

MỘT SỐ CÔNG CỤ HỖ TRỢ TỰ ĐỘNG HÓA PHÂN LOẠI VÀ BIÊN MỤC TÀI LIỆU

ThS Đinh Thúy Quỳnh

Trường Đại học Văn hóa Hà Nội

Tóm tắt: Giới thiệu một số công cụ điện tử hỗ trợ công tác xử lý thông tin tại các thư viện và cơ quan thông tin nhằm giúp cán bộ thư viện rút ngắn thời gian xử lý đồng thời phát huy hiệu quả trong việc tự động hóa các khâu công tác thư viện.

Từ khóa: Phân loại tài liệu; công cụ tra cứu; tự động hóa; biên mục; xử lý thông tin

Automatic document classification and cataloguing tools

Abstract: The article introduces several e-tools to assist the information processing at libraries and information centres as well as to improve automatic library activities performance.

Keywords: Document classification; search tools; automatic; cataloguing; information processing.

Mở đầu

Xử lý thông tin luôn đóng một vai trò quan trọng trong hoạt động của các cơ quan thông tin-thư viện (TT-TV). Một nghiên cứu đã chỉ ra rằng: “Chỉ tính riêng về lĩnh vực khoa học tự nhiên và kỹ thuật, hành tinh chúng ta hàng năm cho ra đời khoảng năm vạn tạp chí với 4,5 triệu bài báo, 110 nghìn cuốn sách, 30 vạn bản mô tả sáng chế phát minh, 25 vạn báo cáo khoa học và thiết kế thử nghiệm, nửa triệu catalog công nghiệp,... Khối lượng các tài liệu này, cứ khoảng 10 đến 15 năm, lại tăng lên gấp đôi; với độ tăng ấy, cứ sau 1000 năm, tổng số ấn phẩm khoa học kỹ thuật lại tăng gấp 100 lần” [1, tr.93]. Sự bùng nổ này đã dẫn tới hiện tượng thông tin trùng lặp, lạc hậu, kém chất lượng, khó kiểm soát. Nếu các thư viện và cơ quan thông tin không tiến hành xử lý thông tin sẽ dẫn đến tình trạng “khủng hoảng thông tin”.

Xử lý thông tin chính là quá trình biến đổi thông tin từ dạng thức ban đầu thành

những dạng thức mới nhằm đáp ứng nhu cầu người dùng tin một cách nhanh chóng, chính xác và đầy đủ. Kết quả của quá trình xử lý thông tin sẽ tạo ra các sản phẩm và dịch vụ thông tin, là cơ sở để người dùng tin cũng như các cơ quan TT-TV có thể lưu trữ, quản lý và khai thác thông tin có hiệu quả. Chính vì vậy, xử lý thông tin luôn là một trong những quy trình nghiệp vụ được các thư viện và cơ quan thông tin quan tâm phát triển.

Đặc biệt, trong những năm trở lại đây, cùng với sự phát triển vượt bậc của KH&CN, nhiều thành tựu tiên tiến của công nghệ thông tin đã được ứng dụng vào lĩnh vực TT-TV nói chung và công tác xử lý thông tin nói riêng. Trong đó, phải kể đến một số công cụ điện tử đã được sử dụng để hỗ trợ cho công tác xử lý thông tin nhanh hơn và đạt kết quả chính xác hơn. Các công cụ này, cùng với thời gian, đã có những sự thay đổi về định dạng để phù hợp với nhu cầu ngày càng cao

của các cơ quan TT-TV: từ vi phim, cho đến CD-ROM và ngày nay là Web.

