

Chủ đề

4

BÀI 10

Giới thiệu các hệ cơ sở dữ liệu

LƯU TRỮ DỮ LIỆU VÀ KHAI THÁC THÔNG TIN PHỤC VỤ QUẢN LÝ

SAU BÀI HỌC NÀY EM SẼ:

- Biết được sự cần thiết phải lưu trữ dữ liệu và khai thác thông tin từ dữ liệu lưu trữ cho các bài toán quản lý.



Các công việc quản lý trong thực tế rất đa dạng: quản lý nhân viên, tài chính, thiết bị,... tại các cơ quan, tổ chức; quản lý chỗ ngồi trên các chuyến bay, tàu xe tại các phòng bán vé; quản lý hồ sơ bệnh án tại bệnh viện; quản lý học sinh và kết quả học tập trong các trường.

Để quản lý kết quả học tập, như em biết, phải quản lý điểm của từng môn học bao gồm điểm đánh giá (ĐĐG) thường xuyên, ĐĐG giữa kì, ĐĐG cuối kì,... Theo em, hoạt động này có cần lưu trữ dữ liệu không? Nếu có, đó là những dữ liệu gì?

1. CẬP NHẬT DỮ LIỆU

Hoạt động

Thảo luận về ghi chép điểm môn học

Giáo viên dạy môn Toán dùng cuốn sổ điểm môn học để ghi lại điểm của từng học sinh lớp 11A (Bảng 10.1). Hãy cùng thảo luận để xác định xem có thể khai thác được những thông tin gì từ sổ điểm môn học này. Ngoài việc ghi điểm vào sổ điểm, có thể có những công việc nào khác?



Việc ghi điểm vào sổ điểm được thực hiện thường xuyên, mỗi khi có ĐĐG thường xuyên, giữa kì hay cuối kì. Việc ghi chép này là gọi là lưu trữ dữ liệu điểm.

Bảng 10.1. Bảng điểm môn Toán lớp 11A (học kì I)

STT	Họ và tên	ĐĐG thường xuyên				ĐĐG giữa kì	ĐĐG cuối kì
1	Dương Hồng Anh	8	7		7	9	8
2	Lương Việt Anh		5	6	5	5	5
3	Nguyễn Kì Duyên	7		8	0	10	9
4	Bùi Quý Dương	5	7		7	6	6
5	Đỗ Hồng Dương		7	8	6	6	7
	...						

Việc ghi chép điểm có thể có sai sót, nhầm lẫn. Vì vậy:

- Có thể phải xoá một dữ liệu điểm. Chẳng hạn xoá điểm 0 trong điểm đánh giá thường xuyên của Nguyễn Kì Duyên vì nghỉ học có lí do.
- Có thể phải sửa một dữ liệu nào đó. Chẳng hạn trong danh sách có tên Dương Hồng Anh nhưng thực tế tên chính xác phải là Dương Hoàng Anh.

Bảng 10.2. Bảng điểm môn Toán lớp 11A (học kì I) sau khi chỉnh sửa

STT	Họ và tên	ĐDG thường xuyên			ĐDG giữa kì	ĐDG cuối kì
1	Dương Hoàng Anh	8	7	7	9	8
2	Lương Việt Anh		5	6	5	5
3	Nguyễn Kì Duyên	7		8	10	9
5	Bùi Quý Dương	5	7	7	6	6
6	Đỗ Hồng Dương		7	8	6	7
	...					

Việc thêm, xoá và chỉnh sửa dữ liệu tương tự như trên là những công việc thường được thực hiện với dữ liệu của tất cả các bài toán quản lí và chúng được gọi chung là *cập nhật dữ liệu*.

2. TRUY XUẤT DỮ LIỆU VÀ KHAI THÁC THÔNG TIN

Rõ ràng việc ghi chép điểm các môn học như trên không chỉ nhằm mục đích lưu trữ. Từ dữ liệu lưu trữ này, có thể:

- Lập được danh sách học sinh với điểm học kì từ cao xuống thấp.
- Tìm kiếm và lập danh sách học sinh có điểm học kì $\geq 7, \geq 8$.

