

# KHAI PHÁ DỮ LIỆU TRONG THƯ VIỆN SỐ

TS Đậu Mạnh Hoàn

Trung tâm Học liệu, Trường Đại học Quảng Bình

**Tóm tắt:** Dựa trên những đặc trưng của kỹ thuật khai phá dữ liệu và tổ chức thư viện số, bài viết trình bày các khả năng khai thác của kỹ thuật khai phá dữ liệu trong thư viện số sau đó đề xuất các áp dụng của kỹ thuật khai phá dữ liệu trong thư viện số trên khía cạnh: Cải thiện tốc độ; nâng cấp chất lượng dịch vụ thông tin của thư viện; hỗ trợ các quyết định của thư viện; dịch vụ thông tin cá nhân; tự động hóa xử lý thông tin; hỗ trợ các công việc khác nhằm nâng cao hiệu quả sử dụng cũng như khai thác thư viện số.

**Từ khóa:** Khai phá dữ liệu; thư viện số.

## Data mining in digital library

**Abstract:** Based on the characteristics of data mining and digital library, the article discusses some benefits and possibilities of data mining in digital library, including: speed improvement; information service improvement; decision - making support; personal information service support; information processing automation, and other activities to improve the performance and utility of digital library.

**Keywords:** Data mining; digital library.

## 1. Giới thiệu

Khai phá dữ liệu đã trở thành lĩnh vực nghiên cứu nóng, thu hút sự quan tâm của nhiều người, nhiều tổ chức trong nước và thế giới. Mục đích của khai phá dữ liệu là tìm ra mô hình phân loại hữu ích trong các dữ liệu phục vụ cho mục đích dự đoán hoặc mô tả và khai thác thông tin trong các lĩnh vực khác nhau. Sự ra đời và phát triển của khai phá dữ liệu đã đáp ứng và giải quyết được nhiều vấn đề ứng dụng trong đời sống, mang lại hiệu quả cao. Khai phá dữ liệu là kỹ thuật có nhiều ưu việt, được sử dụng rộng rãi trong các lĩnh vực và cũng được ứng dụng rộng rãi trong không gian của thư viện số. Trên phương diện xây dựng thông tin kỹ thuật số, tối ưu hóa xử lý thông tin tự động, chất lượng dịch vụ thông tin và phát triển kinh doanh, vai trò của khai phá dữ liệu trong tự động hóa thư viện kỹ thuật số là rất quan trọng.

Thư viện số là một nguồn tài nguyên thông tin số vô cùng to lớn trong đó có các phương tiện truyền thông với nhiều nội dung đa dạng khác nhau. Nó còn là một hệ thống thông tin kỹ thuật số được ra đời bởi sự hỗ trợ của nhiều công nghệ cao và hiện đại, là thế hệ tiếp theo của phương thức quản lý tài nguyên thông tin Internet, là một loại cơ chế dịch vụ thuận tiện cung cấp thông tin cho độc giả. Thư viện kỹ thuật số đã giải quyết vấn đề căn bản hiện nay trên Internet, đó là giải quyết được sự phân tán, thiếu đồng nhất và bất tiện trong sử dụng khi khai phá tài nguyên số. Bằng cách áp dụng các kỹ thuật mới, thư viện số không chỉ thay đổi các dịch vụ thư viện truyền thống, phương tiện truyền thông và phương pháp quản lý, mà nó cũng sử dụng các nguồn tài nguyên kỹ thuật số một cách phong phú, đồng thời khai phá lợi thế các kỹ thuật đó để phát triển các ý tưởng tạo ra công cụ mới, tăng cường các chức năng của thư

viện số, cung cấp các dịch vụ cho giáo dục, các ứng dụng trong đời sống và trong nghiên cứu khoa học. Trong tương lai, thương mại điện tử sẽ trở thành dịch vụ cốt lõi của thế giới số, điều này làm tăng trưởng lợi điểm của các thư viện số và cũng là động lực để thư viện số ngày càng phát triển.