1. Web Dewey

Web Dewey là phiên bản Bảng phân loại thập phân Bách Khoa DDC điện tử do OCLC phát triển. Đây là phiên bản DDC truy cập qua web thay cho phương thức truy cập từ CD-ROM trước đây (Địa chỉ truy cập: <http://dewey.org/webdewey/login/login.html>). Điều này đã tạo điều kiện thuận lợi cho người dùng trong việc truy cập và sử dụng dịch vụ. Phiên bản ban đầu của Web Dewey được OCLC phát hành vào năm 2003, cùng thời điểm với việc công bố bản in DDC 22. Đến năm 2010, khi bản in DDC 23 ra mắt, Web Dewey đã được nâng cấp lên Web Dewey 2.0. Về cơ bản, Web Dewey được coi là bản DDC đầy đủ dưới dạng tư liệu điện tử.

Phiên bản Web Dewey 2.0 hiện nay đang là công cụ hỗ trợ rất đắc lực cho các cán bộ chuyên trách phân loại tài liệu tại các cơ quan TT-TV. Web Dewey 2.0 có những tính

năng nổi bật:

- Về nội dung:

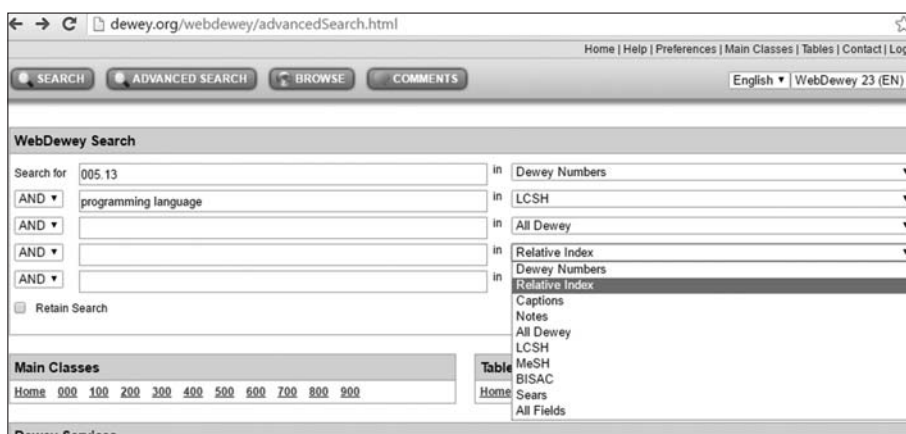
Chứa toàn bộ các ký hiệu phân loại của DDC 23 (trong đó bao gồm cả những phát triển mới, những chỉ số mới và các thuật ngữ chỉ mục điện tử phụ trợ). Các nội dung này liên tục được cập nhật.

- Về giao diện:

Sử dụng giao diện Web 2.0 thân thiện, dễ sử dụng. Đặc biệt, người dùng còn có thể tùy biến màn hình hiển thị bằng một nhấp chuột duy nhất.

- Về chức năng:

- Sử dụng nhiều cách thức tìm kiếm: tìm lướt (browse), tìm thông thường (search), tìm nâng cao (advanced search). Tùy theo yêu cầu tìm mà người dùng có thể tìm kiếm theo một yếu tố nhất định như: theo chỉ số phân loại, đề mục chủ đề, chỉ mục quan hệ,... hoặc cùng một lúc kết hợp nhiều yếu tố thông qua việc sử dụng các toán tử Boolean: AND, OR, NOT (Hình 1).



Hình 1. Tìm kiếm kết hợp nhiều yếu tố

- Cho phép ánh xạ các ký hiệu của DDC với các đề mục chủ đề của Thư viện Quốc hội Hoa Kỳ LCSH (Library of Congress Subject Headings), đề mục chủ đề y học- MeSH (Medical Subject Headings), đề mục chủ đề

