

CƠ SỞ DỮ LIỆU TRỰC TUYẾN VÀ NGÀNH CÔNG NGHIỆP KHAI THÁC THÔNG TIN TRỰC TUYẾN: HIỆN TRẠNG VÀ XU HƯỚNG

Hoàng Hương | 30/11/2011 | Bài viết chọn lọc, Headlines, Sản phẩm, dịch vụ thông tin | 6 phản hồi



Kể từ đầu những năm 1970, một nhóm các công ty, cơ quan liên bang, và các tổ chức nghề nghiệp có lợi nhuận và phi lợi nhuận ở Mỹ đã có sẵn một lượng thông tin trực tuyến là thông tin số, thông tin có chất lượng cao cho các thư viện và người sử dụng thư viện. Đặc điểm quan trọng của nhóm này là nó phát triển rất nhanh chóng và không có sự phân hoá về vai trò. Tuy nhiên, cùng với sự phát triển về hệ thống viễn thông hiện đại, ngành công nghiệp này có nhiều những biến đổi không thể lường hết được. Nhu cầu thông tin khoa học, kỹ thuật, kinh doanh, y tế, thông tin pháp luật... đang hiện hữu trong cuộc sống của

chúng ta từng ngày, từng giờ. Các nhu cầu thông tin được lấp đầy chủ yếu do các cơ quan, tổ chức sản xuất thông tin chuyên nghiệp.

Nói chung, chúng ta đều thấy rằng ngành công nghiệp khai thác thông tin trực tuyến bắt đầu tại Mỹ và Mỹ đã là một quốc gia chiếm ưu thế trên thị trường. Trong suốt gần 40 năm qua, tỷ lệ nội dung trực tuyến đến từ Mỹ luôn chiếm phần nửa trên thị trường toàn thế giới. Tuy nhiên, một số các cơ quan, tổ chức ở Châu Âu cũng đóng vai trò quan trọng như ESA-IRS, Dialog...và nước Nhật cũng tham gia tích cực vào ngành công nghiệp này. Và cũng với xu hướng toàn cầu hoá ngành công nghiệp này cũng lan toả ra tới mọi quốc gia. Mặc dù vậy, ngành công nghiệp này hiện nay đang bị đe dọa bởi những nguồn thông tin khổng lồ miễn phí trên internet và các công cụ tìm kiếm trực tuyến luôn sẵn sàng cho người sử dụng ở bất kỳ nơi đâu, bất kỳ không gian nào.

Là một quốc gia đang phát triển, các cơ quan thông tin thư viện của Việt Nam sẽ đạt được những lợi thế gì và đối mặt với những thách thức nào trong một cuộc chơi khi bản thân chúng ta là người đến sau?

Phần 1: Những nhà sản xuất nguồn tin số nổi tiếng và vai trò của họ

Các công ty, cơ quan chính phủ, và các tổ chức phi lợi nhuận từ nhiều ngành tham gia vào ngành công nghiệp sản xuất và khai thác thông tin trực tuyến trong mấy chục năm qua hoạt động chồng chéo, thay đổi vai trò và chức năng. Những nhiệm vụ của ngành công nghiệp dịch vụ thông tin trực tuyến bao gồm: thu thập thông tin, tạo biểu ghi, lập hồ sơ chỉ mục, tập hợp biểu ghi vào CSDL, tải CSDL vào máy tính, làm cho

các nguồn tin sẵn sàng trên web, tạo ra các giao diện để dễ truy cập, cấp phép cho các nhà phân phối, tiếp thị sản phẩm, đào tạo người sử dụng, hỗ trợ khách hàng, hoặc bất kỳ sự kết hợp nào những nhiệm vụ trên đang được thực hiện chông chéo và phức tạp. Mặc dù sự đa dạng to lớn, một đặc tính của tất cả các nhà đầu tư thành công trong ngành công nghiệp này đã hiểu rằng rất ít người xem trực tuyến tìm kiếm thông tin như là một mục tiêu cuối cùng của riêng mình, hoặc tìm thấy nó đặc biệt thú vị, nó chỉ là một phương tiện để hoàn thành một số mục tiêu trực tiếp. Vì vậy, các thành viên của ngành công nghiệp đã nỗ lực phấn đấu để tổ chức nội dung trực tuyến trong CSDL theo những cách dễ hiểu và họ đã làm việc để đào tạo người dùng sử dụng tốt nhất các sản phẩm của họ.

