

BIG DATA VÀ XU HƯỚNG ỨNG DỤNG TRONG HOẠT ĐỘNG THÔNG TIN - THƯ VIỆN

ThS Nguyễn Lê Phương Hoài
Viện Thông tin Khoa học xã hội

• **Tóm tắt:** Big Data là một thuật ngữ được sử dụng để chỉ những bộ dữ liệu khổng lồ, chủ yếu không có cấu trúc, được thu thập từ nhiều nguồn khác nhau. Big Data có nhiều tác động, ứng dụng và được xem như một yếu tố quyết định đến việc phát triển, mang lại lợi thế cạnh tranh cho tổ chức. Bài viết tổng quan lược sử các quan điểm về Big Data, đồng thời nhấn mạnh vào các xu hướng ứng dụng trong hoạt động thông tin - thư viện.

• **Từ khóa:** Big Data; dữ liệu lớn; hoạt động thư viện.

BIG DATA APPLICATION IN LIBRARY AND INFORMATION ACTIVITIES

• **Abstract:** Big Data is a term used to refer to huge, mostly unstructured datasets, collected from a variety of sources. Big Data has many impacts, applications and is considered as a decisive factor in the development, bringing competitive advantages to the organization. The overview paper summarizes the views on Big Data and emphasizes application trends in library and information activities.

• **Keywords:** Big Data; library activities.

1. LƯỢC SỬ CÁC QUAN ĐIỂM VỀ BIG DATA

Hiện nay, chưa có một định nghĩa chính xác cho thuật ngữ Big Data. Big được ghi nhận lần đầu tiên trong báo cáo "Application-controlled demand paging for out-of-core visualization" của Michael Cox và David Ellsworth được trình bày tại Hội nghị IEEE lần thứ 8 (vào tháng 10 năm 1997). Ý tưởng đưa ra ban đầu là dung lượng thông tin đã tăng quá lớn tới mức các bộ nhớ máy tính dùng để xử lý không còn đủ số lượng cần khảo sát, do vậy các kỹ sư cần cải tạo các công cụ để có thể phân tích được tất cả các thông tin [11].

Tháng 8 năm 1999, Steve Bryson, David Kenwright, Michael Cox, David Ellsworth, và Robert Haimes đăng bài "Visually exploring gigabyte data sets in real time" trên Tạp chí Communications of the ACM. Đây là bài viết đầu tiên sử dụng thuật ngữ "Big Data". Các tác giả nhận định: "Những chiếc máy tính mạnh là lợi thế cho việc khảo sát nhiều lĩnh vực, cũng có thể là bất lợi; tính toán nhanh chóng tạo ra một lượng lớn dữ liệu. Nếu trước kia bộ dữ liệu megabyte đã từng được coi là lớn, thì bây giờ chúng ta có thể tìm thấy những bộ dữ liệu của cá nhân vào khoảng 300 GB" [15].

Tháng 11 năm 2000, Francis X. Diebold trình bày tại Đại hội Thế giới lần thứ VIII của Hiệp hội Kinh tế lượng bài viết "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting". Trong bài viết này, tác giả khẳng định: "Gần đây, nhiều ngành khoa học như vật lý, sinh học, khoa học xã hội, vốn đang buộc phải đương đầu với khó khăn - đã thu được lợi từ hiện tượng Big Data và đã gặt hái được nhiều thành công. Big Data chỉ sự bùng nổ về số lượng (và đôi khi, chất lượng), khả năng liên kết cũng như độ sẵn sàng của dữ liệu, chủ yếu là kết quả của những tiến bộ gần đây và chưa từng có trong việc ghi lại dữ liệu và công nghệ lưu trữ" [4].