BISAC, SEARS,... Người dùng có thể tra cứu cách thức sử dụng các đề mục chủ đề tại liên kết ký hiệu của DDC. Các đề mục chủ đề này được mô tả theo cấu trúc biểu ghi MARC (Hình 2). Các dữ liệu được tổ chức theo

```

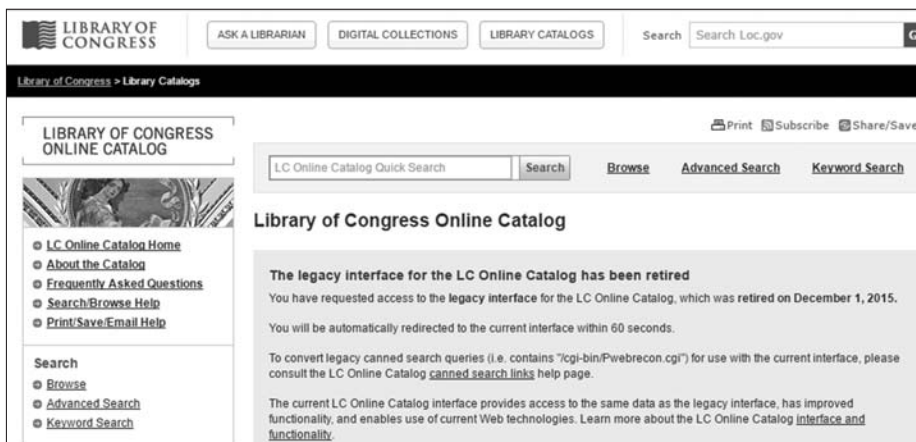
ocd00116847
OCoLC-D
20160120163650.0
100204aaaaaaaa
## OCLCD $b eng $c OCLCD
0# ddc $c 23 $e eng
## 005.1 $e 005.1-005.5 $g * $j Programming
1# $i Class here $t application frameworks, application programming, computer algorithms, integrated development environments, software engineer
2# $i Class $t algorithms discussed solely from a theoretical perspective, without regard to computer implementation $i , in $a 518.1
2# $i Class $t a specific application of programming within computer science $i with the application in $a 005.4 $c 005.8 $i or $a 006 $i , e.g., $t evolu
programming $e 006.3823 $i , $t programming of computer graphics $e 006.66
0# $i For $t programming for specific types of computers, for specific operating systems, for specific user interfaces $i , see $a 005.2
2# $i See Manual at $a 005.1-005.2 vs. 005.42
2# $i See Manual at $a 005.1 vs. 005.3
    
```

Hình 2. Biểu ghi mô tả theo cấu trúc MARC

cấu trúc này sẽ giúp cho việc trao đổi dữ liệu giữa các hệ thống được thuận tiện, chính xác.

- Có chức năng liên kết tới mục lục tra cứu

trực tuyến (Link to OPAC). Chức năng này giúp cán bộ biên mục có thể tham khảo các biểu ghi của Thư viện Quốc hội Hoa Kỳ có liên quan đến chủ đề tài liệu đang xử lý (Hình 3).



Hình 3. Kết nối với TVQH Hoa Kỳ thông qua chức năng Link to OPAC

Web Dewey được OCLC phát triển và liên tục cập nhật, chỉnh sửa các tính năng. Điều này đã giúp cho người sử dụng có được công cụ mạnh để tối ưu hóa công tác phân loại tài liệu. Để sử dụng phiên bản Web Dewey, người sử dụng phải mua quyền truy cập từ nhà cung cấp OCLC. Chi phí mua bản quyền sử dụng tùy thuộc vào số lượng người sử dụng được đăng ký. Các cá nhân và tổ chức có thể đăng ký dùng thử trong vòng 30 ngày theo địa chỉ: <https://www.oclc.org/forms/webdewey-free-trial.en.html>

2. PowerCat

PowerCat là một sản phẩm do Công ty phần mềm Hiện đại thiết kế. PowerCat ứng dụng công nghệ tìm kiếm đa luồng để truy cập và tải về đầy đủ thông tin của tài liệu từ nhiều CSDL thư viện trên thế giới. Đây là một công cụ hữu ích không chỉ cho các cán bộ biên mục mà còn hữu ích đối với cả cán bộ bổ sung bởi các tính năng nổi bật [6]:

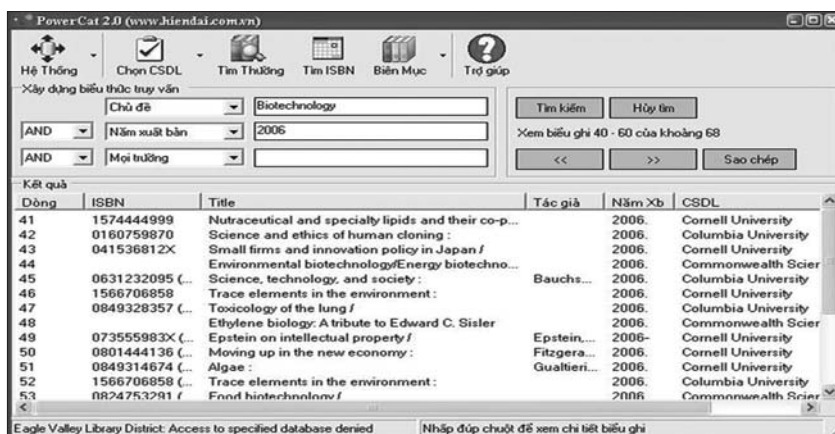
- Tìm kiếm cùng một lúc trên nhiều CSDL: Với việc sử dụng công nghệ tìm

kiểm đa luồng, PowerCat cho phép tìm kiếm nhanh chóng trên danh mục lập sẵn của gần 1000 CSDL thư viện trên thế giới trong cùng một lệnh tìm kiếm (Hình 4).



Hình 4. Danh mục các CSDL thư viện trong liên kết

- Sử dụng nhiều phương thức tìm kiếm: kết hợp các yếu tố với nhau (tối đa là ba yếu tố). Với phương thức tìm kiếm này, người sử dụng có thể dễ dàng tìm kiếm và tải về các biểu ghi thư mục cần thiết (Hình 5).
- Tìm thông thường: Cũng giống như Web Dewey, PowerCat cho phép người sử dụng có thể tìm kiếm theo một yếu tố hoặc

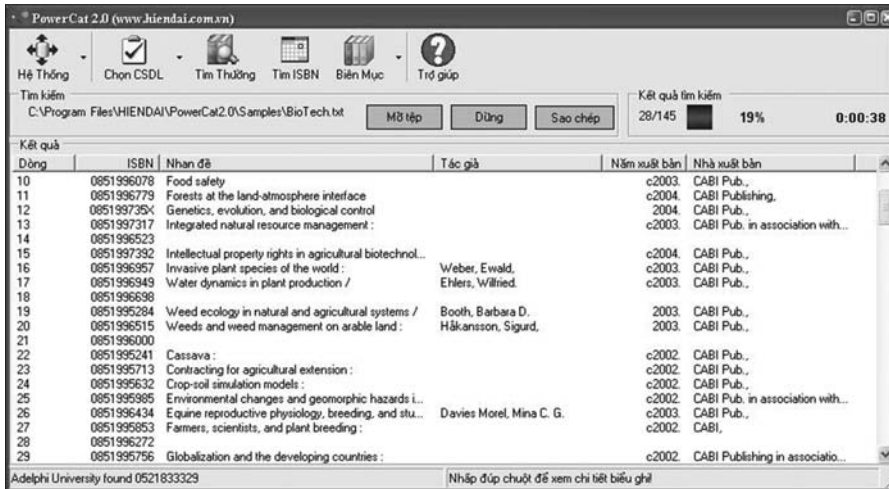


Hình 5. Kết quả tìm theo cách thông thường

- Tìm theo ISBN: Đây là phương thức tìm kiếm tiện lợi và hiệu quả. Bởi thông thường, khi muốn tìm kiếm thông tin cho một danh mục tài liệu, người dùng sẽ phải tìm kiếm thông tin của từng tài liệu trong CSDL rồi sau đó mới lập thành danh mục. PowerCat cho phép tìm kiếm và tải về nhanh chóng danh mục các biểu ghi thư mục theo danh sách ISBN được đưa vào, hay nói cách khác là tìm tài liệu theo danh
- mục ISBN. Phương thức này giúp cán bộ biên mục có thể tiết kiệm được nhiều thời gian trong quá trình tìm kiếm (Hình 6).