Mặc dù không thể tách bạch chính xác ai làm gì và vai trò của họ như thế nào vì cùng với thời gian, vai trò của các công ty cũng thay đổi nhanh chóng, nhưng chúng ta có thể kể đến tên tuổi của một số công ty, tổ chức thương mại Dịch vụ Khôi phục thư mục (BRS), đã trở thành BRS Ovid công nghệ sau này (nghiên cứu cơ sở dữ liệu trong tất cả các ngành); Thông Tin Nhóm Cambridge (cơ sở dữ liệu CSA), Chemical Abstracts Service CAS (STN quốc tế), chỉ số Dow Jones & Company (cơ sở dữ liệu Factiva, cơ sở dữ liệu doanh nghiệp Dow Jones, Wall Street Journal trực tuyến); Dun & Bradstreet -D & B (cơ sở dữ liệu khác nhau của Dun); EBSCO aggregator nội dung học thuật; Elsevier (ScienceDirect, EMBASE, Ei Engineering Village); Factiva-được thành lập như là một công ty liên doanh của Dow Jones và Reuters và hiện nay thuộc sở hữu của Dow Jones (kinh doanh, tài chính, và cơ sở dữ liệu tin tức); Gale (Thư viện tham khảo ảo và học thuật thương mại khác và thư mục phổ biến, cơ sở dữ liệu); FIZ Chemie Berlin (hóa chất); GBI (kinh doanh thông tin và tin tức); LexisNexis (các tập tin liên quan đến kinh doanh, pháp lý, và tin tức); ProQuest (báo chí, kinh doanh, học thuật, và cơ sở dữ liệu của người tiêu dùng); Questel (cơ sở dữ liệu bằng sáng chế, nhãn hiệu hàng hoá và khoa học); Platter được mua từ Wolters Kluwer và bây giờ chỉ tồn tại như OvidSP (ban đầu là cơ sở dữ liệu trên đĩa CD- ROM); Swets (tạp chí điện tử toàn văn) và Thomson (một công ty lớn, sở hữu qua nhiều năm sản xuất khác như cơ sở dữ liệu quan trọng và các nhà cung cấp như thoại, ISI Web of Science, Derwent Bằng sáng chế thế giới, Reuters, Westlaw, và Dialog DataStar),...

Trong nhóm của các doanh nghiệp thương mại lớn cũng đã có được các dịch vụ trực tuyến nhắm mục tiêu đến người dùng cuối như CompuServe, America Online, Prodigy. Các công ty này đi tiên phong trong việc phát triển các dịch vụ tìm kiếm đơn giản cho những người tìm kiếm thông tin không phải là chuyên gia, mở rộng bộ sưu tập tài liệu thích hợp cho mục đích sử dụng hàng ngày như đi lại, mua sắm, và sở thích, họ giáo dục cho công chúng nhận thức được sức mạnh và tiện lợi của việc tìm kiếm thông tin trực tuyến lần đầu tiên thông qua các kỹ thuật tiếp thị hàng loạt.

Các tổ chức phi lợi nhuận lớn trong ngành công nghiệp bao gồm JSTOR (bộ sưu tập của tạp chí backfiles); OCLC (FirstSearch, WorldCat, netlibrary) và Thư viện nghiên cứu nhóm RLG, mà bây giờ thuộc sở hữu của OCLC (thư viện hồ sơ, bao gồm đại diện đa phương tiện). Trong điều kiện của nhà sản xuất cơ sở dữ liệu, nhiều hiệp hội phi lợi nhuận được dự thi đầu vào ngành công nghiệp khai thác thông tin trực tuyến, nói chung là một hệ quả tự nhiên của hoạt động xuất bản truyền thống của họ.

Chính thời gian dài các cơ quan chính phủ và các sản phẩm của họ đã bao gồm Viện Thông tin Khoa học

và kỹ thuật (CISTI) Canada, Thư viện nông nghiệp Quốc gia (Agricola), Thư viện Y học Quốc gia (MEDLINE), Dịch vụ thông tin kỹ thuật quốc gia (NTIS); Văn phòng Giáo Dục Hoa Kỳ (ERIC).

Ngày nay, với sự tham gia tích cực của thư viện và các trung tâm thông tin, nhất là các thư viện trường học, các viện nghiên cứu thì vai trò của các đơn vị tham gia sản xuất thông tin đã thay đổi. Người dùng tin là các nhà nghiên cứu, các giảng viên ...lại trở thành người sản sinh thông tin để cung cấp cho các nhà sản xuất thông tin trực tuyến này. Và bản thân họ, như một hệ quả tất yếu, cũng quảng cáo và truyền bá thông tin cho tới những người sử dụng khác – đó là sinh viên và đồng nghiệp của họ.

Ở Việt Nam, gần đây xuất hiện hình thức kinh doanh tài liệu trực tuyến thông qua việc mua bản quyền truy cập và bán cho một bên thứ ba hoặc người dùng – hình thức mà một số cơ quan thông tin thư viện lớn đang làm. Cũng có những nhóm/công ty nhỏ tập hợp tài liệu trực tuyến (theo hình thức chia sẻ) rồi bán quyền truy cập như: www.thuvien247.net ; www.tailieu.vn; <http://www.ebook.edu.vn/> ; <http://ebook4u.vn> ... nhưng nhìn chung các nhóm/công ty này còn đang hoạt động đơn lẻ, tài liệu chưa thực sự phong phú và không có bản quyền.

