. WSD không chỉ đưa ra các thông tin về tài liệu tìm được mà còn hỗ trợ bạn đọc truy cập đến tài liệu toàn văn ngoài thư viện thông qua xác thực định danh là người dùng của thư viện mà bạn đọc đang sử dụng dịch vụ.

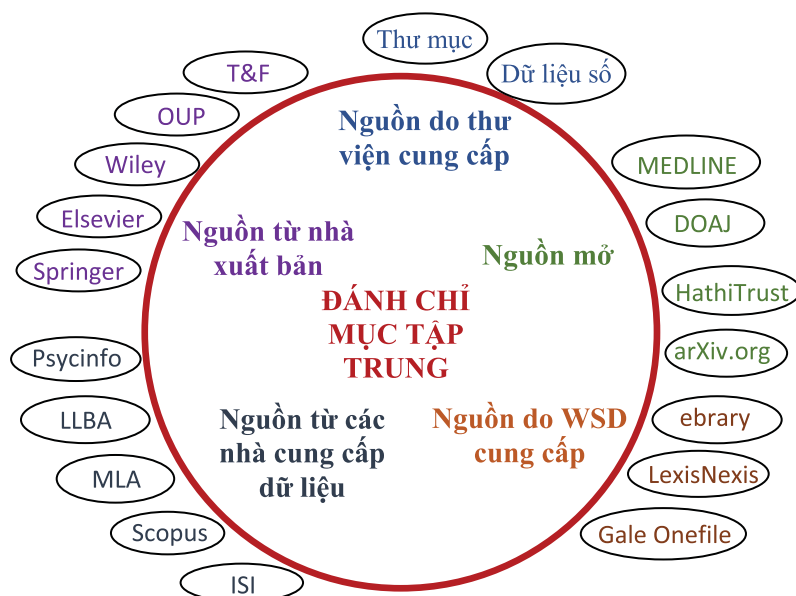
- Có khả năng tùy biến và cá nhân hóa cho người dùng đầu cuối. Mỗi người dùng đều được cung cấp một tài khoản để xác nhận quyền truy cập cũng như tùy biến các kết quả tìm của mình. Dựa trên nền Web 2.0, người dùng có thể tổ chức thông tin, lưu trữ và kết xuất kết quả tìm kiếm, cũng như đặt thẻ (tag), xếp hạng hoặc chia sẻ để giới thiệu tài liệu tìm được. Các thư viện cũng có thể gắn logo hoặc các thông tin về thư viện trên giao diện của dịch vụ WSD.

Hệ thống chỉ mục dữ liệu tập trung (Central index)

Hệ thống sẽ tiến hành đánh chỉ mục đối với toàn bộ nội dung mà thư viện cung cấp, đó có thể là từ nhà xuất bản gốc, nhà xuất bản cấp hai hay nhà tích hợp nội dung và những thông tin nội sinh của thư viện. Có thể nói dịch vụ chỉ mục dữ liệu tập trung chính là trái tim của hệ thống WSD. Nó tạo ra sự kết nối người dùng với các nguồn tài nguyên thông tin. Hệ thống này thực chất là tạo ra một CSDL tổ hợp hay một mục lục trung tâm chứa các biểu ghi của tất cả nguồn tài nguyên mà thư viện quản lý cũng như những siêu liên kết đến kho tài nguyên số trực tuyến. Hình 4 mô tả các nguồn thông tin được đánh chỉ mục tập trung trong hệ thống WSD.

Các nguồn thông tin bao gồm: nguồn trong thư viện, nguồn mở miễn phí trên Internet, nguồn của chính các nhà cung cấp dịch vụ WSD, nguồn từ các cung cấp dữ liệu chuyên nghiệp và nguồn từ các nhà xuất bản.

Nghiên cứu - Trao đổi



Hình 4. Mô hình đánh chỉ mục tập trung của WSD [6]

Đối với một hệ thống đánh chỉ mục tập trung thì độ lớn của CSDL đánh chỉ mục không quan trọng bằng các yếu tố như: phạm vi và độ sâu của nội dung thông tin, loại hình tài liệu, tài liệu gốc toàn văn hay độ phong phú của siêu dữ liệu. Một hệ chỉ mục hiệu quả là giúp các thư viện cung cấp thông tin phù hợp với nhu cầu bạn đọc dựa trên những nội dung mà thư viện đang có. Hệ thống phải hiểu được bạn đọc cần gì và đưa ra những gợi ý phù hợp. Chính vì thế mà ở một khía cạnh nào đó WSD được coi là trí tuệ nhân tạo của *năng lực thông tin* (information literacy).

