

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

LÊ THỊ YÊN

**KỸ THUẬT TÌM KIẾM ÂM THANH THEO
NỘI DUNG**

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS.TS ĐẶNG VĂN ĐỨC

Thái Nguyên - 2014

MỞ ĐẦU

1. Đặt vấn đề

Ngày nay, con người sống, làm việc và hoạt động giao tiếp thông qua các dữ liệu đa phương tiện. Công nghệ thông tin truyền thông, mạng máy tính và các giao thức truyền thông phát triển mạnh mẽ, kết hợp với khả năng mô tả, đồ họa phong phú của các trình duyệt đã mang lại sự đa dạng về các dữ liệu cho người dùng đầu cuối.

Do đó, đòi hỏi làm thế nào để tổ chức và cơ cấu một lượng rất lớn các dữ liệu đa phương tiện để có thể dễ dàng nhận được thông tin cần thiết một cách nhanh chóng tại bất kỳ thời điểm nào.

Từ đó, cơ sở dữ liệu đa phương tiện được xây dựng để trở thành một công cụ quản lý, lưu trữ và truy cập một lượng lớn các đối tượng đa phương tiện. Đó chính là cơ hội cũng như là nguyên nhân để các công nghệ về cơ sở dữ liệu đa phương tiện phát triển và ứng dụng rộng rãi trong đời sống kinh tế xã hội.

Các dữ liệu đa phương tiện gồm có: văn bản, hình ảnh tĩnh, hình ảnh động, âm thanh, âm nhạc, video... Hiệu quả của các ứng dụng đa phương tiện phụ thuộc vào sức mạnh của cơ sở dữ liệu đa phương tiện, cụ thể là cấu trúc, cách tổ chức, khả năng truy cập nhanh, chính xác... Công nghệ đa phương tiện được ứng dụng trong nhiều trường hợp như: elearning, hội thảo video, thư điện tử, hiện thực ảo, trò chơi điện tử... Việc tìm hiểu bản chất cũng như là các đặc trưng, các thuộc tính, các kỹ thuật số hoá của từng loại dữ liệu đa phương tiện là yêu cầu để triển khai và ứng dụng công nghệ đa phương tiện vào đời sống.

Trong đó, việc tìm hiểu các đặc trưng, phương pháp số hoá, phương pháp trích chọn, tìm kiếm của dữ liệu âm thanh trong cơ sở dữ liệu âm thanh hiện đang được quan tâm đặc biệt bởi các đặc thù của dữ liệu âm thanh như:

đa dạng thông dụng với người dùng, thân thiện với mọi đối tượng, truyền tải một lượng lớn thông tin trong khoảng thời gian ngắn, ứng dụng nhiều trong đời sống, đó chính là lí do tôi chọn đề tài ***“Kỹ thuật tìm kiếm âm thanh theo nội dung”***

2. Đối tượng và phạm vi nghiên cứu

- Các khái niệm cơ bản về cơ sở dữ liệu đa phương tiện.
- Các khái niệm cơ bản về đặc trưng âm thanh.
- Một số kỹ thuật ứng dụng phát triển cơ sở dữ liệu âm thanh.

3. Hướng nghiên cứu của đề tài

- Nghiên cứu giải thuật liên quan đến các kỹ thuật tìm kiếm âm thanh trong cơ sở dữ liệu âm thanh.
- Nghiên cứu giải pháp công nghệ cài đặt chương trình thử nghiệm.

4. Những nội dung nghiên cứu chính

Nội dung nghiên cứu của luận văn bao gồm:

- Giới thiệu về cơ sở dữ liệu đa phương tiện
- Các đặc trưng âm thanh và cơ sở dữ liệu âm thanh
- Xây dựng chương trình thử nghiệm hệ thống tìm kiếm âm thanh.

5. Phương pháp nghiên cứu

Tổng hợp các tài liệu đã được công bố về dữ liệu âm thanh.

Thực nghiệm một số thuật toán biến đổi trong xử lý âm thanh

Nhận xét, đánh giá kết quả thử nghiệm.

