

NGÂN HÀNG DỮ LIỆU

THS. TRẦN XUÂN HÀO
NGUYỄN CÔNG NHẬT
Đại học Vinh

1. Dữ liệu và thông tin

Dữ liệu và thông tin là hai khái niệm khác nhau nhưng thường được dùng lẫn lộn. Đối với một người, một bộ phận của cơ quan hay một hệ thống nào đó, dữ liệu là các số liệu hoặc tài liệu cho trước. Đó có thể là một thông điệp, một văn bản một hình ảnh một lời nói hay tín hiệu nào đó được thể hiện, truyền đạt bằng nhiều hình thức và phương tiện khác nhau. Thông tin là dữ liệu đã được xử lý (chế biến) thành dạng dễ hiểu, tiện dùng, nghĩa là có giá trị đối với người nhận tin trong việc ra quyết định. Dữ liệu được ví như nguyên liệu thô của thông tin. Thông tin do người này, bộ phận này phát ra có thể lại được người khác, bộ phận khác coi như dữ liệu để xử lý thành thông tin phục vụ cho những mục đích khác. Đó là lý do tại sao hai từ dữ liệu và thông tin có thể dùng thay thế cho nhau.

Dữ liệu được coi là nguyên liệu để chế biến thành thông tin dưới dạng tiện dùng và có ích trong việc hỗ trợ ra quyết định. Ngày nay, ngân hàng dữ liệu hay hệ CSDL dựa trên máy tính điện tử là phương tiện chủ yếu để quản lý dữ liệu một cách thành công.

2. Cơ sở dữ liệu

2.1. Khái niệm

Cơ sở dữ liệu (CSDL) là một tập hợp dữ liệu được sắp xếp và có thể truy cập, khai thác, hiển thị thông tin theo một cấu

trúc logic. Chẳng hạn CSDL sách hoặc tạp chí trong các cơ quan thông tin thư viện, với CSDL này người dùng có thể truy cập theo nhiều cách khác nhau để có thể tìm đến một hoặc nhiều tài liệu như tìm theo tác giả, nhan đề, từ khoá,... Trước khi tạo một CSDL, kỹ thuật viên phải tạo đường dẫn thư mục dữ liệu (data dictionary) chứa các thông tin hoặc tính chất logic quan trọng của CSDL.

Cũng có thể hiểu CSDL là một sưu tập các thông tin biểu diễn một số khía cạnh của thế giới thực. Là tập hợp dữ liệu có mối liên kết logic với một số ý nghĩa vốn có. CSDL được thiết kế, xây dựng, lưu trữ dữ liệu cho mục đích riêng, nó là ý muốn của nhóm người sử dụng.

2.2. Cấu trúc logic của CSDL

- Trong tin học, mỗi CSDL có thể có một tệp (file - một đơn vị chứa dữ liệu có tên xác định) hay nhiều tệp có thông tin liên quan lẫn nhau. Đơn vị dữ liệu nhỏ nhất trong CSDL là ký tự (character). Một tập hợp logic các ký tự tạo thành một trường (field), chẳng hạn như trường tác giả hoặc trường từ khoá. Một tập hợp logic các trường tạo thành một biểu ghi (record), tập hợp logic các bản ghi tạo thành một tệp (file), tập hợp logic các tệp tạo thành một CSDL. Dữ liệu được truy cập thông qua cấu trúc logic, chứ không phải cấu trúc vật lý.

CSDL phải là một tập hợp các dữ liệu

mang tính hệ thống chứ không phải là các dữ liệu rời rạc, không có mối quan hệ với nhau. Các dữ liệu này phải có cấu trúc và tập hợp các dữ liệu này phải có khả năng đáp ứng các nhu cầu khai thác của nhiều người sử dụng một cách đồng thời. Đó cũng chính là các đặc trưng của CSDL.