Có hai khái niệm trong chỉ mục tập trung đó là đánh chỉ mục trước (pre- harvested index) và đồng sở hữu nội dung thông tin (mutually licensed content). Các siêu dữ liệu và tài liệu toàn văn được thu gom một cách hệ thống và định kỳ và được xử lý trước để phục vụ cho việc tìm kiếm thông tin sau này. Dữ liệu được thu thập từ nhiều nguồn khác nhau và được đưa vào hệ thống chỉ mục tập trung để xử lý. Đối với nguồn thông tin có bản quyền (của thư viện và các nhà cung cấp WSD) thì người dùng phải xác thực định danh để truy cập đến tài liệu toàn văn.

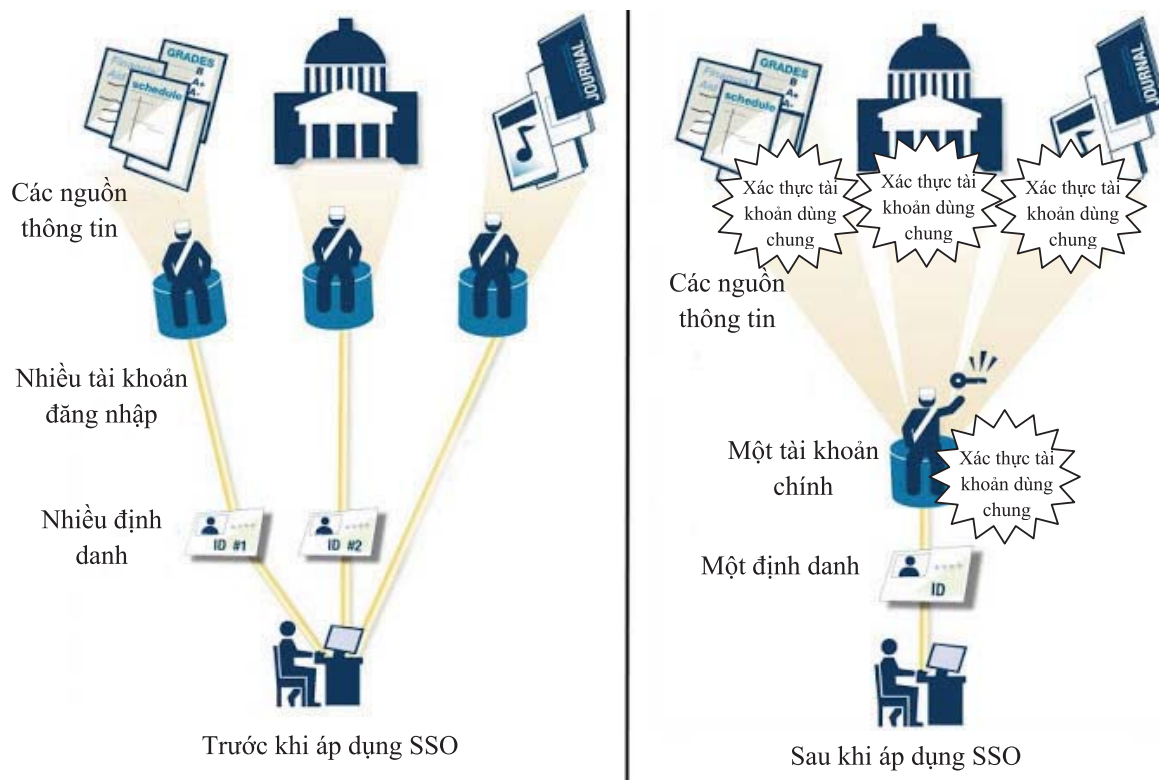
Đăng nhập một lần - Single sign on

Dịch vụ đăng nhập một lần - single sign on (SSO) là một tính năng giúp bạn đọc thuận tiện hơn khi sử dụng các dịch vụ thư viện. Bạn đọc có tài khoản của thư viện chỉ cần đăng nhập một lần là có thể sử dụng nhiều dịch vụ khác nhau của thư viện. Thông thường các dịch vụ trong thư viện được thực hiện bởi nhiều ứng dụng từ nhiều nhà cung cấp giải pháp khác nhau. Khi bạn đọc truy cập đến các dịch vụ hay tài nguyên thông tin do thư viện cung cấp thường phải xác thực quyền truy cập vào hệ thống. Nếu không có sự thống nhất trong thiết kế hệ thống, bạn đọc phải đăng nhập nhiều lần khi sử dụng các dịch vụ khác nhau trong thư viện. Ví dụ, người dùng phải đăng nhập để biết thông tin về mượn trả hay gia hạn mượn tài liệu. Khi khai thác CSDL số trong thư viện, người dùng lại phải đăng nhập lần nữa để xác thực định danh, việc này cũng sẽ diễn ra tương tự nếu người dùng khai thác nguồn thông tin trực tuyến mà thư viện mua bản quyền truy cập. Điều này sẽ gây phiền toái cho bạn đọc. SSO giúp bạn đọc xác thực quyền của mình đối với toàn bộ dịch vụ mà thư viện cung cấp

Nghiên cứu - Trao đổi

với một lần đăng nhập duy nhất. Ngoài việc giúp bạn đọc tiện lợi trong sử dụng dịch vụ của thư viện, SSO còn giúp các thư viện tăng

cường tính bảo mật của hệ thống và dễ dàng trong việc quản trị cũng như giảm chi phí trong quản lý.