Việc tìm kiếm, sắp xếp hay lọc ra các dữ liệu theo những tiêu chí nào đó từ dữ liệu đã có thường được gọi là *truy xuất dữ liệu*.

Một ví dụ phức tạp hơn về truy xuất dữ liệu là lập bảng điểm cuối học kì I (Bảng điểm lớp 11A) đòi hỏi ghép dữ liệu từ các bảng điểm môn học theo danh sách lớp (Hình 10.1).

DANH SÁCH LỚP 11A

STT	Họ và tên
1	Dương Hoàng Anh
2	Lương Việt Anh
3	Nguyễn Kì Duyên
...	

ĐDG CUỐI KÌ I MÔN TOÁN LỚP 11A

STT	Họ và tên	ĐDG cuối kì
1	Dương Hoàng Anh	8
2	Lương Việt Anh	5
3	Nguyễn Kì Duyên	9
...		

ĐDG CUỐI KÌ I MÔN ĐỊA LÍ LỚP 11A

STT	Họ và tên	ĐDG cuối kì
1	Dương Hoàng Anh	6
2	Lương Việt Anh	8
3	Nguyễn Kì Duyên	7
...		

BẢNG ĐDG CÁC MÔN CUỐI HỌC KÌ I LỚP 11A

STT	Họ và tên	Toán	Tin học	Vật lí	Hoá học	Sinh học	Ngữ văn	Lịch sử	Địa lí
1	Dương Hoàng Anh	8	10	10	9	10	6	8	6
2	Lương Việt Anh	5	6	6	7	6	7	8	8
3	Nguyễn Kì Duyên	9	10	10	10	8	6	7	7
...									

Hình 10.1. Ví dụ truy xuất dữ liệu

Cũng có những công việc khác, chẳng hạn lập bảng phân loại kết quả học tập như Bảng 10.3, đòi hỏi phải phân tích, thống kê, tính toán từ dữ liệu đã có để được thông tin cần thiết. Những công việc kiểu như vậy được gọi là khai thác thông tin từ dữ liệu đã có.

Bảng 10.3. Thống kê kết quả trung bình năm học môn Toán

STT	Phân loại	Số lượng	Tỉ lệ
1	Giỏi	8	20%
2	Khá	16	40%
3	Trung bình	16	40%
4	Yếu	0	0%

Từ bảng điểm của lớp 11A, tương tự như với các môn học, có thể tìm kiếm hoặc thống kê theo các tiêu chí khác nhau, để khai thác thông tin phục vụ đánh giá phân loại, so sánh kết quả các môn học của lớp,...

Quản lý kết quả học tập của học sinh từ bảng điểm môn học đến bảng điểm chung của lớp chỉ là một ví dụ cho thấy nhu cầu lưu trữ dữ liệu và khai thác thông tin rất đa dạng. Trong thực tế, nhiều lĩnh vực hoạt động khác, khối lượng dữ liệu được lưu trữ và khai thác thường xuyên lớn hơn rất nhiều, chẳng hạn ở bệnh viện là dữ liệu về các bệnh nhân đến khám chữa bệnh, các loại thuốc, vật tư y tế được mua, sử dụng; ở ngân hàng là dữ liệu về khách hàng, lượng tiền gửi vào rút ra hàng ngày; ở các trung tâm dự báo thời tiết là các dữ liệu về những thay đổi nhiệt độ, độ ẩm, hướng và cường độ gió;... Chính vì thế, ngày nay dữ liệu đã trở thành đối tượng nghiên cứu của một lĩnh vực khoa học liên ngành mới nổi đó là Khoa học dữ liệu.

- Bài toán quản lý là bài toán phổ biến trong thực tế. Cần phải tổ chức lưu trữ dữ liệu để phục vụ các yêu cầu quản lý đa dạng.
- Dữ liệu lưu trữ có thể được cập nhật thường xuyên, được truy xuất theo nhiều tiêu chí khác nhau để thu được các thông tin hữu ích.



Cập nhật dữ liệu là gì? Tại sao dữ liệu cần được cập nhật thường xuyên?