Tháng 2 năm 2001, Doug Laney - nhà phân tích của Tập đoàn Meta, công bố nghiên cứu "3D Data Management: Controlling Data Volume, Velocity, and Variety". Laney cho rằng, những thách thức và cơ hội nằm trong việc tăng trưởng dữ liệu có thể được mô tả bằng mô hình "3Vs": tăng về số lượng lưu trữ (Volume), tăng về tốc độ xử lý (Velocity) và tăng về chủng loại (Variety) [3]. Một thập kỷ sau, mô hình "3Vs" đã trở thành thuật ngữ được chấp nhận rộng rãi trong việc xác định dữ liệu lớn ba chiều. Nhiều công ty và tổ chức

trong lĩnh vực công nghệ thông tin tiếp tục sử dụng mô hình “3Vs” này để định nghĩa Big Data.

Tháng 12 năm 2008, Randal E. Bryant, Randy H. Katz, và Edward D. Lazowska công bố bài viết “Big-Data Computing: Creating Revolutionary breakthroughs in Commerce, Science and Society”, trong đó miêu tả: “Cũng như công cụ tìm kiếm đã làm thay đổi cách chúng ta tiếp cận thông tin, các hình thức khác của sử dụng Big Data có thể sẽ làm thay đổi cách hoạt động của các công ty, các nhà nghiên cứu khoa học, các học viên y tế, quốc phòng và tình báo,... Sử dụng Big Data có lẽ là đổi mới lớn nhất trong công nghệ máy tính suốt một thập kỷ qua. Chúng tôi chỉ mới bắt đầu nhìn thấy tiềm năng của nó trong việc thu thập, sắp xếp và xử lý dữ liệu của tất cả các tầng lớp xã hội. Một khoản đầu tư dù khiêm tốn của chính phủ liên bang sẽ thúc đẩy phát triển và mở rộng nó” [13].

Tháng 2 năm 2010, Kenneth Cukier đăng trên tờ *The Economist Newspaper* bài viết “Data, data everywhere”. Cukier viết: “... thế giới chứa một số lượng thông tin số lớn đến mức không tưởng, và càng ngày càng được nhân rộng với tốc độ nhanh hơn bao giờ hết... Hiệu quả đã được thể hiện ở khắp mọi nơi, từ kinh doanh đến khoa học, từ chính phủ đến nghệ thuật. Các nhà khoa học và kỹ sư máy tính đã đặt ra một thuật ngữ mới cho hiện tượng này: Big Data” [8].

Tháng 5 năm 2012, Danah Boyd và Kate Crawford đưa ra luận điểm của họ trong bài “Critical Question for Big Data” trên tờ *Information, Communications and Society*. Các tác giả định nghĩa Big Data như là “một hiện tượng văn hóa, công nghệ và học thuật dựa trên sự tương tác của: 1) Công nghệ tối đa hóa sức mạnh tính toán và độ chính xác thuật toán để thu thập, phân tích, liên kết, và so sánh các tập dữ liệu lớn; 2) Phân tích: tạo ra trên dữ liệu lớn để xác định mô hình cho tuyên bố kinh tế, xã hội, kỹ thuật và pháp lý; 3) Thần thoại: niềm tin phổ biến rằng dữ liệu lớn cung cấp một hình thức cao hơn của trí thông minh và kiến thức có thể tạo ra những hiểu

biết mà trước đây không thể, với hào quang của sự thật, khách quan, chính xác” [2].

Sau đó, Gartner - công ty nghiên cứu và tư vấn công nghệ thông tin - bổ sung thêm rằng “Big Data ngoài 3 tính chất (số lượng, tốc độ xử lý và chủng loại) thì còn phải cần đến các dạng xử lý mới để đưa ra quyết định khám phá sâu vào sự vật/sự việc và tối ưu hóa các quy trình làm việc” [5]. Cùng quan điểm đó, Tan Jee Toon cho rằng Big Data là khối lượng dữ liệu rất lớn được tạo ra từ mọi thứ xung quanh chúng ta, từ các thiết bị kỹ thuật số như di động, video, hình ảnh, tin nhắn tới các thiết bị cảm biến, các máy móc được kết nối tới các trang web và mạng xã hội. Big Data có đặc điểm là được sinh ra với khối lượng (volume), tốc độ (velocity), độ đa dạng (variety) và tính xác thực (veracity) rất lớn [16].