Hình 5. Sơ đồ mô tả đăng nhập trước và sau khi áp dụng SSO [2]

SSO kết hợp với WSD là một giải pháp tổng thể trong quản trị thư viện trong thời đại số và Internet. Mục tiêu là mang đến cho người dùng sự dễ dàng và thuận tiện trong việc tiếp cận các dịch vụ của thư viện. Bên cạnh đó giải pháp này giúp các thư viện quản trị hệ thống của mình dễ dàng hơn và nâng cao năng lực cung cấp thông tin của thư viện.

2.3. Một số nhà cung cấp dịch vụ WSD trên thế giới

WSD là dịch vụ mới so với lịch sử ứng dụng công nghệ thông tin trong lĩnh vực thư viện. Tuy nhiên, dịch vụ này đang được rất nhiều thư viện quan tâm. Hiện nay, trên thế giới có một số dịch vụ WSD nổi tiếng có thể

kể đến như: Summon (của ProQuest), Primo (của Ex Libris), EBSCO Discovery Service (của EBSCO) và WorldCat Discovery Services (của OCLC). Có thể thấy rằng những tên tuổi này đều là những nhà cung cấp thông tin và giải pháp công nghệ chuyên nghiệp. Các nhà cung cấp này có sự cạnh tranh thị phần với nhau, song trong một số trường hợp họ phải hợp tác với nhau vì có những thư viện sử dụng nguồn dữ liệu từ nhiều nhà cung cấp [7].

Dịch vụ Summon của ProQuest hiện có hơn 500 thư viện trên thế giới sử dụng, khách hàng tiêu biểu của dịch vụ này là Đại học Harvard, Hoa Kỳ và Đại học Victoria, New Zealand [9].

Nghiên cứu - Trao đổi

ProQuest là một nhà cung cấp CSDL toàn văn về nhiều lĩnh vực khoa học, ngoài ra còn cung cấp các giải pháp, ứng dụng và các sản phẩm cho thư viện. Trong đó Summon là một giải pháp tìm kiếm tập trung nổi bật. Summon là một giải pháp công nghệ mới, dễ sử dụng với tiêu chí chỉ cần một cổng tìm kiếm duy nhất có thể truy cập đến tất cả các CSDL phân tán những như các CSDL đơn lẻ trực tuyến khác. Mục tiêu của Summon là giúp các thư viện tạo lập một môi trường học thuật trong đó hỗ trợ bạn đọc của thư viện tiếp cận tới các nguồn thông tin mà họ cần thông qua một cổng thông tin duy nhất của thư viện [1].