2.3. Phân loại CSDL

Mô hình CSDL là sự trừu tượng hoá môi trường thực. Mỗi loại mô hình CSDL đặc trưng cho một cách tiếp cận dữ liệu khác nhau của những nhà phân tích thiết kế CSDL. Mỗi loại mô hình dữ liệu đều có những ưu điểm và những mặt hạn chế của nó, nhưng vẫn có những mô hình dữ liệu nổi trội và được nhiều người quan tâm nghiên cứu. Dưới đây là tiến trình phát triển của các mô hình dữ liệu.

Vào những năm 60, thế hệ đầu tiên của CSDL ra đời dưới dạng mô hình thực thể kết hợp, mô hình mạng và mô hình phân cấp.

Vào những năm 70, thế hệ thứ hai của CSDL ra đời. Đó là mô hình dữ liệu quan hệ do E.F. Codd phát minh. Mô hình này có cấu trúc logic chặt chẽ. Đây là mô hình đã và đang được sử dụng rộng khắp trong công tác quản lý dữ liệu trên phạm vi toàn cầu. Việc nghiên cứu mô hình dữ liệu quan hệ nhằm vào lý thuyết chuẩn hoá các quan hệ và là một công cụ quan trọng trong việc phân tích thiết kế các hệ CSDL hiện nay. Mục đích của nghiên cứu này nhằm bỏ đi các phần tử không bình thường của quan hệ khi thực hiện các phép cập nhật, loại bỏ các phần tử dư thừa.

Sang thập kỷ 80, mô hình CSDL thứ ba ra đời, đó là mô hình cơ sở dữ liệu hướng đối tượng, mô hình cơ sở dữ liệu phân tán, mô hình cơ sở dữ liệu suy diễn,...

2.4. Yêu cầu đối với CSDL

Như trên đã nói, có nhiều loại mô hình CSDL, mỗi loại có những yêu cầu riêng để đáp ứng nhu cầu người dùng. Nhưng nhìn chung, một CSDL phải có được những ưu điểm sau:

- Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác nhau.
- Khả năng chia sẻ thông tin cho nhiều người sử dụng và nhiều ứng dụng khác nhau cùng một lúc.
- Giảm sự trùng lặp thông tin xuống mức thấp nhất và bảo đảm được tính nhất quán và toàn vẹn dữ liệu.
- Dễ dàng trong việc lưu trữ, trao đổi thông tin.

Tuy nhiên, để đạt được các ưu điểm trên, CSDL đặt ra những vấn đề cần phải giải quyết. Đó là:

- Tính chủ quyền của dữ liệu.
- Tính bảo mật và quyền khai thác thông tin của người sử dụng.
- Tranh chấp dữ liệu.
- Phản ánh được thực tế hiện thời, được cập nhật thường xuyên.
- Đảm bảo dữ liệu khi có sự cố.

Để đáp ứng những yêu cầu đó thì người thiết kế và quản trị CSDL phải là những người hiểu biết về tin học, về các hệ quản trị CSDL và hệ thống máy tính. Họ là người tổ chức CSDL (khai báo cấu trúc CSDL, ghi nhận các yêu cầu bảo mật cho các dữ liệu cần bảo vệ,...), do đó họ phải nắm rõ các vấn đề kỹ thuật về CSDL để có thể phục hồi dữ liệu khi có sự cố. Họ là những người cấp quyền khai thác CSDL, do vậy họ có thể giải quyết được các tranh chấp dữ liệu nếu có.

3. Hệ quản trị CSDL

Để giải quyết tốt tất cả các vấn đề đặt ra cho một CSDL như đã nêu trên: tính

chủ quyền, cơ chế bảo mật hay phân quyền hạn khai thác CSDL, giải quyết tranh chấp trong quá trình truy nhập dữ liệu, và phục hồi dữ liệu khi có sự cố... thì cần phải có một hệ thống các phần mềm chuyên dụng. Hệ thống các phần mềm đó được gọi là hệ quản trị CSDL. Đó là các công cụ hỗ trợ tích cực cho các nhà phân tích, thiết kế CSDL và những người khai thác CSDL. Cho đến nay có khá nhiều hệ quản trị CSDL mạnh được đưa ra thị trường như: SQL-Server, DB2, Sybase, Paradox, Informix, Oracle... với các chất lượng khác nhau. Cũng có thể hiểu hệ quản trị CSDL là một hệ thống phần mềm tổng quát cho phép người dùng định nghĩa, thao tác, và xử lý dữ liệu trong một CSDL để xuất ra những thông tin có nghĩa.