Sản xuất CSDL

Một số công ty, tổ chức, và các cơ quan chính phủ trong ngành công nghiệp này tạo ra những bộ sưu tập cơ sở dữ liệu trên máy tính có thể đọc được các loại thông tin khác nhau. Sản xuất CSDL bằng việc thu thập và đánh chỉ mục nội dung từ nhiều nhà xuất bản chính khác nhau. Không thể liệt kê tất cả các loại thông tin được thu thập vào cơ sở dữ liệu, nhưng các hạng mục chính bao gồm các trích dẫn thư mục, tóm tắt, toàn văn bản của các bài báo có cùng chủ đề và các tạp chí, tài liệu của chính phủ, các báo cáo kỹ thuật, giấy tờ hội nghị, chi tiết kỹ thuật, bản thảo chưa xuất bản, sách, hình ảnh, hình ảnh đồ họa, thông cáo báo chí, và các dữ liệu thị trường, cũng như hàng loạt các dữ liệu số cho khoa học, kinh doanh, sử dụng cá nhân, và các ứng dụng khác. Cơ sở dữ liệu cũng có thể chứa thông tin thư mục, các vụ án của tòa án, dữ liệu tài chính của công ty, báo cáo nghiên cứu thị trường, và các báo cáo phân tích đầu tư.

Trong số hàng ngàn cơ sở dữ liệu công khai có rất nhiều dư thừa và lặp đi lặp lại nội dung. Điều này làm tăng thêm thách thức đối với người tiêu dùng (các thư viện và người dùng cuối) để lựa chọn trong số đó, cũng như một thách thức cho các nhà sản xuất để có sản phẩm của họ nhận thấy trên thị trường.

Ví dụ: Tạp chí Library management có thể tìm thấy ở 4 CSDL mà trường RMIT đã mua bao gồm: Emerald current, Proquest/INFORM complete, Proquest Central, Proquest/INFORM global...mặc dù tìm hiểu kỹ thì thấy thời gian xuất bản online của mỗi CSDL lại khác nhau. Emerald current thì từ 1994 đến nay, các CSDL khác thì từ 1992-2010, phải chăng đó là cách mà các nhà sản xuất tạo lợi thế cạnh tranh và buộc người dùng phải mua CSDL của họ?

Xu hướng hồi cố dữ liệu vẫn đang được các nhà sản xuất nội dung số chú ý, nhiều loại báo, tạp chí đang được số hoá cho đến tận thế kỷ 19 như *Christian Science Monitor* trở lại đến năm 1803 và *New York*

Times trở lại đến 1851. Vì vậy, nhiều cơ sở dữ liệu trực tuyến ngày nay có chứa 25 hoặc nhiều hơn các năm nội dung điện tử có thể truy cập trực tuyến, bao gồm cả văn bản phổ biến theo định dạng HTML hoặc PDF.

Các nhà cung cấp trực tuyến

Kể từ khi lần đầu tiên xuất hiện vào đầu những năm 1970, phân khúc của ngành công nghiệp phân phối thông tin cơ sở dữ liệu trực tuyến đã được gọi bởi các tên gọi khác nhau. Trên toàn thế giới các nhà cung cấp tên tuổi bao gồm các máy chủ trực tuyến, ngân hàng dữ liệu, dịch vụ tìm kiếm, các nhà cung cấp cơ sở dữ liệu trực tuyến, các nhà cung cấp nội dung điện tử, dịch vụ truy vấn thông tin, và dây chuyền sản xuất thông tin số. Các cơ sở dữ liệu được tạo ra bởi các nhà sản xuất, các nhà cung cấp đóng góp giá trị gia tăng bằng cách chế biến, chuẩn bị đóng gói và cung cấp, bằng các khả năng tìm kiếm đặc biệt của các phần mềm, nhà cung cấp dịch vụ tìm kiếm trực tuyến, hoặc phân phối đĩa CD-ROM và các sản phẩm cơ sở dữ liệu khác đến người sử dụng. Họ cũng cung cấp các dịch vụ như đặt hàng tài liệu trực tuyến, nhận yêu cầu trực tuyến, và liên kết hoặc nhóm các cơ sở dữ liệu với cơ sở dữ liệu khác. Các nhà cung cấp cơ sở dữ liệu thông thường phục vụ thị trường thư viện, đặc biệt là doanh nghiệp và các thư viện đại học.