## **2. Khai phá dữ liệu và thư viện số**

### **2.1. Khai phá dữ liệu**

Khai phá dữ liệu (Data mining) là một khái niệm bao hàm nhiều kỹ thuật nhằm phát hiện ra các thông tin có giá trị tiềm ẩn trong các tập dữ liệu lớn (các kho dữ liệu). Về bản chất, khai phá dữ liệu liên quan đến việc phân tích các dữ liệu và sử dụng các kỹ thuật để tìm ra các mẫu hình có tính chính quy trong kho dữ liệu lưu trữ [1, 2].

Khai phá dữ liệu là lĩnh vực nghiên cứu trong khoa học máy tính nói chung, trong trí tuệ nhân tạo, xử lý tri thức thông minh nói riêng. Khai phá dữ liệu là bước chính của quy trình khai phá tri thức trong CSDL (Knowledge Discovery in Database- KDD). Nó đề cập đến việc khai phá hoặc dữ liệu hoặc tri thức từ một số lượng lớn các dữ liệu và thông tin [3]. Khai phá dữ liệu được ứng dụng trong bất kỳ loại hình cung cấp dịch vụ lưu trữ thông tin nào, chẳng hạn như: kho dữ liệu, CSDL quan hệ, CSDL giao dịch, CSDL hướng đối tượng, CSDL quan hệ hướng đối tượng, CSDL không gian, CSDL thời gian, CSDL văn bản, CSDL đa phương tiện, CSDL Web, CSDL tri thức. Đặc biệt, sự thành công của hướng nghiên cứu khai phá văn bản đã góp phần to lớn trong sự phát triển của thư viện số. Kỹ thuật mới này được sử dụng để dự đoán các thông tin tiềm ẩn từ các CSDL có quy mô lớn hoặc từ kho dữ liệu, khai phá các phương thức tiềm năng, tự động dự đoán tri thức và hành vi, và tự động khám phá các mẫu trước đó chưa biết. Các mẫu chiết xuất được từ khai phá dữ liệu tri thức có thể được thể hiện như khái niệm, quy tắc, các mẫu và tri thức.

Quá trình khai phá dữ liệu có thể được chia thành 3 giai đoạn: giai đoạn chuẩn bị dữ liệu, giai đoạn khai phá tri thức và giai đoạn trình bày và thể hiện kết quả [4, 5]. Quá trình khai phá tri thức được lặp đi lặp lại với sự tham gia của người sử dụng. Có sự khác biệt nhất định giữa khai phá dữ liệu và khai phá tri thức. Nói chung, khai phá dữ liệu là một bước cụ thể trong quá trình khai phá tri thức. Nó sử dụng các thuật toán đặc biệt để trích xuất các mô hình từ CSDL, từ đó thông qua quá trình giải thích của hệ thống và việc đánh giá của các mô đun, chuyển đổi vào hệ thống tri thức mà người sử dụng có thể hiểu được. Tuy nhiên, một cách tổng quát, khai phá tri thức thường được coi là quá trình bao gồm: chuẩn bị dữ liệu, trích xuất mô hình, biểu diễn tri thức và các bước khác. Khai phá dữ liệu và truy vấn thông tin có sự khác biệt nhất định, điều đó được thể hiện ở chỗ: truy vấn thông tin xác định bởi mục tiêu, nhu cầu của người sử dụng và cần phải đặt ra một cách rõ ràng trong yêu cầu truy vấn, còn khai phá dữ liệu là ngẫu nhiên, kết quả của nó là không thể đoán trước được. Mục tiêu của truy vấn thông tin là để giúp người dùng tìm ra các tài liệu đáp ứng các yêu cầu truy vấn từ một số lượng lớn các tài liệu, còn khai phá dữ liệu là để khám phá những kiến thức tiềm ẩn trong các tài liệu. Tuy nhiên, hai kỹ thuật này là bổ sung cho nhau. Có thể sử dụng những kết quả nghiên cứu của khai phá dữ liệu để cải thiện độ chính xác và hiệu quả của truy hồi thông tin và cải thiện sự biểu hiện của kết quả tìm kiếm, do đó hệ thống truy hồi thông tin được phát triển lên một tầm cao mới. OLAP (On-line analytical processing- OLAP) là một công cụ phân tích, nhưng cả hai đều có sự khác biệt rõ ràng. Công nghệ phân tích xử lý trực tuyến không phải là một quá trình tự động. Khi người dùng đặt câu hỏi, hệ thống phân tích xử lý trực tuyến có trách nhiệm trích xuất thông tin chi tiết về vấn đề này. Quá trình khai phá dữ liệu thường là tự động, người dùng không cần phải trình bày các vấn đề chính xác, đây chính là ưu điểm trong quá