Năm 2014, Gartner đưa ra khái niệm mới về Big Data bằng mô hình “5Vs”, gồm: Volume (khối lượng), Velocity (tốc độ), Variety (tính đa dạng), Veracity (tính xác thực) và Value (giá trị). Trong đó: Volume là khối lượng Big Data được tạo ra mỗi ngày. Công nghệ Big Data có thể lưu trữ và sử dụng những tập dữ liệu trong các hệ thống phân tán, nơi mà dữ liệu chỉ được lưu trữ một phần tại các điểm khác nhau và được tập hợp bởi phần mềm. Velocity là tốc độ dữ liệu mới được tạo ra và tốc độ dữ liệu chuyển động. Công nghệ Big Data có thể phân tích dữ liệu ngay khi chúng được tạo ra mà không cần lưu giữ chúng trong các cơ sở dữ liệu. Variety là các kiểu khác nhau của dữ liệu. Công nghệ Big Data cho phép lưu trữ các loại dữ liệu có cấu trúc truyền thống (được lưu trữ trong các bảng hoặc các cơ sở dữ liệu quan hệ) và phi cấu trúc (bao gồm các thông điệp, trao đổi của mạng xã hội, các hình ảnh, dữ liệu cảm biến, video, tiếng nói...). Veracity là tính hỗn độn hoặc tính tin cậy của dữ liệu. Công nghệ Big Data và phân tích dữ liệu cho phép kiểm soát những loại dữ liệu này. Value là giá trị của dữ liệu. Việc tiếp cận Big Data sẽ chỉ có ý nghĩa khi chúng ta chuyển được dữ liệu thành những thứ có giá trị. Đây là khái niệm đầy đủ về 5 tính chất của Big Data [5].

2. XU HƯỚNG ỨNG DỤNG BIG DATA TRONG HOẠT ĐỘNG THÔNG TIN - THƯ VIỆN

Ngày nay, một lượng lớn dữ liệu số có thể được tạo ra bởi những hệ thống liên kết với các mạng xã hội. Theo Howe D. (2008): “Chỉ riêng trong các lĩnh vực nghiên cứu khoa học, lượng dữ liệu đang tăng lên với tốc độ tăng trưởng theo cấp số nhân” [7]. Số lượng dữ liệu ngày càng tăng là xu thế phát triển trong nhiều lĩnh vực khác nhau và dữ liệu lớn (Big Data) được sử dụng rộng rãi trong các lĩnh vực, tổ chức với nhiều mục đích khác nhau. Các công ty sử dụng Big Data để tìm hiểu hành vi tiêu dùng của khách hàng, để đề xuất mua sắm cá nhân hóa cho từng đối tượng dựa trên thông tin thu thập được (Ebay, Facebook, Google...). Các cơ sở nghiên cứu khoa học sử dụng Big Data để tìm ra những khám phá khoa học mới, ví dụ như xây dựng bản đồ gene của con người,... Việc sử dụng Big Data trong hoạt động thông tin - thư viện đã bắt đầu được quan tâm nghiên cứu. Campbell D. Grant, Cowan Scott R. (2016) phân tích nghịch lý về quyền riêng tư khi xem xét lại giá trị cốt lõi của thư viện trong thời đại Big Data và dữ liệu liên kết [1]. Kim Young Seok (2017) phân tích Big Data trong các hoạt động và dịch vụ thư viện công cộng bằng phương pháp khuôn mặt Chernoff [9]. Gerrard D., Mooney J. và Thompson D. (2017) xem xét kiến trúc của các hệ thống bảo quản kỹ thuật số hỗ trợ phân tích dữ liệu, các bộ tài nguyên được bảo tồn ở quy mô lớn và phân tích dự báo việc sử dụng Big Data của các nhà nghiên cứu trong thời gian tới [6]. Waqar Ahmed và Kanwal Ameen (2017) tổng quan các khái niệm về Big Data và đo lường xu hướng liên quan của Big Data trong lĩnh vực quản lý thông tin và thư viện [17]. Ye Chunlei (2017) nghiên cứu về công nghệ chính của dịch vụ Big Data trong thư viện đại học [18]. Zhan Ming, Widén Gunilla (2018) nghiên cứu vai trò của thư viện công cộng trong thời đại Big Data [20]. Li Shuqing; Jiao Fusen; Zhang Yong; Xu Xia (2019) nghiên cứu các vấn đề và thay đổi của thư viện số trong thời đại Big Data từ góc độ dịch vụ người dùng tin [10],... Các nghiên cứu bằng những cách tiếp cận khác nhau đã đề cập đến nhiều vấn đề của hoạt động thư viện