Ex Libris là nhà cung cấp các giải pháp tự động hóa cho thư viện. Ex Libris tập trung vào xây dựng giải pháp tổng thể cho thư viện bao gồm: khai thác, quản trị và phân phối các nguồn tài liệu trong thư viện kể cả tài liệu in ấn và tài liệu số. Dịch vụ Primo của Ex Libris đang được trên hơn 2.100 thư viện sử dụng. Đây là một trong những dịch vụ WSD phổ biến nhất hiện nay. Primo giúp người dùng khám phá tri thức từ nhiều nguồn khác nhau và chuyển giao trực tiếp đến người dùng. Dịch vụ này giúp người dùng cá nhân hóa phương thức tìm kiếm, đánh giá và sắp xếp kết quả tìm kiếm và chia sẻ lên mạng xã hội. Cũng như các dịch vụ WSD khác, đánh chỉ mục tập trung chính là điểm mạnh của Primo trong việc giúp người dùng tìm kiếm thông tin trong các CSDL phân tán [5].

EBSCO Discovery Service (EDS) là một dịch vụ WSD của EBSCO - nhà cung cấp dữ liệu và giải pháp tự động hóa cho thư viện. Với triết lý hướng tới sự tiện lợi và đơn giản hóa việc tìm kiếm thông tin cho người dùng, EDS cung cấp công cụ tìm kiếm đa nguồn (OPACs, tạp chí điện tử, sách điện tử và các CSDL toàn văn) trong một cửa sổ tìm kiếm duy nhất, có khả năng phân loại và sắp

xếp kết quả tìm kiếm theo mức độ ưu tiên (streamlined search) và trình bày trên một giao diện thân thiện và đồng nhất. EBSCO phối hợp với các nhà cung cấp thông tin để tích hợp vào dịch vụ EDS nhằm mục tiêu hỗ trợ người dùng tìm kiếm, truy cập những thông tin có chất lượng tốt nhất, phục vụ hiệu quả cho công việc của mình [4].

OCLC (Trung tâm thư viện máy tính trực tuyến) cung cấp dịch vụ WorldCat discovery services với khả năng tìm kiếm và truy cập đến 1,8 tỷ tài liệu của 74.000 thư viện trên thế giới. Đây có thể coi là một ứng dụng điện toán đám mây (cloud-based application). Ngoài khả năng cung cấp truy cập đến nguồn tài liệu số trực tuyến, dịch vụ này còn hỗ trợ mượn liên thư viện đối với các tài liệu in ấn. Đúng như tên gọi của dịch vụ, WorldCat Local kết nối người dùng chặt chẽ với chính thư viện của họ, tạo ra một kênh chuyển giao thông tin dựa trên những dịch vụ mà thư viện cung cấp, đơn giản hóa dịch vụ mượn tài liệu của thư viện. Cũng như các dịch vụ WSD khác, người dùng được trải nghiệm việc tìm kiếm thông tin thông qua một hộp tìm kiếm duy nhất và kết quả được phân loại, sắp xếp và đồng bộ hóa theo nhu cầu của người dùng [8].

Mỗi dịch vụ đều có những thế mạnh riêng của mình và không có dịch vụ nào là tối ưu cho từng thư viện cụ thể. Việc áp dụng các dịch vụ này phụ thuộc vào quy mô, đối tượng phục vụ và tiềm lực tài chính của mỗi thư viện.

Ngoài các dịch vụ tìm kiếm có bản quyền, hiện nay cũng có rất nhiều các giải pháp mã nguồn mở WSD dành cho thư viện. Có thể kể đến như: Blacklight, Fac-Back-OPAC, LibraryFind, Rapi, Scriblio, SOPAC và VuFind. Các giải pháp mã nguồn mở cũng tiếp cận phương pháp tìm kiếm tương tự như WSD bản thương mại. Điểm khác nhau giữa giải pháp mã nguồn và WSD thương mại là

các nhà cung cấp giải pháp thương mại thường cung cấp gói thông tin đi kèm (có sự thỏa thuận đối với nhà cung cấp thông tin), trong khi đó, nếu thư viện quyết định sử dụng mã nguồn mở, họ phải tùy biến mã nguồn này và xây dựng thỏa thuận với các nhà cung cấp dịch vụ thông tin.