trình khai phá những kiến thức ẩn.

Khai phá dữ liệu là một quá trình hoàn chỉnh, từ việc xác định vùng đối tượng, chuẩn bị dữ liệu, tiến trình khai phá và trình bày đến phân tích kết quả [6, 7].

## 2.2. Thư viện số

Hiện nay, có rất nhiều tổ chức nghiên cứu trong và ngoài nước đã xây dựng các hệ thống thư viện số với quy mô lớn, các hệ thống này phù hợp với mô hình tương tác dịch vụ của thư viện và hệ thống dịch vụ cá nhân. Nó dựa trên lợi ích cá nhân và nhu cầu học tập của người dùng thông qua các trang web để thu thập, tổ chức và liên kết các vùng tài nguyên kỹ thuật số mà cá nhân quan tâm đến. Khi người dùng truy cập một lần nữa hệ thống thư viện số, họ có thể nhận nội dung mới nhất liên quan đến yêu cầu của họ. Dịch vụ này hoạt động như sau, đầu tiên cán bộ thư viện chuyên nghiệp tạo ra một danh sách tài nguyên thư viện trực tuyến kỹ thuật số hoặc loại tài nguyên theo chủ đề cho người sử dụng, tiếp theo hệ thống cung cấp cho người dùng một tài khoản đăng nhập, người dùng sử dụng tài khoản đăng nhập để vào trang web của thư viện, nơi sẽ cung cấp một danh sách các yêu cầu cần thiết để lựa chọn tài nguyên của mình cũng như các nguồn tài nguyên Web khác, nó được bổ sung vào hệ thống thư viện. Ngoài ra, hệ thống thư viện cũng cung cấp các thông tin liên lạc mới nhất, thư mục cá nhân và các chức năng đánh dấu, các dịch vụ khác, người dùng cũng có thể tùy chỉnh các dịch vụ mà họ yêu cầu. Hệ thống dựa trên thông tin đăng ký của người dùng và nội dung cho mỗi người dùng để tạo ra một tập tin chính sách, bao gồm cả thông tin cá nhân của người sử dụng và các nguồn lực kỹ thuật số trên đại diện cho người dùng lựa chọn một danh sách các thông số. Tập tin này được lưu trữ dưới các hình thức trong ổ cứng máy tính hoặc máy chủ cơ sở dữ liệu của người sử dụng, khi người dùng truy cập một trang, hệ thống thư viện sẽ được trích xuất, và theo đó tùy chọn mới nhất được bổ

sung vào nội dung [8].

Trên thế giới, thư viện điện tử đã được áp dụng rất sớm. Các hệ thống tùy biến tài nguyên, tùy biến trang, các nguồn tài nguyên mới nhất, quản lý cá nhân, bộ sưu tập liên kết cá nhân, tổ chức thư viện và chức năng hỗ trợ truy vấn đã được áp dụng. Thư viện số trong tương lai chắc chắn sẽ phát triển nhanh chóng. Việc áp dụng công nghệ khai phá dữ liệu trên các nguồn thông tin rộng lớn sẽ là một sự lựa chọn lớn của các công cụ khai phá tri thức và các thuật toán, cá nhân hoá dịch vụ thư viện số trở thành một phần không thể thiếu trong xây dựng hỗ trợ kỹ thuật [9].