trong thời đại Big Data. Bàn về xu hướng ứng dụng của Big Data trong hoạt động thư viện tuy chưa có một nghiên cứu đầy đủ và trực tiếp nào, nhưng có thể tổng hợp thành các xu hướng chính như sau:

Một là, tổ chức lưu trữ, bảo quản dữ liệu

Marydee Ojala nhận định: “Các thư viện đã phải đối mặt với Big Data từ trước khi thuật ngữ này xuất hiện và trở nên thông dụng như ngày nay. Bộ sưu tập các tài nguyên số được bảo tồn đang trở thành nguồn Big Data trong các thư viện. Khối lượng và tính đa dạng dữ liệu đang tăng lên nhanh chóng đòi hỏi các thư viện phải có phương pháp tổ chức lưu trữ, bảo quản dữ liệu hợp lý” [12]. Nguồn dữ liệu thư viện bao gồm: nguồn dữ liệu mô tả tài liệu thư viện, nguồn tài nguyên số hóa tài liệu thư viện, nguồn tài liệu số thư viện bổ sung qua việc mua hay sử dụng chung, nguồn dữ liệu khảo sát thư viện, dữ liệu định tính, dữ liệu tương tác xã hội,... Trước đây, các thư viện đều lưu trữ dữ liệu ngoại tuyến trên các cuộn băng, đặt trong các cơ sở lưu trữ. Trước tác động của công nghệ thông tin và cuộc Cách mạng công nghiệp lần thứ tư, các thư viện đã xem xét lại cách thức lưu trữ truyền thống và triển khai một giải pháp lưu trữ tiết kiệm chi phí hiệu quả. Dữ liệu được lưu trữ theo hai cách, cả trên các thiết bị ngoại tuyến (thẻ nhớ SD, ổ cứng ngoài, ổ đĩa flash) và lưu trữ trực tuyến trên đám mây. Với phương thức kết hợp sử dụng băng từ để bảo quản lưu trữ, sử dụng ổ đĩa cho các thông tin thường xuyên được yêu cầu, và sử dụng lưu trữ đám mây cho các Big Data. Các thư viện hướng đến việc xác định khả năng truy cập tài nguyên thư viện (bao gồm cả tài nguyên vật chất và dữ liệu), xác định nhu cầu của người dùng tin và lập kế hoạch vòng đời cho tài nguyên thư viện. Trong thời gian tới, khi các yêu cầu mới thúc đẩy việc sử dụng Big Data, các thư viện hướng tới việc thu nhận, tổ chức lưu trữ dữ liệu (lưu trữ vật lý trong các máy chủ hoặc trong các cơ sở dữ liệu), bảo tồn dữ liệu và phổ biến dữ liệu, làm cho dữ liệu có sẵn trong thư viện trở nên dễ dàng tiếp cận hơn thông qua các sản phẩm trực quan. Các thư viện tiến tới xây dựng, tạo lập hệ thống bảo quản

kỹ thuật số (bảo tồn cả tài nguyên số và siêu dữ liệu mô tả) có thể phát triển trong nhiều năm tới để đáp ứng với các yêu cầu mới.