3. Giải pháp cho các thư viện Việt Nam

3.1. Mua các giải pháp WSD từ nước ngoài

Giải pháp nhanh nhất cho các thư viện đại học Việt Nam đó là mua dịch vụ WSD của nước ngoài. Với kinh nghiệm triển khai lâu năm, tối ưu về mặt công nghệ và khả năng kết nối, có mối liên hệ rộng với các nhà cung cấp thông tin chuyên nghiệp, các giải pháp nước ngoài hiệu quả hơn về mặt tìm kiếm và mở rộng truy cập đến các nguồn tin chất lượng. Đối với các thư viện có tiềm lực thì đây là lựa chọn tối ưu.

Tuy nhiên, khó khăn của phần lớn của các thư viện đại học Việt Nam là kinh phí hoạt động hàng năm hạn chế. Do vậy, đầu tư một hệ thống WSD từ nước ngoài là thực sự ngoài tầm với của các thư viện. Với việc phải mua bản quyền theo năm, các dịch vụ này đòi hỏi các thư viện cần có tiềm lực tài chính mạnh để duy trì.

Một số chú ý khi quyết định đầu tư một hệ thống WSD đó là cần xem xét các tính năng của đánh chỉ mục tập trung và lớp khai thác. Công cụ đánh chỉ mục tập trung có giúp đưa ra toàn bộ những nguồn thông tin mà thư viện đang nắm giữ không? Các kết quả tìm kiếm có được xếp hạng và phù hợp với yêu cầu người dùng không? Hệ thống có khả năng tìm kiếm được các CSDL của các nhà cung cấp mà thư viện mua bản quyền không? Hệ thống có cung cấp khả năng tìm kiếm toàn văn cho các bài báo tạp chí không? Hệ thống có các mức tìm kiếm cơ bản hoặc nâng cao hay không, có hạn chế người dùng tìm kiếm nội dung mà họ

mong muốn? Giao diện người dùng có thân thiện, dễ sử dụng và điều khiển được không? Người dùng có khả năng nâng cao hiệu quả tìm kiếm thông qua các tiêu chí và các gợi ý của hệ thống? Những câu hỏi này cần được trả lời “có” mới có thể xem xét đầu tư một hệ thống WSD.

3.2. Tự phát triển WSD trong nước

Câu hỏi đặt ra là các công ty công nghệ và các thư viện Việt Nam có thể phát triển hệ thống WSD được không? Câu trả lời là “có”. Về mặt công nghệ, xây dựng hệ thống tìm kiếm WSD là hoàn toàn khả thi đối với năng lực hiện tại của các công ty công nghệ. Một số công ty đã bắt tay vào phát triển hệ thống tìm kiếm tập trung và đang chạy thử nghiệm và cho kết quả bước đầu. Ngoài việc tự phát triển công nghệ này, các công ty hay thư viện hoàn toàn có thể sử dụng mã nguồn mở để phát triển thêm hoặc tùy biến theo nhu cầu thực tế.

Vấn đề còn lại là ở chính các thư viện. Đó là sự mong muốn áp dụng cũng như thỏa thuận hợp tác trong việc chia sẻ nguồn thông tin. Đối với các nguồn tin phân tán trong nội bộ một thư viện (local resources), việc áp dụng WSD là dễ dàng và khả thi. Tuy nhiên, đối với nguồn tin từ xa (remote resources) thì việc tìm kiếm và khai thác là một trở ngại lớn. Để hệ thống tìm kiếm tập trung đạt hiệu quả thì nhất thiết phải có sự hợp tác giữa các thư viện. Đó là sự thỏa thuận trọng việc cho phép các máy tìm kiếm có thể truy cập và tìm kiếm trong các CSDL. Cao hơn nữa đó là thỏa thuận trong việc truy cập các nguồn tài nguyên số. Nếu mới chỉ dừng ở bước tìm kiếm, mà chưa thể lấy được tài liệu về sử dụng thì không phải là một hệ thống WSD hoàn chỉnh. Điều này lại phụ thuộc và sự thỏa thuận giữa các thư viện với nhau.