Có một loại nhà cung cấp được gọi là bên cung cấp thứ 3, họ tập hợp danh mục tất cả các CSDL của các nhà sản xuất và cung cấp quyền truy cập tới tất cả các cơ sở dữ liệu họ đã mua. Đó là một trong những hình thức một số cơ quan thông tin thư viện của Việt Nam đang làm như đã nêu ở trên. Một số nhà xuất bản phát hành xuất bản phẩm của họ thông qua nhiều hơn một tập hợp dữ liệu, và từ một năm để có thể thay đổi tập hợp. Do đó, gói tài nguyên thông tin được cung cấp bởi bất kỳ một tập hợp là khác nhau từ một trong những năm tiếp theo. Ba tập hợp lớn tại thời điểm này là EBSCO, Gale, và ProQuest và họ bán quyền truy cập theo năm cho người dùng.

Ví dụ: Gói CSDL IPro do Đại học FPT đặt mua tạo thời điểm mua năm 2011 bao gồm hơn 10,000 sách liên quan đến công nghệ thông tin và được cập nhật hàng tuần những cuốn sách mới cho đến hết năm 2011. Đến 2012, thư viện cần phải mua quyền truy cập năm 2012 để được truy cập đến những cuốn sách xuất bản trong năm 2012...Như vậy, các nhà xuất bản đang luôn tận dụng công nghệ để “làm mới” bộ sưu tập của họ.

Ảnh hưởng từ chính phủ

Ngay từ khi bình minh của kỷ nguyên thông tin trực tuyến, chính phủ Mỹ đã đóng vai trò chính trong nghiên cứu và phát triển, cũng như phát động các ngành công nghiệp. Nhiều cơ quan chính phủ, bao gồm cả Văn phòng phân tích kinh tế, Dịch vụ Nghiên cứu Quốc hội, Bộ Quốc phòng, Bộ Năng lượng, Thư viện Quốc gia nông nghiệp, Thư viện Y khoa Quốc gia (NLM) và các bộ phận khác của Viện Y tế quốc gia, Dịch vụ thông tin kỹ thuật quốc gia, Chứng khoán, Ủy ban trao đổi, Văn phòng Giáo dục Hoa Kỳ (sau này là Bộ Giáo dục Hoa Kỳ), và những người khác đã tạo ra cơ sở dữ liệu quan trọng và được sử dụng rộng rãi. Một số trong số họ đã cung cấp dịch vụ tìm kiếm trực tuyến của họ và ngày nay tất cả họ đều duy trì các trang

web thông tin tìm kiếm thêm. Các cơ sở dữ liệu được gọi là Agricola, ERIC, MEDLINE, và NTIS nằm trong số đầu tiên được cung cấp trực tuyến các dịch vụ phục hồi thương mại đầu tiên của những năm 1970, với tốc độ truy cập tương đối thấp, tất cả vẫn còn được sử dụng rộng rãi đến ngày nay và thường là trong phiên bản miễn phí trên trang web của cơ quan, cũng như các phiên bản có sẵn từ các nhà cung cấp có thêm chức năng thương mại, tính năng, và khả năng tương tác.

Ngoài các cơ quan đặc biệt sản xuất và cung cấp truy cập trực tiếp vào cơ sở dữ liệu trực tuyến, chính phủ Mỹ đã đóng một vai trò quan trọng trong việc xây dựng các ngành công nghiệp thông tin Mỹ là cạnh tranh mạnh mẽ, sáng tạo hàng đầu thế giới. Một số lượng lớn tài liệu theo định hướng xây dựng cơ sở dữ liệu và các công ty viễn thông, các nhà cung cấp trực tuyến được hưởng lợi từ một số loại hỗ trợ trong những năm 1960 và 1970 từ chính phủ Mỹ trong các hình thức hợp đồng dịch vụ hoặc tài trợ nghiên cứu các trường đại học và phòng thí nghiệm tư nhân. Ví dụ, Bộ Quốc phòng và Văn phòng Giáo Dục Hoa Kỳ hỗ trợ nghiên cứu ban đầu trên hệ thống Dialog Lockheed, NLM đóng góp vào sự phát triển của ORBIT cùng với Dialog, sử dụng nhiều nhất các hệ thống đầu thu hồi trực tuyến. Hệ thống ngày nay được biết đến như LexisNexis đã bắt đầu tại Mead, tiền thân được cung cấp bởi dự án Air Force Mỹ. Sự hỗ trợ này rất quan trọng bởi vì những hệ thống này rất đắt tiền, đòi hỏi phải truy cập vào lớn, nhà nước đầu tư cho máy móc và các chuyên gia với những lập trình viên cao cấp để xây dựng hệ thống phân tích, thiết kế cơ sở dữ liệu, và kỹ năng quản lý.

Vì lợi ích kinh doanh nên đã có những sự đấu tranh, nhất là các cơ quan, tổ chức tư nhân về quyền cung cấp thông tin và tính cạnh tranh với các cơ quan do nhà nước đầu tư. Trong tháng 11 năm 2002, các dịch vụ PubSCIENCE trực tuyến do bộ năng lượng sản xuất đã ngưng, PubSCIENCE là một công cụ Web-dựa trên công khai có sẵn để truy cập các bài báo xuất bản trong tạp chí Peer. Thư viện mạnh mẽ ủng hộ tiếp tục dịch vụ, nhưng ngành công nghiệp thông tin vận động chống lại nó, họ tranh cãi rằng đó là cạnh tranh với hai khu vực tư nhân – Scirus (thuộc sở hữu của Elsevier) và Infotrieve. Trong trường hợp này, các lợi ích thương mại chiếm ưu thế.