## 3. Khả năng khai phá dữ liệu trong thư viện số

Dựa vào cấu trúc và nội dung của thư viện số chúng tôi đề xuất các khả năng khai thác của khai phá dữ liệu trong thư viện số như sau:

### 3.1. Khai phá cấu trúc thư viện số

Thư viện số được thiết kế trên cấu trúc các trang web, nó sử dụng các ngôn ngữ thiết kế web cùng với các siêu liên kết để tổ chức thông tin. Trên cơ sở đó, thông qua các siêu liên kết và tổ chức của trang, các kết nối, các thư mục, nội dung mà chúng liên kết đến chúng ta có thể khám phá ra các kiến thức mới và bổ ích. Các kỹ thuật khai phá trang web (web mining) được khai thác một cách triệt để để thu được các thông tin mới và ý nghĩa nhất.

### 3.2. Khai phá người sử dụng thư viện số

Khi người sử dụng khai thác tài nguyên trên thư viện số, một phiên giao dịch sẽ ghi lại tất cả các lần người sử dụng trình duyệt web theo thời gian để hình thành cơ sở dữ liệu giao dịch, kết quả là chúng ta có thể thu thập và lưu trữ lại các thông tin của người sử dụng như là các bộ sưu tập đặc biệt thông qua chế độ duyệt web, từ đó sử dụng các kỹ thuật để khai phá thông tin. Sử dụng các thuật toán khai phá luật kết hợp để tìm các giao dịch tập hợp có tần số truy cập vượt quá

một ngưỡng nhất định, sau đó sử dụng kết quả này để phân loại dữ liệu [10]. Trên cơ sở đó tìm ra bộ tính năng trong cơ sở dữ liệu giao dịch mà được truy cập thường xuyên, và lưu trữ chúng trong cơ sở kiến thức của máy chủ bằng cách sử dụng luật kết hợp của các phân tích liên kết. Khi người dùng duyệt một trang, các đại lý mạng trước đây kết nối các trang có liên quan theo quy định, để cải thiện tốc độ phản ứng. Sử dụng khai phá web để có được mô hình chuỗi các truy cập của người dùng trước đó và thực hiện truyền các trang người dùng có thể đọc theo dự đoán.

### **3.3. Khai phá nội dung trong thư viện số**

Dựa trên nội dung các trang web, nội dung có trong thư viện số bao gồm: văn bản có cấu trúc, văn bản phi cấu trúc, các loại văn bản, các bảng, dữ liệu đa phương tiện, âm thanh, ảnh,... Có thể khai phá nội dung từ thư viện kỹ thuật số thông qua các hình thức sau:

- Sử dụng kỹ thuật tóm tắt văn bản để khai phá các tóm tắt (abstract) từ các file dữ liệu. Đây là phần nội dung quan trọng và là trọng tâm của mỗi tài liệu, nó phản ảnh nội dung chính của tài liệu đó. Tóm tắt văn bản là kỹ thuật vô cùng hữu ích trong quá trình tìm hiểu xem liệu một văn bản dài có thể đáp ứng nhu cầu của người sử dụng và có nội dung đáng đọc để biết thêm thông tin hay không. Với văn bản lớn, kỹ thuật tóm tắt tài liệu thường sẽ được thực hiện và tóm tắt trong thời gian người sử dụng tiến hành đọc đoạn đầu tiên. Mục đích cốt lõi của việc tóm tắt là để giảm độ dài và chi tiết của một tài liệu trong khi vẫn giữ được điểm chính của nó và ý nghĩa tổng thể.

- Phân loại văn bản: Tự động phân loại văn bản trên cơ sở tài liệu người dùng, kết quả phân loại sẽ phục vụ các tìm kiếm và khai thác của người sử dụng. Kết quả phân loại còn là tiền đề để giải quyết một lớp bài toán phục vụ các thao tác trên văn bản.