6. Ý nghĩa khoa học và thực tiễn

- Luận văn nghiên cứu kỹ thuật tìm kiếm âm thanh theo nội dung.
- Cài đặt thử nghiệm các kỹ thuật xử lý âm thanh.
- Giải quyết bài toán xử lý âm thanh trong cơ sở dữ liệu âm thanh đa phương tiện.

7. Bố cục của luận văn

Luận văn bao gồm 3 chương cùng với phần Mở đầu, phần Kết luận, phần Mục lục, phần Tài liệu tham khảo.

Chương 1: Giới thiệu về cơ sở dữ liệu đa phương tiện

Trình bày một số khái niệm về CSDL đa phương tiện nói chung và CSDL âm thanh nói riêng. Các vấn đề cơ bản được trình bày bao gồm Kiến trúc tổng quan của hệ thống CSDL đa phương tiện, các loại dữ liệu đa phương tiện và mô hình của chúng. Các nhiệm vụ phát triển hệ thống CSDL đa phương tiện. Giới thiệu tình hình nghiên cứu trong và ngoài nước về vấn đề liên quan.

Chương 2: Các đặc trưng âm thanh và cơ sở dữ liệu âm thanh

Trình bày tổng quan một số phương pháp, trích chọn đặc trưng âm thanh. Tiếp theo là nghiên cứu các thuộc tính và đặc trưng chính của âm thanh, bao gồm các đặc trưng trong miền thời gian biên độ, trong miền biến đổi và trong miền ảnh phổ. Các thuộc tính và đặc trưng chính của CSDL đa phương tiện, phân lớp âm thanh phục vụ tìm kiếm dữ liệu âm thanh trong CSDL âm thanh.

Chương 3: Xây dựng chương trình thử nghiệm hệ thống tìm kiếm âm thanh

Giới thiệu bài toán thử nghiệm. Dữ liệu thử nghiệm. Các công cụ phần mềm hỗ trợ phát triển CSDL âm thanh. Thiết kế hệ thống. Viết chương trình thử nghiệm. Dự định sử dụng MatLab để xây dựng chương trình demo.

CHƯƠNG I: GIỚI THIỆU VỀ CƠ SỞ DỮ LIỆU ĐA PHƯƠNG TIỆN

1.1. Các dữ liệu đa phương tiện

Đa phương tiện (multimedia) là một phương pháp giới thiệu thông tin trên máy tính bằng cách sử dụng nhiều phương tiện truyền thông tin như: Text (văn bản), graphic (biểu đồ, đồ thị), animation (hoạt hình), image (ảnh chụp), video (hình ảnh), audio (âm thanh), hoặc kết hợp các media với nhau (video + audio + văn bản diễn giải)... [2]

Người ta thường phân media thành hai loại dựa trên quan hệ của chúng với thời gian. Đó là:

- Static media: Không có chiều thời gian. Thông tin không liên quan tới thời gian. Ví dụ cho loại này là văn bản, hình họa, ảnh chụp.
- Dynamic media: Có chiều thời gian. Thông tin có quan hệ chặt chẽ với thời gian và thông tin phải được trình diễn với thời gian xác định. Ví dụ các loại audio, video, animation, game online...

So với dữ liệu truyền thông như văn bản và số, dữ liệu đa phương tiện có một số đặc điểm rất khác biệt, đó là:

- *Kích thước và số lượng dữ liệu đồ số*

- Kích thước dữ liệu lớn: dữ liệu đa phương tiện có kích thước lớn hơn nhiều so với các kiểu dữ liệu số và văn bản thông thường. Một văn bản thô có 200 từ (khoảng 1000 ký tự) chỉ có kích thước là 1kByte, nhưng nếu lưu văn bản đó bằng định dạng ảnh GIF thì kích thước gấp khoảng 10 lần. Một giọng nói đơn sắc được lưu với định dạng .WAVE trong thời gian 1 phút có kích thước khoảng 2640 kByte (đã nén) hoặc xấp xỉ 6-8 MB (chưa nén). Một cảnh video rất ngắn chứa hàng trăm bức ảnh với kích thước có thể lên đến hàng chục MB..., xem bảng minh họa:

| Kiểu | Mô tả | Kích thước |
|---------------|---|------------|
| Plain text | khoảng 200 từ (1000 ký tự) | 1 kByte |
| Tệp Winword | khoảng 200 từ (1000 ký tự) | 15 kByte |
| Ảnh GIF | khoảng 200 từ (1000 ký tự, 210 x 100mm) | 10 kByte |
| Âm thanh WAVE | Giọng nói (1 phút, 22KHz, 16 bit, mono) | 2640 kByte |