- Dễ dàng hiệu chỉnh biểu ghi theo tiêu chuẩn MARC: PowerCat cho phép cán bộ xử lý có thể dễ dàng xem và hiệu chỉnh biểu ghi tải về theo tiêu chuẩn MARC trước khi đưa vào CSDL của thư viện mình (Hình 7).

- Xuất dữ liệu tùy biến: PowerCat hỗ trợ các tiêu chuẩn MARC, tạo thuận lợi



Hình 6. Kết quả tìm theo danh mục ISBN

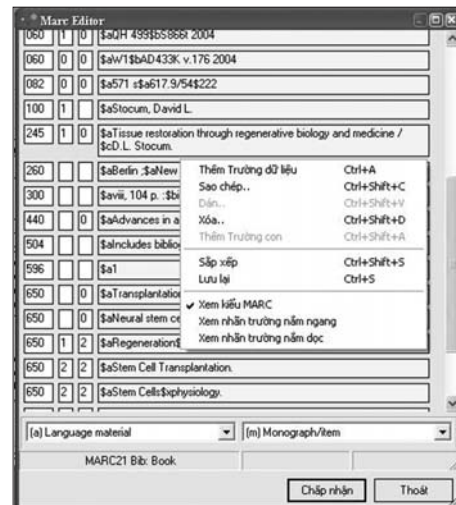
cho việc trao đổi dữ liệu giữa các thư viện như ISO2709, MARCXML,... Các thông tin được kết xuất theo đúng định dạng tiêu chuẩn quốc tế. Ngoài ra, PowerCat có thể xuất ra Excel các thông tin tùy biến theo cấu hình, không giới hạn số trường dữ liệu được chọn.

- Tra giá tự động từ Amazon: Ngoài các tính năng kể trên, PowerCat còn tích hợp với dịch vụ của nhà cung cấp tài liệu trực tuyến hàng đầu thế giới Amazon cho phép tải về các thông tin thương mại theo danh mục tài liệu có đánh chỉ số ISBN như: giá cả, các thông tin thư mục về tài liệu, thông tin nhà cung cấp,... Tính năng này đã giúp cho không chỉ cán bộ biên mục mà cả cán bộ bổ sung có thể dễ dàng tra cứu và lập danh mục tài liệu cần thiết.

- Hỗ trợ biên mục độc lập (biên mục ngoại tuyến): PowerCat có thể làm việc như là một công cụ biên mục độc lập. Trong trường hợp cần thiết, PowerCat có thể làm việc không cần kết nối với máy chủ của thư viện.

3. Mercury Z39.50 Client

Mercury Z39.50 Client là một phần mềm miễn phí sử dụng công nghệ tìm kiếm đa luồng. Do vậy, đối với những thư



Hình 7. Chức năng hiệu chỉnh biểu ghi tải về

viện không có điều kiện cài đặt các phần mềm thương mại đều có thể tự tải và cài đặt Mercury Z39.50 Client để sử dụng tại thư viện của mình (Địa chỉ tải ứng dụng: <http://www.basedowtech.com/projects/mzc/download>). Mercury Z39.50 Client cũng có những tính năng cần thiết để giúp người dùng có thể tìm kiếm thông tin một cách thuận tiện:

- Tìm kiếm cùng một lúc trên nhiều CSDL khác nhau: Cũng giống như PowerCat, Mercury Z39.50 Client sử dụng công nghệ tìm kiếm đa luồng, cho phép

người dùng có thể tìm kiếm thông tin nhanh chóng từ nhiều CSDL thư viện trên thế giới trong cùng một lệnh tìm kiếm. Phần mềm này cũng hỗ trợ việc cài đặt thêm hoặc xóa bỏ bớt CSDL thư viện trong danh mục sẵn có của mình (Hình 8).