Việt Nam cũng có những CSDL và một số cơ quan cũng đã xây dựng ngân hàng dữ liệu trả lời khách hàng, ví dụ như dịch vụ 1080, nhưng những CSDL, ngân hàng dữ liệu của Việt Nam còn nhỏ và hoạt động còn manh mún. Một số cơ quan thông tin thư viện được hưởng ngân sách quốc gia để xây dựng các CSDL phổ biến cho công chúng nhưng khả năng và mức độ phổ biến lại yếu ớt và nhỏ lẻ.

Đóng góp của thư viện

Thư viện cũng là thành viên hoặc ít nhất là các đối tác trong ngành công nghiệp thông tin. Ngay từ khi ngành công nghiệp này bắt đầu và cho đến ngày nay, thư viện có chức năng như một lực lượng bán hàng không chính thức và không được các nhà cung cấp, nhà sản xuất cơ sở dữ liệu thanh toán. Họ đã thử nghiệm tính năng mới và cung cấp phản hồi cho các nhà phát triển CSDL. Họ đã dạy cho người sử dụng cuối cùng giá trị của nội dung chất lượng cao và đào tạo họ để tìm kiếm. Các nhà nghiên cứu thư viện đã tiến hành nghiên cứu để so sánh và đánh giá tính năng, chức năng và hiệu quả của CSDL cả về mặt công

nghệ và nội dung. Thư viện cũng xây dựng cổng thông tin và biên mục tài liệu để nâng cao khả năng tiếp cận của người sử dụng. Thư viện còn phân tích nhu cầu thông tin, hành vi và sở thích của người sử dụng để có những cải tiến phù hợp và thúc đẩy nhu cầu thông tin của người dùng. Nhìn chung thư viện mang đến cho cộng đồng người dùng cuối một nguồn dữ liệu có giá trị, đóng góp rất lớn cho sự phát triển và tiếp thị ngành công nghiệp thông tin trực tuyến.

Ví dụ: Các thư viện đại học thường có các khoá học hướng dẫn sử dụng và khai thác cơ sở dữ liệu trực tuyến cho bạn đọc của mình vào đầu mỗi học kỳ và các khoá học nâng cao kỹ năng sử dụng cho giảng viên và cán bộ nghiên cứu vào những thời gian khác nhau theo nhu cầu.

Vậy thư viện được hưởng lợi gì từ công việc của mình?

Quan sát viên

Mỗi năm từ năm 1975 và 2000, Martha E. Williams đã theo dõi sự tăng trưởng hiện tượng của ngành công nghiệp này và báo cáo các số liệu ở tất cả các cuộc họp trực tuyến quốc gia, định kỳ trong *các bài viết trong Bản tin của Hiệp hội Khoa học thông tin và Cơ sở dữ liệu* Directory Gale. CSDL công khai sẵn có tăng trong thời gian đó từ 300 đến 13.000 hoặc thậm chí hơn nữa, nếu tính tất cả các tập tin cá nhân trong những bộ sưu tập của cơ sở dữ liệu được cung cấp bởi các nhà cung cấp như LexisNexis và Westlaw. Williams đã báo cáo rằng số lượng của các nhà sản xuất cơ sở dữ liệu tăng từ 200 đến hơn 4000. Số lượng các nhà cung cấp đã tăng 100-3000. Số lượng các bản ghi trong cơ sở dữ liệu trực tuyến tăng vọt 52.000.000-15.250.000.000

Trong khi tốc độ phát triển trung bình vẫn tương đối ổn định, sự đa dạng của các loại nội dung mở rộng không chỉ dừng lại ở trích dẫn, tóm tắt, mà còn có tin tức, thông tin tiểu sử, tên hóa học và cấu trúc, dữ liệu bất động sản, công thức nấu ăn, chuỗi thời gian, hình ảnh, mô tả, hoặc danh sách của tất cả mọi thứ tưởng tượng, cũng như ngày càng gia tăng trong thập kỷ qua, toàn văn các bài viết đầy đủ của tạp chí và các bài báo, báo cáo kỹ thuật, và các loại khác nhiều tài liệu. Williams cũng báo cáo doanh thu ngành công nghiệp, các nhà cung cấp và cơ sở dữ liệu các số liệu phổ biến nhất, và các ngành công nghiệp chẳng hạn như tốc độ truyền, thời gian tìm kiếm trung bình, và chi phí trung bình cho mỗi tìm kiếm.