- Số lượng dữ liệu đồ sộ: người ta ước tính, chỉ riêng trên WWW có số lượng lên đến hàng tỉ ảnh, hàng trăm triệu bài hát MP3 và vài chục triệu phim video.

- *Một số dữ liệu đa phương tiện phụ thuộc thời gian*

Audio và video có thêm chiều thời gian. Khi trình diễn audio và video thì chất lượng của chúng phụ thuộc chặt chẽ vào tốc độ trình diễn. Ví dụ, video phải được trình diễn với tốc độ 25 đến 30 hình/giây để có thể cảm nhận được hình ảnh chuyển động trơn tru.

- *Tìm kiếm dựa trên cơ sở tương tự*

Trong cơ sở dữ liệu quan hệ, phương pháp tìm kiếm truyền thống đối với dữ liệu dạng văn bản và số là tìm kiếm chính xác, hay còn gọi là "exact search". Đối với dữ liệu đa phương tiện, người dùng thường đặt ra yêu cầu tìm kiếm một đối tượng tương tự theo nội dung mà họ đưa ra. Ví dụ, một nghiên cứu khoa học cho biết con người có khả năng nhận biết một bài hát thông qua giai điệu (humming) tốt hơn thông qua tên bài hát. Mặt khác, có rất nhiều bài hát có cùng tên và chỉ khác nhau về giai điệu. Vì vậy, việc tìm kiếm một bài hát dựa trên giai điệu sẽ đáp ứng tốt hơn nhu cầu đầy tiềm năng của ngành công nghiệp giải trí.

Tuy nhiên, việc tìm kiếm tương tự có thể phải dựa trên các đặc trưng phức tạp (ví dụ, video có thể chứa văn bản, âm thanh, hình ảnh...).

- *Đồng bộ*

Một số ứng dụng đa phương tiện sử dụng hệ thống thời gian thực. Hệ thống thời gian thực là hệ thống mà trong đó sự đúng đắn của việc thực hiện

thao tác không chỉ phụ thuộc vào việc thu được kết quả đúng mà còn phải đưa ra kết quả đúng thời điểm. Ví dụ, các tệp phim, bài giảng, truyền hình trực tiếp, hội nghị, hội thảo qua mạng (video conference), xem video theo yêu cầu (video on demand) ... thì yêu cầu hình ảnh phải được đồng bộ với âm thanh.

- *Chất lượng dịch vụ (Quality of Service- QoS)*

QoS là một tập các yêu cầu về chất lượng đối với các hoạt động tổng thể chung của một hoặc nhiều đối tượng. Các tham số QoS mô tả tốc độ và độ tin cậy của việc truyền dữ liệu như thông lượng, trễ, tỷ lệ lỗi... Các ứng dụng đa phương tiện khi truyền qua mạng thường đòi hỏi yêu cầu cao về QoS, nhất là các dịch vụ đa phương tiện tương tác thời gian thực như điện thoại internet, hội thảo qua mạng. Các dịch vụ này thường đòi hỏi khắt khe về độ trễ (tối đa là vài trăm ms). Để xác định QoS, người ta dựa vào các tham số sau đây:

- Độ trễ: là khoảng thời gian cực đại để truyền dữ liệu.
- Jitter: là độ biến đổi độ trễ.
- Thông lượng: là tổng số dữ liệu cực đại được truyền đi trên một đơn vị thời gian.
- Tỷ số mất tin: là số dữ liệu cực đại bị mất trên một đơn vị thời gian.