Một trở ngại của các giải pháp WSD phát

Nghiên cứu - Trao đổi

triển trong nước đó là khả năng khai thác các nguồn thông tin có bản quyền từ các nhà cung cấp thông tin chuyên nghiệp trên thế giới. Thường thì các nhà cung cấp dịch vụ WSD đều cạnh tranh nhau. Do vậy, để hợp tác với họ là khó vì họ vẫn đóng vai trò là nhà cung cấp thông tin cũng như giải pháp WSD. Đối với một số thư viện có khả năng mua được các CSDL trực tuyến, một trong những thỏa thuận quan trọng là cho phép WSD có thể tìm kiếm được thông tin trong CSDL mà thư viện mua bản quyền.

4. Kết luận

Với sự gia tăng không ngừng của tài nguyên số và xu thế khai thác thông tin trực tuyến đã và đang đặt ra cho các thư viện một bài toán khó đó là hỗ trợ người dùng tìm kiếm

thông tin nhanh, dễ dàng và hiệu quả trong một khối lượng thông tin khổng lồ. Sẽ không còn có cán bộ thư viện túc trực hỗ trợ bạn đọc trong tìm kiếm thông tin như khi bạn đọc đến trực tiếp thư viện, thay vào đó bạn đọc phải tự khai thác thông tin. Điều này nảy sinh nhu cầu là cần có một hệ thống tìm kiếm đủ thông minh để biết bạn đọc cần gì và tìm kiếm thông tin ở đâu. WSD ra đời với kỳ vọng đáp ứng được nhu cầu này. WSD là một dịch vụ hỗ trợ cao về công nghệ chi phí và nhân lực. Do vậy, các thư viện cần có sự chuẩn bị cho các yếu tố này để sẵn sàng xây dựng và triển khai dịch vụ WSD. Việc hợp tác giữa các thư viện để xây dựng dịch vụ WSD là một điều kiện tiên quyết rất quan trọng trong chiến lược phát triển chung của hệ thống thư viện đại học.

Tài liệu tham khảo

1. Burke, J. (2010). Discovery versus Disintermediation: the new reality driven by today's end-user. Retrieved from http://www.vala.org.au/vala2010/papers2010/VALA2010_57_Burke_Final.pdf.
2. Codeproject (2013). Basics of Single Sign on (SSO). Retrieved from <http://www.codeproject.com/Articles/429166/Basics-of-Single-Sign-on-SSO>.
3. Diamond, T.; Price, S. and Chandrasekar, R. (2013). Actions Speak Louder than Words: Analyzing large-scale query logs to improve the research experience. Code4Lib Journal. Retrieved from <http://journal.code4lib.org/articles/8693>.
4. Ebsco (2014). Ebsco discovery services. Retrieved from <http://www.ebscohost.com/discovery>.
5. Exlibris (2014). Primo - Empowering libraries to address user needs. Retrieved from <http://www.exlibrisgroup.com/category/PrimoOverview>.
6. Hoepfner, A. (2012). The ins and outs of evaluating web-scale discovery services. Computers in Libraries, 3(32), pp. 6-11.
7. OCABD - Orbis Cascade Alliance Board of Directors (2014). Alliance Board letter to EBSCO and Ex Libris. Retrieved from https://www.orbis-cascade.org/file_viewer.php?id=2205
8. Oclc (2014). WorldCat Local. Retrieved from <http://oclc.org/worldcat-local.en.html>.
9. ProQuest (2014). Discovery Services. Retrieved from <http://www.proquest.com/libraries/academic/discovery-services>.
10. Thompson, J. (2014). Implementing web-scale discovery services: A practical guide for librarians. London: Rowman & Littlefield Publishers
11. Vaughan, J. (2011). Investigations into library web scale discovery services. Information Technology and Libraries. Retrieved from http://digitalscholarship.unlv.edu/lib_articles/44.

(Ngày Tòa soạn nhận được bài: 08-1-2015; Ngày phản biện đánh giá 14-03-2015; Ngày chấp nhận đăng: 22-05-2015).

Đại học Quốc gia Hà Nội tài trợ cho nghiên cứu này. Mã số: QG.15.49