Phần 2: Xu hướng hiện tại

Một ngành công nghiệp luôn thay đổi

Đặc điểm nổi bật của ngành công nghiệp này phức tạp và không ổn định. Sự phát triển mạnh mẽ của các nhà sản xuất, các công cụ tìm kiếm sẵn có trên internet như google, yahoo, altavista... làm cho tốc độ khai thác, phổ biến và sản sinh thông tin ngày càng khủng khiếp dẫn đến một số hệ quả như kết quả tìm kiếm thông tin không thể kiểm soát và rất khó đánh giá, sàng lọc – nội dung thông tin sao chép và trùng lặp nhiều. Trước bối cảnh đó, thế hệ web 3.0 dự kiến sắp ra đời sẽ tác động mạnh mẽ trong việc tổ chức, sắp

xếp tìm kiếm và khai thác thông tin có sẵn trên internet một cách trật tự hơn – điều đó sẽ thực sự ảnh hưởng rất mạnh đến các nhà sản xuất thông tin trực tuyến.

Mua bán sáp nhập

Thời gian gần đây có rất nhiều các công ty có sự sáp nhập, bán lại CSDL của mình cho một công ty kinh doanh lớn hơn hoặc có quan hệ đối tác để cùng tham gia vào những dự án đấu thầu cho những cơ quan, tổ chức có nhu cầu sử dụng và khai thác thông tin. Các tên tuổi nổi tiếng trong ngành công nghiệp này như CSA, EBSCO, Dialog, Dow Jones, Factiva, Gale, ProQuest, Reuters, Thomson, và Westlaw... đã tham gia trong trong những giao dịch mua bán, sáp nhập có giá trị hàng triệu đến hàng tỉ đô la. Các công ty LexisNexis, ISI, và BIOSIS kết hợp thành tập đoàn lớn hơn, và họ cũng đã được mua bán nhiều lần khó có thể kể hết. Câu chuyện báo cáo về trường hợp của Dialog được mua, bán và kết hợp với các công ty và các sản phẩm khác là một minh họa sự bất ổn và phát triển không kiểm soát của ngành công nghiệp này.

Thư viện có thường xem các giao dịch mua bán này tác động rất lớn đến ngân sách của thư viện vì mỗi lần sáp nhập, mua bán sẽ có tác động rất mạnh đến giá cả, thay đổi thói quen và tâm lý của người sử dụng, đặt biệt thay đổi nội dung (tăng/giảm) nguồn tin có giá trị và quen dùng đối với người sử dụng. Ví dụ, việc chuyển nhượng của Ngân hàng Thông tin năm 1986 từ máy chủ miễn phí thường trực và sản xuất cơ sở dữ liệu (từ Công ty New York Times) LexisNexis xuống cấp chất lượng của lập chỉ mục toàn văn. Mười năm sau khi đã được chuyển giao quyền sở hữu của LexisNexis để Reed / Elsevier, toàn văn của tờ New York Times đã trở thành có sẵn trên Dow Jones / Tin tức Retrieval, hộp thoại, UMI ProQuest, OCLC, và Westlaw, cùng với một sự phục hồi mạnh mẽ và chính xác lập chỉ mục để bổ sung tìm kiếm từ khóa của toàn bộ văn bản.

Các xu hướng hợp nhất, mua bán của các công ty này là hình ảnh di động và làm suy yếu lòng trung thành thương hiệu.

Khủng hoảng kinh tế, lo ngại ngân sách

Ngành công nghiệp này phát triển từ mức 10-20% đã giảm xuống còn 3% cũng có thể thấy sự chững lại của ngành công nghiệp này. Việc sáp nhập, mua bán, quan hệ đối tác đang diễn biến phức tạp này làm cho ngành công nghiệp này phát triển theo chiều hướng khó đoán. Sự khủng hoảng kinh tế toàn cầu và sự mất lòng tin vào các nền kinh tế của người tiêu dùng cũng gây ra tâm lý đáng lo ngại cho những nhà sản xuất thông tin trực tuyến khi mà việc cắt giảm chi tiêu luôn thường trực trong đầu người tiêu dùng.

Cạnh tranh từ các thành phần mới

Sự suy giảm trong tăng trưởng của ngành công nghiệp thông tin trực tuyến đã không bị ảnh hưởng bởi tất cả mọi người. Công cụ tìm kiếm trên Web như Google, yahoo... đang được hưởng lợi to lớn để tăng trưởng nhờ các công cụ tìm kiếm hỗ trợ người dùng tìm được những thứ họ mong muốn. Sự cạnh tranh từ

Google Scholar, các nguồn như Wikipedia làm cho các công ty thông tin truyền thống lo lắng khi các thư viện tham gia vào việc tìm kiếm và tổ chức các nguồn thông tin trực tuyến miễn phí cho người dùng của họ. Các ứng dụng web 2.0 làm cho người dùng tin cũng có thể tự tổ chức nguồn thông tin cho họ nếu như họ có một kế hoạch tìm kiếm và khai thác thông tin hợp lý. Sự ra đời của web 3.0 với những tính năng ưu việt và những công cụ tổ chức, tìm kiếm thông tin trực tuyến mang tính logic và khả năng chính xác cao sẽ thực sự là thách thức cho các phương pháp tìm kiếm thông tin truyền thống.