Hai là, cung cấp sản phẩm, dịch vụ thư viện mang tính cá nhân hóa, tùy chỉnh theo người dùng tin

Hiện nay, các thư viện đang có xu hướng cung cấp nhiều tài nguyên và dịch vụ trực tuyến. Nhiều thư viện đang sử dụng các phương tiện truyền thông xã hội như facebook, instagram để quảng bá các sản phẩm và dịch vụ thư viện. Các phương tiện truyền thông xã hội này cho phép các thư viện thu thập và sở hữu dữ liệu người dùng thư viện. Cùng với các dữ liệu khảo sát thư viện, dữ liệu định tính (thông qua phỏng vấn, bảng trò chuyện...), dữ liệu tương tác xã hội (từ các trang truyền thông xã hội)... các thư viện mở rộng bộ sưu tập dữ liệu thư viện và dựa vào các công cụ và phương pháp đánh giá để đánh giá nhu cầu tin của người dùng thư viện. Từ đó, thư viện cung cấp các sản phẩm và dịch vụ thư viện phù hợp với nhu cầu của người dùng. Tuy nhiên, trong thời gian tới, “sự gia tăng của Big Data làm cho một số tác vụ thu thập dữ liệu dễ dàng và nhanh hơn, cho phép các thư viện vượt ra ngoài công việc đơn giản là đếm và tổng hợp các biện pháp thống kê và bắt đầu tham gia vào phân tích dữ liệu phức tạp như phân tích học tập và phân tích hiệu suất nghiên cứu” [19]. “Trong kỷ nguyên Big Data, không chỉ tài nguyên dữ liệu của thư viện số hiện đại có đặc điểm của Big Data mà cả dịch vụ thư viện hiện tại cũng cần sử dụng phương pháp Big Data để đổi mới.” [10]. “Big Data có thể làm thay đổi mô hình cung cấp dịch vụ thư viện trong tương lai và vai trò của thư viện cũng sẽ thay đổi cho phù hợp” [17]. Các nhà nghiên cứu cho rằng việc ứng dụng Big Data có thể tác động đến hoạt động thông tin - thư viện thông qua việc cung cấp các dịch vụ thư viện, chuyển đổi phương thức cung cấp dịch vụ ứng dụng công nghệ thông tin hiện nay sang phương thức gắn liền tự động hóa dịch vụ và tích hợp các hệ thống xử lý. Các hỗ trợ công nghệ đối với việc kết nối trực tiếp và việc sử dụng Big Data sẽ tạo ra các lợi thế

cạnh tranh để thư viện thu hút người dùng tin. Theo Li Shuqing, Jiao Fusen, Zhang Yong, Xu Xia: “Các vấn đề và tiềm năng của các thư viện số trong thời đại Big Data liên quan đến dữ liệu, công nghệ, dịch vụ và người dùng tin. Sử dụng Big Data hiện có và xem xét các đặc điểm về nhu cầu của người dùng tin hiện tại theo quan điểm của người dùng tin, thư viện có thể đưa ra các ý tưởng, phương pháp hiệu quả hơn để cải thiện các dịch vụ hiện có trong các thư viện số” [10]. Đồng thời, nhu cầu cá nhân hóa của người dùng tin trong thời đại Big Data tạo nên yếu tố thúc đẩy sự phát triển của thư viện số từ dịch vụ chia sẻ tài nguyên sang dịch vụ hướng đến người dùng tin. Kim Young Seok cho rằng: “Bằng cách sử dụng các phân tích trong thời gian thực, các thư viện có thể thiết kế các dịch vụ được cá nhân hóa cho từng người dùng tin. Big Data cung cấp thông tin chuyên sâu về hành vi và thông tin cá nhân của người dùng tin, từ đó tạo ra trải nghiệm cá nhân hóa” [9]. Ví dụ, người dùng tin tìm kiếm trên trang OPAC của thư viện thì dữ liệu về những gì người dùng tin gõ ở mục tìm kiếm, tần suất tìm kiếm, số lần tham khảo danh mục tài liệu, số lần xem mô tả tài liệu,... được thu thập và phân tích để tối ưu trải nghiệm, tạo cơ hội lớn để thu hút người dùng tin thông qua cá nhân hóa. Đặc biệt, với các công cụ phân tích dự báo của Big Data, thư viện sẽ nắm được thị hiếu, nhu cầu chính xác để cung cấp các sản phẩm, dịch vụ phù hợp với người dùng tin trong thời gian thực.