- Sử dụng nhiều cách tìm kiếm:

- Tìm đơn giản: Người dùng có thể tìm thông tin mình cần thông qua hai yếu tố cơ bản là Nhan đề và Tên tác giả.

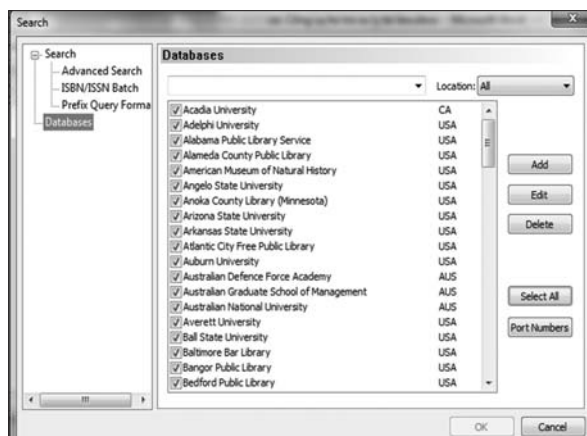
- Tìm nâng cao: Mercury Z39.50 Client cho phép tìm kiếm thông tin một cách chính xác bằng việc sử dụng kết hợp nhiều yếu tố với nhau (tối đa bốn yếu tố) như: Nhan đề, Tác giả, Chủ đề,... thông qua toán tử AND, OR và NOT (Hình 9).

- Tìm theo chỉ số ISBN/ISSN: Đối với các sách, tạp chí nước ngoài, cán bộ xử lý có thể tìm thông tin chính xác về tài liệu mình cần thông qua chỉ số ISBN/ISSN. Tuy nhiên, khác với PowerCat, Mercury Z39.50 Client cho phép mỗi lệnh tìm kiếm chỉ thực hiện được theo một chỉ số ISBN/ISSN nhất định.

- Xuất dữ liệu ra nhiều định dạng: Mercury Z39.50 Client cho phép xuất dữ liệu của các biểu ghi theo nhiều định dạng khác nhau như: Text, XML, MARC,... Tùy theo yêu cầu, thư viện có thể lựa chọn các định dạng phù hợp (Hình 10).

4. Hệ thống tra cứu tự động Đề mục chủ đề- LCSH- LCC- DDC

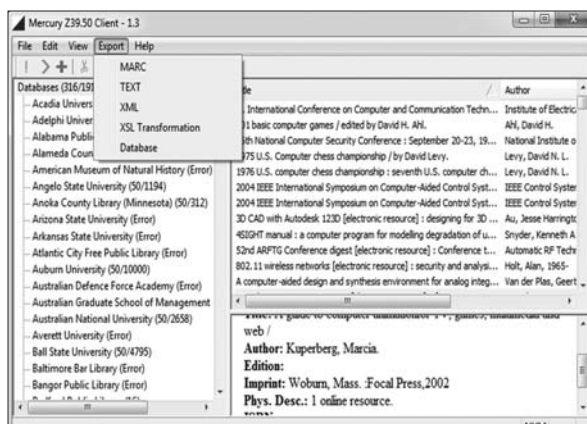
Đây là sản phẩm do nhóm cán bộ thư viện và công nghệ thông tin thuộc Thư viện Tạ Quang Bửu- Trường Đại học Bách Khoa Hà Nội tiến hành nghiên cứu và xây dựng vào năm 2008. Hệ thống này gồm khoảng 30.000 đề mục cho phép cán bộ xử lý có thể tra cứu chéo giữa hai ngôn ngữ tìm tin là đề mục chủ đề và ký hiệu phân loại, trong