- Phân cụm là kỹ thuật được sử dụng để nhóm các tài liệu tương tự nhau trong tập dữ

liệu vào các cụm sao cho các đối tượng cùng cụm là tương đồng. Kết quả phân cụm sẽ giúp cho người sử dụng xác định được các tài liệu tương tự hay có cùng nhóm nội dung.

- Dự đoán và đánh giá đó là tìm ra những thông tin mới, những quyết định mới từ những dữ liệu đã có thông qua quá trình quan sát và xử lý. Kết quả của dự đoán là đưa ra hậu quả của một tình huống nào đó, xác định các lỗi, và định giá chúng từ đó sử dụng để phân tích và đánh giá dữ liệu.

## **4. Áp dụng kỹ thuật khai phá dữ liệu cho thư viện số**

### **4.1. Nâng cao tốc độ**

Nâng cao tốc độ trong mọi hoạt động của thư viện và dịch vụ người dùng là mục tiêu mà thư viện số hướng tới. Nâng cao tốc độ nhằm đáp ứng thời gian truy cập các thông tin cần thiết cho người sử dụng, đáp ứng khả năng trao đổi thông tin, truy xuất thông tin, khai thác các dịch vụ một cách hiệu quả nhất.

### **4.2. Nâng cấp chất lượng dịch vụ thông tin**

Sử dụng các kỹ thuật để xây dựng thư viện phát triển theo hướng phần mềm thông minh, bao gồm dịch vụ truy vấn thông tin. Cải tiến công cụ phục hồi truyền thống thân thiện, dễ hiểu và tương tác theo kết quả. Tìm kiếm thông tin thông minh không chỉ hỗ trợ việc thu hồi khái niệm, tìm kiếm mờ, kết hợp thu hồi và phục hồi đa ngôn ngữ, mà còn có thể nhanh chóng sử dụng các thuật toán phân cụm, phân tích kết quả truy vấn, để thuận tiện cho việc lựa chọn của người sử dụng, và cùng một lúc xác định tìm kiếm thêm trên cơ sở này. Ngoài ra, có thể sử dụng các thuật toán mô hình quan tâm đến đánh giá của người sử dụng tiềm năng, thực hiện phân tích thông tin trực tuyến theo thời gian thực. Trực tiếp phân tích các dữ liệu văn bản và các mối quan hệ giữa chúng, để xác định những kiến thức chưa biết và hữu ích của quá trình.

### **4.3. Hỗ trợ các quyết định của thư viện**

Mức độ quản lý thấp là một trong những



yếu tố cơ bản ảnh hưởng đến sự phát triển của các thư viện. Trước đây, việc đưa ra quyết định của thư viện chủ yếu dựa vào kinh nghiệm, điều này là chủ quan, một chiều, thiếu thông tin, và không thể đáp ứng yêu cầu của thời đại. Công nghệ khai phá dữ liệu có thể cung cấp thông tin bảo đảm cho việc ra quyết định của lãnh đạo quản lý thư viện, cụ thể:

- Khai phá dữ liệu có thể cùng một lúc thu thập dữ liệu nội bộ và thông tin bên ngoài có liên quan đến hệ thống thông tin của thư viện, và sau khi xử lý, chuyển đổi, tạo thành các thông tin tập trung, thống nhất và có sẵn, để tránh việc đưa ra quyết định sai lầm do thiếu thông tin.

- Sử dụng các công cụ hệ thống OLAP kho dữ liệu để so sánh với việc tích hợp các dữ liệu đa chiều, xem xét và xác minh giả thiết của quyết định chính sách, để nâng cao tính khả thi và độ tin cậy của các quyết định, và sử dụng hợp lý các nguồn tài nguyên hạn chế, đồng thời tối ưu hóa phân bổ nguồn lực vào thư viện.

- Sử dụng các công cụ khai phá dữ liệu để tìm ra một mô hình tiềm ẩn từ các dữ liệu lịch sử và dự báo tự động trên cơ sở của mô hình.