1.2. Tổng quan về cơ sở dữ liệu đa phương tiện

1.2.1. Khái niệm

Hệ thống quản trị cơ sở dữ liệu đa phương tiện là hệ thống tổ chức và lưu giữ, bao gồm các dữ liệu truyền thông và các loại dữ liệu trù tượng.

Một định nghĩa khác, theo Libor Janek và Goutham Alluri, hệ thống quản trị cơ sở dữ liệu đa phương tiện là một cơ cấu tổ chức quản lý các kiểu dữ liệu khác nhau, có khả năng thể hiện trong các định dạng trên một phạm vi các nguồn phương tiện đa dạng. [2]

Lượng dữ liệu đa phương tiện phát sinh theo nhu cầu hiện nay được lưu trữ là một con số khổng lồ. Chỉ riêng với dữ liệu video, người ta ước tính có

khoảng 21264 trạm truyền hình phát 16 giờ hàng ngày, sinh ra khoảng 31 tỉ giờ. Tuy nhiên, các hệ quản trị cơ sở dữ liệu đã được sử dụng rộng rãi như cơ sở dữ liệu quan hệ, chủ yếu tập trung vào quản lý các tài liệu văn bản thì không đáp ứng đầy đủ đối với việc quản lý các dữ liệu đa phương tiện, bởi các tính chất cũng như các yêu cầu đặc biệt của chúng như đã nêu ở trên. Do đó, hệ thống quản trị cơ sở dữ liệu đa phương tiện là sự cần thiết để quản lý dữ liệu đa phương tiện một cách có hiệu quả.

1.2.2. Kiến trúc cơ sở dữ liệu đa phương tiện (MMDBMS)

Phát triển một MMDBMS bao gồm các bước sau:

➤ **Bước 1. Thu thập media**

Các dữ liệu media được thu thập từ các nguồn khác nhau như ti vi, CD, www...

➤ **Bước 2. Xử lý media**

Mô tả các đoạn trích media và các đặc trưng của chúng, bao gồm cả lọc nhiễu và tách thô...

➤ **Bước 3. Lưu trữ media**

Dựa vào yêu cầu cụ thể của ứng dụng để lưu dữ liệu và các đặc trưng của chúng vào hệ thống.

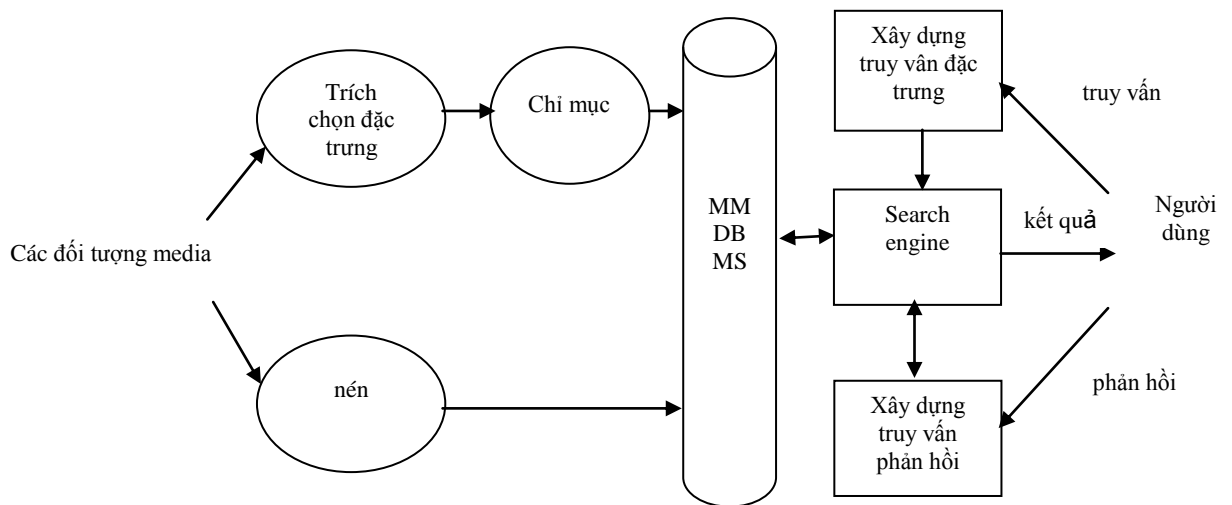
➤ **Bước 4. Tổ chức media**

Tổ chức các đặc trưng để phục vụ việc truy tìm. Ví dụ, chỉ mục các đặc trưng với các cấu trúc giúp khai thác hiệu quả.