Ba là, ứng dụng dịch vụ phân tích dự báo

Giống như hầu hết các ngành khác, phân tích dự báo sẽ là một sự thay đổi lớn, quan trọng trong các cơ quan thông tin - thư viện. Phân tích sử dụng Big Data làm nền tảng cho sự phát triển của thư viện sẽ giúp thư viện hoạt động hiệu quả hơn, đồng thời làm thay đổi căn bản mối quan hệ giữa thư viện và người dùng tin. Theo cách truyền thống, mối quan hệ giữa thư viện và người dùng thư viện khá đơn giản. Người dùng thư viện nộp tiền, làm thẻ thư viện và đổi lại, họ được phục vụ trong các dịch vụ khác nhau của thư viện. Tuy nhiên, mối quan hệ này đang dần thay đổi

và người dùng thư viện không chỉ đơn giản là người sử dụng dịch vụ mà đã trở thành một đối tác trong việc cung cấp dịch vụ của thư viện. Người dùng thư viện cung cấp dữ liệu hành vi người dùng. Thông qua các dữ liệu cá nhân như lịch sử sử dụng tài liệu thư viện, lịch sử tìm kiếm, cách thức, thói quen tìm kiếm,... các công cụ Big Data phân tích dữ liệu, đưa ra thông tin chi tiết, xác định khuynh hướng, nhu cầu sử dụng thư viện, nhu cầu tài liệu trong hiện tại và dự đoán các hoạt động của người dùng thư viện trong tương lai; các công cụ Remarketing trên nền tảng website thư viện hay mạng xã hội thư viện sẽ hỗ trợ các thư viện marketing trực tiếp đến người dùng tin các tài liệu dựa trên sở thích và dự báo nhu cầu,... Simovic Aleksandar (2018) nhận định: “Các công cụ Big Data kết hợp với các thuật toán khoa học cho phép các thư viện dự đoán giá trị tiềm năng của người dùng tin cá nhân và mô hình hoạt động thư viện trong tương lai, giúp dự báo tốt hơn các vấn đề phát sinh trong quản lý các sản phẩm và dịch vụ thư viện để tăng cường khả năng cung cấp thông tin tốt nhất cho người dùng tin” [14]. Về phía thư viện, việc sử dụng tài nguyên Big Data hiện có và xem xét các đặc điểm về nhu cầu của người dùng tin hiện tại theo quan điểm của người dùng tin, có thể đưa ra các ý tưởng và phương pháp hiệu quả hơn để cải thiện các dịch vụ hiện có trong thư viện kỹ thuật số. Đồng thời, căn cứ vào các kết quả phân tích, dự báo, thư viện có thể xác định thời gian, phương thức để quảng cáo cho các sản phẩm và dịch vụ thư viện đến người dùng thư viện. Về phía người dùng thư viện, dựa vào các kết quả dự báo về hành vi tìm kiếm, tra cứu, sử dụng thư viện, các hệ thống khuyến nghị (Recommendation Engine) sẽ gửi đến người dùng tin các tài liệu có thể họ quan tâm.