#### **4.4. Cung cấp dịch vụ thông tin cá nhân**

Việc áp dụng công nghệ khai phá dữ liệu làm cho các dịch vụ thông tin của thư viện hoạt động tốt hơn, giúp nâng cao hiệu quả của các dịch vụ thông tin và thư viện. Sử dụng công nghệ khai phá dữ liệu đối với CSDL duyệt web của người dùng để tìm mô hình sử dụng của người sử dụng và chủ động cung cấp dịch vụ cá nhân theo mô hình quan tâm của người dùng. Những dịch vụ thông tin cần đáp ứng như: Thông báo cho người sử dụng một cách kịp thời khi tìm thấy những thông tin hoặc thư mục dữ liệu mới; Quảng bá thông tin có liên quan khi người dùng truy cập dựa vào mức độ quan tâm của người dùng; Theo dõi nhu cầu của người dùng, phát hiện những nhu cầu mới của người sử dụng; Cung

cấp các báo cáo dự báo tương ứng, phân tích năng động lợi ích của người sử dụng.

Các chức năng này được phản ánh trong trang đăng nhập của người dùng, đây là mục tiêu trong thư viện kỹ thuật số, hoặc tích hợp nó vào trong cổng thông tin có liên quan đến thư viện kỹ thuật số, miễn là người sử dụng tùy biến các dịch vụ tương ứng, theo kết quả của việc khai phá dữ liệu, hệ thống sẽ nhắc nhở người dùng xem các thông tin mới nhất trên trang chủ cá nhân qua e-mail hoặc điện thoại di động thông qua tin nhắn và các hình thức khác, và có thể giúp người sử dụng tự tạo không gian thông tin.

#### **4.5. Tự động hóa xử lý thông tin**

Từ xu hướng phát triển của thư viện kỹ thuật số, các nguồn lực sẽ bao gồm các tài liệu giấy, sách điện tử, tạp chí điện tử, cơ sở dữ liệu đĩa CD, dữ liệu đa phương tiện và dữ liệu Web, và tất cả các loại dữ liệu này mức độ tăng trưởng là rất nhanh. Vì vậy, các yêu cầu tự động hóa đối với việc xử lý thông tin đang ngày càng cấp bách, đặc biệt là việc áp dụng công nghệ khai phá dữ liệu trong quá trình xử lý tự động các dữ liệu văn bản.

Khai phá văn bản hay khai phá dữ liệu văn bản còn được gọi là khám phá tri thức từ cơ sở dữ liệu văn bản. Khai phá văn bản đề cập đến quá trình chiết xuất các mẫu hoặc tri thức từ tài liệu văn bản. Khai phá văn bản là thực hiện tự động hóa xử lý thông tin, trong thư viện kỹ thuật số nó mang lại hiệu quả cao đối với khai thác các tài nguyên văn bản phi cấu trúc để thực hiện truy hồi văn bản, phân loại, tổng hợp, trả lời các câu hỏi, phân tích tâm lý người dùng, và các mục tiêu khác.

#### **4.6. Hỗ trợ các công việc khác**

*Đối với bộ phận cung cấp tài nguyên:* Bộ phận này có thể sử dụng các chức năng của khai phá dữ liệu để phân tích và sử dụng nguồn kinh phí một cách hiệu quả. Làm thế nào để việc sử dụng nguồn kinh phí hạn chế dành cho việc mua sách- đảm bảo về chất lượng và tính hợp lý của hệ thống tài nguyên

thông tin của thư viện. Chính vì vậy, việc định vị chính xác nhu cầu độc giả là một yếu tố quan trọng để nâng cao tỷ lệ sử dụng các nguồn lực. Việc sử dụng phân nhóm khai phá dữ liệu và công nghệ phát hiện độ lệch và phương pháp câu hỏi của độc giả có thể cung cấp nền tảng cơ bản cho việc phân tích phân nhóm, phân tích kết quả khảo sát, và hiểu được nhu cầu của độc giả thông qua việc sử dụng sách, và thông tin phản hồi của độc giả, qua đó đưa ra quyết định phù hợp cho công tác bổ sung nguồn tài liệu.