➤ **Bước 5. Xử lý truy vấn media**

Là quá trình làm cho thích nghi với cấu trúc chỉ mục. Thiết kế các giải thuật tìm kiếm hiệu quả.

Kiến trúc chung cho một MMDBMS được minh họa như sau:



Hình 1.1: Kiến trúc chung của một MMDBMS

Hệ thống cơ sở dữ liệu đa phương tiện có nhiều môđun chức năng khác nhau nhằm hỗ trợ các thao tác trên dữ liệu đa phương tiện. Bao gồm các môđun chính sau đây:

- Giao diện người dung.
- Bộ trích chọn đặc trưng.
- Chỉ số hóa và môtr tìm kiếm.
- Quản lý truyền thông.

Trong đó, có hai thao tác cơ bản là:

- *Bổ sung dữ liệu đa phương tiện mới*

Thao tác bổ sung được thực hiện theo trình tự các bước như sau:

- Bước 1. Dữ liệu đa phương tiện mới được bổ sung thông qua nhiều cách khác nhau như nhập trực tiếp từ bàn phím, từ microphone hay từ bất kỳ thiết bị nhập kỹ thuật số khác. Dữ liệu đa phương tiện cũng có thể được lấy từ các tệp đã lưu sẵn.
- Bước 2. Sau khi dữ liệu đa phương tiện được bổ sung, nội dung của chúng được trích chọn bằng công cụ trích chọn đặc trưng.
- Bước 3. Các dữ liệu đa phương tiện được bổ sung cùng với các đặc trưng của nó, thông qua bộ quản lý truyền tin được gửi về máy chủ.

- Bước 4. Tại máy chủ, các đặc trưng được bố trí về các vị trí phù hợp dựa vào lược đồ chỉ số hóa.

- Bước 5. Các dữ liệu đa phương tiện bổ sung cùng với các đặc trưng và chỉ số hóa phát sinh được lưu vào bộ quản lý lưu trữ.

- *Truy vấn*

Thao tác truy vấn được thực hiện theo trình tự các bước như sau:

- Bước 1. Tại giao diện người dùng, người sử dụng truy vấn thông tin thông qua một thiết bị nhập nào đó, thông qua tệp đã được lưu trước đó hoặc có thể lấy trực tiếp từ cơ sở dữ liệu MMDBMS.

- Bước 2. Nếu truy vấn của người sử dụng không được lấy trực tiếp từ cơ sở dữ liệu trong MMDBMS thì thực hiện như sau:

- + Thực hiện trích chọn đặc trưng truy vấn.

- + Gửi các trích chọn đặc trưng đó đến máy chủ.

- + Mô-đun chỉ số hóa tìm kiếm các mục dữ liệu phù hợp với truy vấn trong cơ sở dữ liệu.

- + Hiển thị kết quả đến người sử dụng thông qua giao diện người dùng.

1.2.3. Đặc trưng của một cơ sở dữ liệu đa phương tiện

Các đặc trưng chủ yếu của MMDBMS bao gồm:

- *Quản lý dữ liệu đa phương tiện đã được lưu trữ*: các dữ liệu đa phương tiện được lưu trữ để quản lý gồm cả các thiết bị bên trong và bên ngoài máy tính, ví dụ dữ liệu lưu trữ trên CD ROM...

- *Các phương pháp tìm kiếm dựa theo mô tả*: ví dụ, người dùng có thể đưa ra một mô tả để tìm kiếm "tiếng chuông điện thoại"...

- *Giao diện người dùng độc lập với thiết bị*: người dùng không cần biết cách thức lưu trữ dữ liệu đa phương tiện như thế nào.

- *Giao diện người dùng độc lập với các định dạng*: các truy vấn dữ liệu đa phương tiện có thể độc lập với định dạng dữ liệu. Nó cho phép có