Bốn là, mở rộng dịch vụ chăm sóc người dùng tin

Dịch vụ chăm sóc người dùng tin có vai trò quan trọng trong sự thành công của các thư viện, đặc biệt trong môi trường thư viện điện tử, thư viện số. Các thư viện đang cố gắng để hiểu được người dùng thư viện, giúp họ dễ dàng liên hệ với thư viện để đáp ứng

các thắc mắc, nhu cầu của mình. Big Data hỗ trợ các thư viện trong việc hình thành một hệ thống chăm sóc người dùng thư viện linh hoạt, tạo ra giá trị từ quá trình xây dựng mối quan hệ thân thiết với người dùng thư viện. Cùng với Big Data, hệ thống trả lời tự động (như Chatbot) không cần sự trợ giúp của con người phát triển tương ứng, giúp tăng hiệu quả phân tích dữ liệu Big Data. Hiện nay, nhiều thư viện sử dụng Chatbot để giao tiếp, trao đổi với người dùng thư viện, tiếp nhận các ý kiến cụ thể của người dùng về sản phẩm và dịch vụ của thư viện cũng như các nhu cầu của người dùng. Khi người dùng truy cập vào website hay fanpage của thư viện, họ có rất nhiều thắc mắc và muốn được giải đáp. Chatbot sẽ đưa ra các gợi ý, hỗ trợ từng bước một, cung cấp thông tin về các sản phẩm, dịch vụ của thư viện cho người dùng. Chatbot được thiết kế và phát triển để vừa là nền tảng trò chuyện giữa Chatbot với người dùng vừa cung cấp các tính năng thu thập dữ liệu người dùng thư viện từ các đoạn đối thoại. Qua những dữ liệu người dùng thu thập được, công cụ phân tích dữ liệu Big Data tiến hành phân tích, xác định những nhu cầu, mong muốn của người dùng thư viện và từ đó thư viện có thể đưa ra các phản hồi thích hợp và xây dựng mối quan hệ lâu dài với người dùng thư viện. Bên cạnh đó, Chatbot nhắc nhở người dùng thư viện về việc sử dụng thư viện như thời hạn trả tài liệu, thời hạn đối thể thư viện hay gợi ý nhiều dịch vụ thư viện khác mà người dùng có thể cần đến trong quá trình sử dụng,... Đặc biệt, Chatbot giúp thư viện chủ động hỗ trợ 24/7, tăng trải nghiệm tối đa cho người dùng thư viện mọi lúc. Chatbot lưu lại lịch sử đối thoại, thông tin người dùng làm cơ sở cho các thư viện sử dụng phân tích Big Data có thể phân tích được chính xác các khúc mắc của người dùng thư viện và cả những sản phẩm hay dịch vụ còn thiếu trong chính thư viện. Chatbot hỗ trợ các thư viện khai thác Big Data phục vụ người dùng. Trong tương lai, số thư viện sử dụng Chatbot để chủ động chăm sóc người dùng tin được dự báo sẽ tăng lên nhanh chóng bởi những tính năng và lợi ích mà Chatbot mang lại. Cùng với đó, thông qua dữ liệu người dùng,

các thư viện có thể phân tích, dự đoán các vấn đề có thể xảy ra trong quá trình cung cấp các sản phẩm, dịch vụ thông qua phân tích cảm nhận của người dùng thư viện và thực hiện các giải pháp kịp thời.

Có thể thấy, Big Data giúp tối ưu hóa hoạt động thư viện bằng việc thu thập, phân tích thông tin, tăng trải nghiệm của người dùng tin bằng cách cá nhân hóa thư viện số. Cùng với đó, Big Data có thể giúp các thư viện tiến hành phân tích dự báo, tìm ra các đặc điểm chung dự báo thị hiếu đọc, tình trạng sử dụng thư viện của người dùng tin bằng việc kết hợp các cơ sở dữ liệu. Không chỉ vậy, Big Data tạo dịch vụ mới dựa vào trải nghiệm của người dùng tin trong quá trình sử dụng thư viện.