3. THU THẬP DỮ LIỆU TỰ ĐỘNG



Hầu hết các hoạt động quản lý truyền thống đều phải nhập dữ liệu thủ công. Trong bối cảnh tự động hóa trên cơ sở máy tính, đặc biệt là sự xuất hiện của các hệ thống kết nối vạn vật (IoT) nói riêng và Cách mạng Công nghiệp 4.0 nói chung, rất nhiều hoạt động quản lý đã thực hiện việc thu thập dữ liệu tự động.



Hình 10.2. Mã vạch

Ở các siêu thị lớn, mỗi ngày có hàng nghìn khách mua hàng, nhiều người mua hàng chục mặt hàng trong số hàng trăm mặt hàng có trong siêu thị. Nếu nhập dữ liệu thủ công, phải mất nhiều thời gian cho mỗi đơn hàng. Việc này có thể gây tắc nghẽn tại các quầy thanh toán, làm giảm năng suất bán hàng. Để khắc phục tình trạng này, người ta đã tạo các mã vạch mang thông tin về mặt hàng dán trên bao bì. Tại các quầy thanh toán, người thu tiền không cần nhập dữ liệu, chỉ cần đưa hàng qua đầu đọc mã vạch. Đầu đọc sẽ đọc mã và gửi cho máy tính để lập đơn hàng. Toàn bộ dữ liệu về hàng hóa cùng doanh thu được lưu trữ tự động. Dựa trên các dữ liệu đó, máy tính sẽ giúp lập các báo cáo doanh thu, thống kê, tổng hợp, phân tích để cải thiện hoạt động kinh doanh, báo tồn kho (để đặt thêm hàng khi cần), lượng hàng tồn trên quầy (để bổ sung từ kho) và nhiều thông tin có ích khác để quản lý siêu thị.

Một ví dụ khác đó là việc lưu trữ chỉ số tiêu thụ điện. Người ta thay các đồng hồ đo điện với hộp số cơ khí trước đây bằng công tơ điện tử. Các công tơ điện tử có thể đều đặn gửi về công ty điện lực không chỉ lượng điện tiêu thụ mà còn cả các giá trị tức thời của điện áp, dòng, tần số và độ lệch pha,... Với công tơ điện tử, nhân viên điện lực không cần phải ghi số thủ công hàng tháng rồi nhập vào máy tính, giảm bớt nhiều công sức làm hoá đơn tiền điện. Tuy nhiên lợi ích lớn nhất mà ngành điện có được từ giải pháp này là có thể quản lý kỹ thuật qua phân tích dữ liệu từ các công tơ điện tử liên tục gửi về. Chẳng hạn có thể biết khu vực nào đang mất điện, khu vực nào đang quá tải,... để đưa ra những quyết định phù hợp.

- Quản lý là hoạt động rất phổ biến. Mục đích chính của quản lý là xử lý thông tin để đưa ra các quyết định. Vì vậy việc thu thập, lưu trữ dữ liệu có ý nghĩa quan trọng hàng đầu.
- Việc thu thập dữ liệu tự động mang lại nhiều lợi ích, không chỉ giảm bớt công sức thu thập mà còn cung cấp một khối lượng dữ liệu lớn giúp nâng cao hiệu quả của việc ra các quyết định cần thiết.



Hãy nêu tầm quan trọng của việc thu thập và lưu trữ dữ liệu đối với các bài toán quản lý.



LUYỆN TẬP

1. Quản lý điểm chỉ là một ứng dụng quản lý trong trường học. Hãy tìm thêm các nhu cầu quản lý khác trong nhà trường và chỉ ra hoạt động quản lý đó cần những dữ liệu nào.
2. Người ta thường nói, ở bất cứ nơi nào có một tổ chức là nơi ấy có nhu cầu quản lý. Hãy kể tên một vài bài toán quản lý mà em biết.



VẬN DỤNG

1. Hãy cho một ví dụ về một bài toán quản lý và nêu những dữ liệu mà hoạt động quản lý đó cần thu thập.
2. Tại các trạm bán xăng, việc thu thập dữ liệu về lượng xăng bán và doanh thu được thực hiện như thế nào?