Ngành công nghiệp truyền thống đang cố gắng để phân biệt sản phẩm của họ, thông qua việc nhấn mạnh vào chiều sâu, chất lượng và lựa chọn, và các công cụ chuyên ngành. Tuy nhiên, các quan sát viên của ngành công nghiệp này đang dự đoán rằng các công ty phát triển công cụ tìm kiếm sẽ thống trị ngành công nghiệp thông tin trực tuyến trong những năm tới.

Sử dụng nhiều thông tin hơn

Thư viện đã và đang ký hợp đồng sử dụng thông tin trực tuyến nhiều hơn để thông phục vụ người dùng. Các công cụ tìm kiếm thông tin trực tuyến phát triển sẽ làm tăng vai trò của các cán bộ thư viện, những người tham gia vào đàm phán hợp đồng mua bán CSDL, marketing, đào tạo người sử dụng ...

Tuy nhiên, sẽ có nhiều những thay đổi về vai trò của các thành phần tham gia vào ngành công nghiệp này cũng như những hình thức mới sẽ tiếp tục biến đổi theo tốc độ phát triển của khoa học công nghệ.

Thư viện và Người dùng cuối

Mặc dù có rất nhiều người dùng cuối ngày nay cho rằng họ có thể tự tìm kiếm thông tin sẵn có miễn phí trên internet, hoặc những nguồn tin thư viện có cũng tương tự như trên web mà họ có thể tìm kiếm ở bất kỳ đâu. Nhưng trên thực tế thư viện vẫn phải đóng một vai trò quan trọng trong việc xây dựng các kỹ năng đọc, viết thông tin của người dùng cuối.

Vấn đề pháp lý

Căng thẳng tiếp tục về quyền sở hữu của nội dung cơ sở dữ liệu là vấn đề bản quyền tài liệu. Ngay cả khi thông tin luôn sẵn có trong một cơ sở dữ liệu trực tuyến, nhà sản xuất hoặc nhà cung cấp có thể cố gắng để hạn chế quyền của người sử dụng như tải về máy tính hoặc chia sẻ nó với những người khác. Khi nhiều khía cạnh của các vấn đề quyền điện tử vẫn chưa được giải quyết, cán bộ thư viện và người sử dụng lo lắng rằng những thông tin trực tuyến có sẵn về mặt lý thuyết có thể chi phí lưu trữ hoặc hạn chế sử dụng trong một phạm vi nhất định theo quy định pháp luật. Ở Việt Nam, vấn đề bản quyền tài liệu số hoá vẫn là một câu hỏi chưa có lời giải đáp.

Toàn cầu hóa

Trong khi ngành công nghiệp thông tin đã được toàn cầu gần như từ khi thành lập, ngày càng nhiều công

ty sẽ toàn cầu, đặt ra những thách thức nghiêm trọng để phục vụ khách hàng đa quốc gia, địa phương hóa thông tin, làm việc với chính phủ, đối phó với sự kiểm duyệt, hạn chế về tiếp cận theo chính sách của các quốc gia mà Việt Nam không phải là ngoại lệ.

Lập chỉ mục và chất lượng

Sau hơn 30 năm của cuộc thảo luận, tranh luận vẫn tiếp tục trên lập chỉ mục của con người so với chỉ mục tự động so với người dùng tạo ra gắn thẻ, hoặc một số sự kết hợp của cả ba. Một thách thức khác là mở rộng số lượng và chất lượng của lập chỉ mục của thông tin số dạng đồ họa và hình ảnh.

Hơn nữa, ngay cả các nhà cung cấp bảo mật vẫn còn đang bối rối bởi dữ liệu rác (ví dụ, chức năng quét lỗi hoặc sai lầm lập chỉ mục) trong cơ sở dữ liệu của họ, và nội dung đó không phản ánh dữ liệu mà người dùng cần. Đối với một số ứng dụng, hiện tại có nghĩa là đang được cập nhật đến từng phút.

Tìm kiếm trên nhiều định dạng và nền tảng

Hệ thống khai thác thông tin trực tuyến mở đang ngày càng trở lên phức tạp cả về hình thức lưu trữ thông tin: âm thanh, hình ảnh, ký hiệu... trên các phần mềm nền khác nhau (window, oracle, sql...) và các định dạng tệp dữ liệu khác nhau như: PDF, html, jpg, doc... Hơn nữa, các nhà cung cấp cần phải hỗ trợ một loạt các nền tảng như CD-ROM (đang được loại bỏ), nối mạng CD-ROM, Web, PDA, và điện thoại.