TÀI LIỆU THAM KHẢO

1. Campbell D. Grant, Cowan Scott R. (2016). "The Paradox of privacy: revisiting a core library value in an age of big data and linked data", *Library trends*, Vol. 64, No. 3, P. 492-811.
2. Danah Boyd, Kate Crawford (2012). *Critical Question for Big Data*, Information, Communications and Society.
3. Doug Laney (2001). "3D Data Management: controlling Data Volume, Velocity, and Variety", *Application Delivery Strategies*, Meta Group. File:949.
4. Francis X. Diebold (2000). "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting", Discussion of Reichlin and Watson paper, in *Economics and Econometrics*, Eighth World Congress of the Econometric Society.
5. Gartner (2013). *Survey Analysis: Big Data Adoption in 2013 shows substance behind the hype*.
6. Gerrard, D., Mooney, J., Thompson, D. (2017). "Digital Preservation at Big data scale: proposing a step - change in preservation system architectures", *Library Hi Tech*, <http://doi.org/10.1108/LHT-06-2017-0122>, truy cập ngày 17/10/2019.
7. Howe D. (2008). "The future of biocuration", *Nature* 455, P. 47-50.
8. Kenneth Cukier (2010). "Data, data everywhere", *A special report on managing information*, *Economist Newspaper*, Volume 394.
9. Kim Young Seok (2017). "Big data analysis of public library operations and services by using the chernoff face method", *Journal of Documentation*, Vol. 73, No. 3, P. 466-480.
10. Li Shuqing; Jiao Fúen; Zhang Yong; Xu Xia (2019). "Problems and Changes in Digital Libraries in the Age of Big data from the perspective of user services", *Journal of Academic Librarianship*, Vol. 45.
11. Michael Cox, David Ellsworth (1997). "Application - Controlled Demand Paging for Out - of - Core Visualization", Report NAS-97-010, NASA Ames Research Center.
12. Marydee Ojala (2018). "Big Data and AI: technology, transparency, and trust", <http://www.infotoday.com>, truy cập ngày 14/11/2019.
13. Randal E. Bryant, Randy H. Katz, và Edward D. Lazowska (2008). "Big - Data Computing: Creating Revolutionary breakthroughs in Commerce, Science and Society", *Computing Community Consortium*, <http://www.cra.org/cccl/initiatives>, truy cập ngày 14/11/2019.
14. Simovic Aleksandar (2018). "A Big Data smart library recommender system for an educational institution", *Library Hi Tech*, Bradford, Vol. 36, Iss. 3, tr.498-523
15. Steve Bryson, David Kenwright, Michael Cox, David Ellsworth, Robert Haines (1999). "Visually exploring gigabyte data sets in real", *Communications of the ACM*, Vol. 42, No.8, tr.83-90.
16. Tan Jee Toon (2014). "Dữ liệu lớn, nhân tố thay đổi "cuộc chơi" của doanh nghiệp", <http://vneconomy.vn/cuoc-song-so/du-lieu-lon-nhan-to-thay-doi-cuoc-choi-cua-doanh-nghiep-20140422025542917.htm>, truy cập ngày 14/11/2019.
17. Waqar Ahmed, Kanwal Ameen (2017). "Defining big data and measuring its associated trends in the field of information and library management", *Library Hi Tech News*, p. 21-24.
18. Ye Chunlei (2017). "Research on the key technology of big data service in university library", the Institute of Electrical and Electronics Engineers, Inc. Conference Proceedings, Piscataway.
19. Yu Jen Chien (2016). "Library Data, Big Data or Better Data: Challenges from the Field", *ASIST Meeting, Proceeding of ASIST annual meeting*, Vol 53, No. 1.
20. Zhan Ming, Widén Gunilla (2018). "Public libraries: roles in big data", *The Electronic library*, Vol. 36, No.1, P. 133-145.

(Ngày Tòa soạn nhận được bài: 26-12-2019;
Ngày phản biện đánh giá: 10-03-2020; Ngày chấp
nhận đăng: 15-03-2020).