4. Ngân hàng dữ liệu và quản lý dữ liệu

4.1. Hệ thống thông tin

Hệ thống thông tin (HTTT) là một tập hợp các yếu tố có liên quan với nhau cùng làm nhiệm vụ thu thập, xử lý, lưu trữ, truyền đạt và phổ biến thông tin để hỗ trợ việc ra quyết định, phân tích tình hình, lập kế hoạch, điều phối và kiểm soát các hoạt động trong một cơ quan, đơn vị.

Các HTTT có thể là hoàn toàn thủ công hay dựa trên máy tính. Trong một HTTT, ngoài máy tính còn có con người, các phương tiện thông tin liên lạc, các quy tắc, thủ tục, phương pháp và mô hình toán học để xử lý dữ liệu, quản lý, phân phát và sử dụng thông tin. Hầu hết các hệ thống thông tin đều được gọi là hệ thống thông tin quản lý bởi vì nó phục vụ cho công tác quản lý.

4.2. Ngân hàng dữ liệu

Trong HTTT người ta lưu trữ và quản lý dữ liệu trong những kho dữ liệu, đó là nơi cất giữ dữ liệu một cách có tổ chức sao cho có thể tìm kiếm nhanh chóng

những dữ liệu cần thiết. Nếu kho dữ liệu này được đặt trên các phương tiện ghi nhớ của máy tính và được bảo quản và khai thác thông qua các chương trình máy tính thì nó còn được gọi là ngân hàng dữ liệu hay hệ CSDL. Ngân hàng dữ liệu hay hệ CSDL là một hệ thống dùng máy tính để lưu trữ, quản lý tập trung dữ liệu nhằm phục vụ cho nhiều người và nhiều mục đích quản lý khác nhau.

4.3. Hệ thống quản lý dữ liệu

Bản thân kho dữ liệu cùng với con người và các phương tiện để duy trì sự hoạt động của nó tạo thành hệ thống quản lý dữ liệu. Vì tầm quan trọng lớn lao của hệ thống quản lý dữ liệu mà người ta thường ví nó như một "hệ thống nhớ" hay một "bộ nhớ" của cơ quan.

4.4. Phần mềm quản trị ngân hàng dữ liệu

Tuỳ thuộc vào kiến trúc của cơ sở hạ tầng cũng như kiến trúc của ngân hàng dữ liệu cụ thể mà người ta sử dụng phần mềm quản trị ngân hàng dữ liệu tương ứng, việc sử dụng phần mềm quản trị ngân hàng dữ liệu như vậy mang lại hiệu quả trong việc thiết kế, tổ chức lưu trữ cũng như khai thác thông tin.

Để quản trị ngân hàng dữ liệu trong lĩnh vực thông tin-thư viện người ta hay dùng phần mềm Oracle. Oracle Database là hệ quản trị CSDL tập trung vào điện toán lưới với việc cho phép các máy chủ chi phí thấp và ngân hàng dữ liệu xử lý thông tin nhanh, mở rộng với độ tin cậy cao, hỗ trợ ngân hàng dữ liệu, điều phối giao dịch và quản trị nội dung.

Có khá nhiều phiên bản của Oracle như Oracle Database 9i, Oracle Database 10i... Trong đó Oracle Database 11g mở rộng các khả năng liên kết thông tin, tự động hóa trung tâm dữ liệu và quản trị khối lượng công việc.