*Đối với bộ phận phục vụ:* Sử dụng các phương pháp phân tích kết hợp khai phá dữ liệu để phân tích các dữ liệu mượn trả. Những cuốn sách có số lượng giao dịch lớn sẽ dành vị trí ưu tiên. Những người mượn thường xuyên hoặc những cuốn sách đã bị hư hỏng cần phải có hình thức phản hồi nhanh chóng cho bộ phận cung cấp tài nguyên để tăng số lượng hoặc thay đổi số lượng. Ngoài ra các hoạt động khác cũng được tự động hóa để đồng bộ trong quản lý thư viện.

## **5. Kết luận**

Ngày nay, thư viện số có vai trò và ý nghĩa vô cùng to lớn trong các trường đại học nói chung và trong các tổ chức nói riêng. Thư viện số đã hình thành và ngày càng làm giá trị hiện tại của nó. Khai phá dữ liệu cung cấp và hỗ trợ kỹ thuật cho các tổ chức và quản lý các nguồn tài nguyên kỹ thuật số, thúc đẩy sự mở rộng chất lượng dịch vụ và dịch vụ, và cùng một lúc làm cho phương pháp nghiên cứu các công nghệ khai phá dữ liệu phát triển cả về quy mô lẫn chiều sâu. Tuy nhiên, ngoài các dữ liệu văn bản lớn, thư viện số còn có một số lượng lớn các dữ liệu đa phương tiện. Vì vậy, làm thế nào để sử dụng công nghệ khai phá dữ liệu một cách hiệu quả để kích hoạt nó xử lý các tính năng văn bản có kích thước lớn, mô tả chính xác các đặc tính của thông tin đa phương tiện, chính xác hóa quy trình thực hiện của không gian khái niệm, và cải thiện tốc độ của tất cả các loại thông tin. Trong tương lai, thư viện số chắc chắn sẽ phát triển

nhanh chóng. Công nghệ hiện có từ khai phá tri thức và ứng dụng tri thức thông minh sẽ phát huy tốt hơn sự phát triển của thư viện số. Kỹ thuật khai phá dữ liệu sử dụng các nguồn thông tin rộng lớn của nó trong hoạt động khai phá thông minh, một sự lựa chọn lớn của các công cụ khai phá và các thuật toán, cá nhân hoá dịch vụ thư viện số trở thành một phần không thể thiếu trong xây dựng hỗ trợ kỹ thuật cho thư viện số.

## **TÀI LIỆU THAM KHẢO**

1. Oded Maimon, Lior Rokach (2010). Data Mining and Knowledge Discovery Handbook, Second Edition, Springer New York Dordrecht Heidelberg London, April 2010.
  2. Han, J.; Kamber, M (2006). Data Mining: Concepts and Techniques. Morgan Kaufmann, Burlington, US.
  3. Cios, K. J.; Swiniarski, R. W.; Pedrycz, W.; Kurgan, L. A. The Knowledge Discovery Process (2007). Data Mining: A Knowledge Discovery Approach. Springer, New York, USA, 2007, p 9-24.
  4. Na, C. Data Mining Research Status and Development Direction. Computer and Information Technology, 2006.
  5. Rui, W., Tao, M., Chen, C (2007). Data Mining and Its Application Status Analysis. Computer Application Technology.
  6. Hui, L., Daiji, H (2006). Data mining technology development and its application. Gansu Science and Technology.
  7. Zhao, Z., Yan, F (2001). Commerce technology research and application of data mining study. P.4. Hunan Economic Management Institute.
  8. Xiaolan, F (2011). Data mining application in digital libraries. Heihe Sciences.
  9. Zhao, C.-Y (2007). Data mining technology and its application in University Library. Beijing College of Political Science, .
  10. Shan, Z., Tingting, X (2010). Web-based data mining in e-commerce application. Xiaogan University.
- (Ngày Tòa soạn nhận được bài: 4-2-2017; Ngày phản biện đánh giá: 1-3-2017; Ngày chấp nhận đăng: 28-4-2017).