Bỏ qua nghiên cứu

Nhiều công ty mới bước vào ngành công nghiệp này đang bị chỉ trích vì bỏ qua tất cả các nghiên cứu khoa học thông tin đã được thực hiện trong quá khứ trên nền tảng phát triển, hành vi người dùng, giao diện, và các khía cạnh khác để tối đa hóa hiệu quả của các truy vấn thông tin.

Kỳ vọng và thái độ của người sử dụng

Các tính năng của hệ thống thông tin trực tuyến thế kỷ 21 đang được nâng cấp như World Wide Web cung cấp các chức năng Web 2.0 cho phép cá nhân, tương tác và cộng đồng trực tuyến. Các liên kết siêu dữ liệu và các bộ lọc cho web đang trở thành mối quan tâm lớn của cộng đồng người sử dụng nguồn thông tin trực tuyến, các thư viện cũng dựa vào các công cụ có sẵn để làm những liên kết đến các thư mục và tạp chí điện tử được cấp phép sử dụng miễn phí nhằm đa dạng hoá nguồn thông tin trực tuyến của đơn vị mình.

Một số tổ chức ở các quốc gia lớn còn mạnh dạn đầu tư vào các mô hình thông tin ảo Second life để thiết lập các giải định thông tin và báo cáo cho thấy có khoảng 7 triệu đô la đã dùng trong lĩnh vực này với khoảng 850,000 người sử dụng.

Kết luận

Sau gần 40 năm tăng trưởng tuyệt vời và sự thay đổi to lớn, ngành công nghiệp cơ sở dữ liệu trực tuyến đang phải đối mặt với những thách thức không nhỏ đó là sự sàng lọc thông tin thực sự có giá trị trong khối thông tin khổng lồ mà họ đã tạo ra và cũng không ít những người còn đang “chết đuối” trong khối thông tin miễn phí trực tuyến mà họ tiếp cận hàng ngày.

Thách thức rất lớn đặt ra với thư viện là họ cần phải tiếp tục mua các nguồn tin trực tuyến để phục vụ người dùng, nhưng họ lại không phải là chủ thể sở hữu, họ chỉ được quyền mua quyền khai thác, sử dụng mà thôi – điều đó có nghĩa là nếu như các sản phẩm thông tin trực tuyến có nhiều thay đổi và biến động thì các thư viện sẽ vẫn phải chạy theo với rất nhiều quan ngại về sở hữu, ngân sách, phổ biến thông tin... Chưa kể đến các vấn đề về thông tin cá nhân, độ chính xác, độ phù hợp, tin cậy, và thân thiện vẫn còn thách thức ngay cả những gã khổng lồ của ngành công nghiệp này.

Các cơ sở dữ liệu trực tuyến và khai thác thông tin trực tuyến là một phần của một ngành công nghiệp thông tin lớn hơn, bao gồm một loạt các sản phẩm và dịch vụ thông tin (phương tiện truyền thông và xuất bản, dịch vụ pháp lý, lập trình, xử lý dữ liệu, viễn thông, và nhiều hơn nữa) cũng được mở rộng và phát triển nhanh chóng. Thực tế cho thấy rằng, thế giới đang thay đổi nhanh chóng đối với tất cả mọi người, điều đó có nghĩa là rất khó phân biệt được vai trò của các thành viên tham gia vào ngành công nghiệp này. Cơ sở dữ liệu trực tuyến và khai thác thông tin trực tuyến là một ngành công nghiệp đã được tái phát minh nhiều lần trong gần 40 năm qua và sẽ chắc chắn nhiều lần hơn thế nữa trong tương lai cùng với sự phát triển của khoa học công nghệ, nhất là công nghệ thông tin và truyền thông.

Tác giả: Hoàng Thị Thu Hương

Trường Đại học FPT

Tài liệu tham khảo

1. **Emma Angus, Charles Oppenheim** (2004). Studies of the characteristics of brand names used in the marketing of the information products and services. II. Internet related services, ***Aslib proceedings***, vol.56, No.1, pp12-23
2. **Hoàng Thị Thu Hương** (2010). Tác động của công nghệ web đến hoạt động thông tin – thư viện trường Đại học, ***Thông tin – tư liệu***, số 3, tr. 2-30
3. **Marc L. Resnick, Raquel Montania** (2003). Perceptions of customer service, information privacy, and product quality from semiotic design features in an online web store, ***International journal of human-computer interaction***, vol. 16(2), pp 211-234
4. **Murray A. Young, Paul L. Sauer** (1996). Organizational learning and online consumer information services, ***Journal of consumer marketing***, vol 13.No.5, pp 35-46

5. **Trudi bellardo Hahn** (2010). Online database and information retrieval services industry, *Encyclopedia of library and information science*, 3rd ed., University of Maryland, College Park, USA DOI: 10.1081/E-ELIS3-120044344