Người dùng sẽ có thể xử lý các ứng dụng thực hiện giao dịch, lưu trữ dữ liệu và quản trị nội dung khác. Oracle Database 11g với 400 tính năng có những tính năng nổi bật như tự động hóa nhiều tác vụ phân vùng dữ liệu mới mà thông thường được thực hiện thủ công. Nó cũng chứa đựng nhiều cải tiến hoạt động đáng kể cho XML-DB, một tính năng trong CSDL của Oracle, giúp lưu trữ và xử lý dữ liệu dạng XML chuẩn xác, phù hợp với yêu cầu cụ thể của người sử dụng. Một công cụ đáng chú ý nữa là "Oracle Total Recall", giúp quản trị viên truy vấn dữ liệu trong các bảng được phân định như những khoảng thời gian trong quá khứ. Nói cách khác, tính năng này bổ sung "chiều thời gian" vào dữ liệu để phục vụ việc theo dõi, kiểm tra các thay đổi và mức tương thích của chúng. Bản Oracle Database 11g hỗ trợ triển khai một cách hiệu quả trên diện rộng, từ server phiến mỏng đến các máy chủ SMP (modular symmetric multiprocessor - đa xử lý đối xứng hai chiều) lớn nhất và các mạng liên kết.

Oracle Database 11g đáp ứng các tiêu chí trong việc quản lý thông tin trong thư viện điện tử:

- Là hệ quản trị CSDL thực thi trên trên mô hình khách/chủ (client/server).

- Làm việc trên mạng TCP/IP đảm bảo khả năng kết nối mạng toàn cầu và triển khai các dịch vụ liên quan tới chia sẻ và khai thác các nguồn tin điện tử trên thế giới.

- Làm việc trong môi trường Web, bằng tiếng Việt và tiếng Anh hoặc một số ngôn ngữ phổ biến khác.

- Xây dựng theo kiến trúc nhiều lớp, hệ thống bao gồm các phân hệ chức năng và được tích hợp thành một hệ thống thống nhất.

- Là hệ quản trị CSDL mô hình quan hệ.

- Quản trị và giám sát: Cho phép theo dõi và giám sát được mọi hoạt động trên hệ thống (Ai? Làm gì? Vào lúc nào?)

- An ninh hệ thống: Phần mềm phải hỗ trợ nhiều mức và cơ chế đảm bảo an ninh hệ thống khác nhau. Hệ thống cho phép quản lý người dùng thông qua các tài khoản với mỗi thao tác cụ thể tác động lên mỗi loại dữ liệu.

- Vận hành hiệu quả trên CSDL lớn: Đảm bảo làm việc ổn định và tốc độ truy cập với CSDL lớn (trên hàng chục triệu biểu ghi).

- Khả năng sao lưu/khôi phục dữ liệu: Đảm bảo toàn vẹn dữ liệu cho hệ thống. Khả năng mở rộng: Khả năng bổ sung thêm các phân hệ, tính năng, máy trạm và máy chủ với số lượng người dùng không hạn chế.

Khả năng tự quản cao: Cài đặt dễ dàng, người dùng có khả năng tự đặt cấu hình cho hệ thống với trợ giúp tối thiểu của nhà cung cấp.

5. Kết luận

Ngân hàng dữ liệu đóng một vai trò rất quan trọng trong cuộc sống, mỗi ngân hàng dữ liệu cụ thể thực hiện lưu trữ nội dung thông tin cụ thể về một lĩnh vực nào đó. Trong bối cảnh hiện nay việc tạo ra các ngân hàng dữ liệu trong mọi lĩnh vực của cuộc sống đã tạo sự thuận tiện cho người sử dụng khai thác cũng như cập nhật thông tin. Trong hệ thống thư viện số, việc xây dựng một hệ thống ngân hàng dữ liệu về các nội dung sách, tạp chí, các báo cáo khoa học là hết sức quan trọng. Nó sẽ tạo ra một hệ thống thông tin lớn phục vụ được một lượng lớn độc giả. Cùng với một kiến trúc phần cứng cụ thể và lựa chọn một phần mềm quản lý ngân hàng dữ liệu hiệu quả sẽ

(Xem tiếp trang 55)

có 250 thẻ với 10.000 lượt sách báo lưu hành. Từ nhu cầu đọc sách báo của hội viên, Thư viện tỉnh luân chuyển thêm sách cho CLB và mỗi tháng một lần lại kết hợp với CLB tổ chức các buổi nói chuyện chuyên đề về sức khỏe, giới thiệu sách mới, hoặc giới thiệu một số bài thuốc trên báo – tạp chí theo từng chủ đề của buổi sinh hoạt...