Hình 8. Các CSDL cài sẵn trong Mercury Z39.50 Client



Hình 9. Tìm kiếm theo cách tìm nâng cao



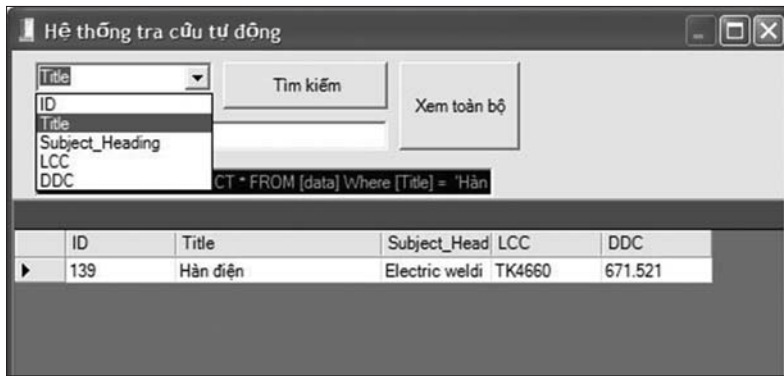
Hình 10. Các định dạng xuất dữ liệu

đó bao gồm: đề mục chủ đề LCSH và đề mục chủ đề dịch sang tiếng Việt, ký hiệu phân loại LCC và DDC.

Với tính năng tham chiếu chéo, hệ thống cho phép cán bộ xử lý có thể tìm được thông tin của ba yếu tố còn lại khi có trước một thông tin về tài liệu như: ký hiệu phân loại LC, ký hiệu phân loại DDC, đề mục chủ đề LCSH, đề mục chủ đề bằng tiếng Việt. Hay nói một cách khác, cán bộ

xử lý tài liệu có thể dễ dàng tham chiếu tới ba ngôn ngữ còn lại khi có thông tin về tài liệu bằng một trong bốn ngôn ngữ trên.

Hệ thống tra cứu tự động Đề mục chủ đề- LCSH-LCC-DDC là công cụ tra cứu rất hữu ích không chỉ cho Thư viện Tạ Quang Bửu- Trường Đại học Bách Khoa Hà Nội mà còn cho rất nhiều thư viện khác có nhu cầu tự động hóa trong công tác phân loại tài liệu (Hình 11).



Hình 11. Giao diện tìm kiếm của hệ thống tra cứu tự động

Kết luận

Như vậy có thể thấy, hiện nay có khá nhiều công cụ hữu hiệu hỗ trợ cho việc tự động hóa công tác xử lý thông tin tại các thư viện và cơ quan thông tin. Các công cụ này không chỉ giúp cho cán bộ xử lý thông tin nhanh hơn mà còn đem lại kết quả chính xác hơn. Tùy vào đặc điểm và điều kiện kinh phí mà mỗi thư viện có thể lựa chọn cho mình công cụ hỗ trợ tự động hóa phân loại và biên mục phù hợp và hiệu quả.

TÀI LIỆU THAM KHẢO

1. Nguyễn Hữu Hùng (2005). *Thông tin: Từ lý luận đến thực tiễn*, Nxb Văn hóa thông tin, Hà Nội.
2. Trần Thị Quý, Đỗ Văn Hùng (2007). *Tự động hóa trong hoạt động TT-TV*, Nxb Đại học Quốc gia Hà Nội, Hà Nội.

3. Nguyễn Văn Thiên (2008). *Xây dựng hệ thống tra cứu tự động bốn ngôn ngữ LCC-DDC- LCSH-Đề mục chủ đề*, Đề tài nghiên cứu khoa học cấp trường, Trường Đại học Bách Khoa Hà Nội, Hà Nội.

4. <http://www.basedowtech.com/projects/mzc>
5. <http://dewey.org/webdewey/login/login.html>.
6. <http://hiendai.com.vn/JetSolutions.aspx?mnuid=144>
7. http://lrc.tnu.edu.vn/upload/collection/brief/40299_13122013931272201238.pdf
8. <https://www.oclc.org/forms/webdewey-free-trial.en.html>

(Ngày Tòa soạn nhận được bài: 05-4-2016; Ngày phản biện đánh giá: 02-6-2016; Ngày chấp nhận đăng: 02-7-2016).