Nhiều hội viên còn tham gia viết bài, làm thơ gửi bài đăng trên các báo. Cá biệt có những hội viên mắt kém, cán bộ thư viện đã tìm và đọc giúp tài liệu cho họ, hiểu rõ tâm lý, sở thích của từng hội viên để phát hiện và giới thiệu cho họ những tài liệu phù hợp. Công tác phục vụ bạn đọc người cao tuổi tại CLB có hiệu quả rõ rệt, được hội viên hoan nghênh và tham gia xây dựng trạm sách ngày một phát triển.

Qua gần 10 năm hoạt động đến nay, trạm sách của CLB là địa điểm thu hút được nhiều người đến đọc sách báo. Đã thành thói quen có nhiều hội viên ngày nào cũng đến CLB để gặp gỡ, giao lưu, trao đổi với nhau những kinh nghiệm, những kiến thức đã thu lượm được trong cuộc sống và trong sách báo, cởi mở tâm tình, giúp nhau sống vui, sống khỏe, sống có ích.

Mặc dù tuổi cao nhưng ham muốn được đọc, được hiểu biết cập nhật thông tin của người cao tuổi là rất lớn. Các cụ cần đọc, cần nghe, cần nắm bắt tình hình xã hội qua mọi phương tiện thông tin và sách báo là một kênh hữu hiệu nhất để các cụ cập nhật.

Phục vụ CLB người cao tuổi tuy là một nhiệm vụ mới, nhưng kết quả đạt được đã khẳng định sự cố gắng không nhỏ của Thư viện Thái Bình đối với bạn đọc. Bạn đọc càng tăng thì trách nhiệm của Thư viện tỉnh càng lớn, Thư viện đã làm tốt chức năng phục vụ người cao tuổi của mình.

NGÂN HÀNG DỮ LIỆU

(Tiếp theo trang 23)

tạo nên một hệ thống ngân hàng dữ liệu cho phép bổ sung và khai thác thông tin với hiệu quả cao nhất.

TÀI LIỆU THAM KHẢO

1. Nguyễn Thị Ngọc Mai, Đoàn Thiện Ngân. Lý Thuyết Cơ sở dữ liệu. - Tập 1-2. - H.: Lao động – Xã hội, 2007.
2. Sam R. Alapati and Charles Kim. OCP Oracle Database 11g: New Features for Administrators Exam Guide - Paperback - Nov 15, 2007.
3. Darl Kuhn, Sam R. Alapati, and Arup Nanda - RMAN Recipes for Oracle Database 11g: A Problem-Solution Approach - Paperback - Aug 20, 2007.

Bài viết cung cấp cho người đọc các khái niệm: Thế nào là dữ liệu, thông tin, cơ sở dữ liệu, cấu trúc logic của cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, ngân hàng dữ liệu... Qua bài viết, tác giả cũng phân tích cụ thể sự cần thiết để xây dựng một ngân hàng dữ liệu dùng cho việc lưu trữ, quản lí dữ liệu trong thư viện và cho phép bổ sung và khai thác thông tin với hiệu quả cao nhất; các phần mềm quản trị ngân hàng dữ liệu đáp ứng được các tiêu chí trong việc quản lí thông tin của một thư viện điện tử.

The article provides definitions on data, information, database, logical structure of databases, database management systems, data banks,... The author also analyses the necessity of establishment of a data bank for archiving and managing of library data which allows effective usage of information; softwares for data bank management in order to meet the demands of managing information in a digital library.