

**Nội dung bài thi***(Được sử dụng tài liệu cá nhân, trừ các thiết bị số)*

**Bài 1:** Trong 1 hệ thống MoD (Movie-On-Demand), có 1 đĩa điều khiển và 5 đĩa từ được dùng để lưu dữ liệu (D0, D1, D2, D3, D4), sử dụng phương pháp lưu trữ RAID-5 với 2 cluster. Mỗi cluster đều gồm 5 đĩa. Cluster 1 có đĩa kiểm tra (parity disk) nằm ở đĩa số  $(i \bmod 5)$ . Cluster 2 có đĩa kiểm tra (parity disk) nằm ở đĩa số  $((i+2) \bmod 5)$ . Trong mỗi cluster, kỹ thuật striping được bắt đầu ở đĩa kế tiếp đĩa kiểm tra. Trong đó,  $i$  là số cuối của mã số sinh viên

- a) Minh họa cách lưu trữ dữ liệu trên hệ thống này cho 3 movie M1 (8 blocks), M2 (7 blocks), M3 (5 blocks). M1 đặt trên cluster 1, M2 và M3 đặt trên cluster 2.
- b) Giả sử 1 đoạn dữ liệu ở D4 bị hỏng, dữ liệu ở các bit tương ứng ở các đĩa khác như sau. Hãy khôi phục lại dữ liệu ở đĩa D4, chỉ rõ cách xác định.

D0	D1	D2	D3	D4
1	0	1	0	?
1	0	1	1	?
0	1	1	0	?
1	0	1	1	?
1	1	0	1	?

**Bài 2:** Cho 1 ảnh đa mức xám được biểu diễn bởi ma trận dưới đây.

- a) Sử dụng phương pháp lấy ngưỡng để thực hiện phân đoạn ảnh thành 2 vùng. Giá trị ngưỡng chọn = 10.

0	1	2	1	1	0	0	0
4	0	0	0	0	10	10	10
3	2	1	20	20	40	40	35
1	1	1	20	20	40	40	35
20	20	20	20	35	1	1	2
20	20	35	35	35	35	4	1
20	20	35	35	35	35	10	10
0	0	0	0	35	10	10	10

- b) Sử dụng kết quả phân đoạn ở trên, hãy tính ma trận đồng xuất hiện với khoảng cách  $d = 1$  và hướng  $\beta = 0^\circ$  cho vùng ảnh của các điểm ảnh có giá trị mức xám lớn hơn ngưỡng.

- c) Khi các ảnh được biểu diễn bằng đặc trưng cục bộ, hãy chỉ ra các cách bạn có thể dùng để đối sánh 2 ảnh.

**Bài 3:**

Giả sử các video về giảng bài trực tuyến được lưu trữ, hoạt động **Giảng Bài** và **Hỏi Đáp** được phát hiện và biểu diễn bởi:

GiảngBàì: {(GiảngViên, Person), (ChủĐề, String) }

HỏiĐáp: {(NgườiHỏi, Person), (NgườiĐáp, Person), (CâuHỏi, String), (CâuTrảLời, String)}

Giả sử dữ liệu được quản lý/lưu trữ trong hệ CSDL với cơ sở dữ liệu video: vdl1 xây dựng theo nguyên tắc tự. Hãy biểu diễn truy vấn sau bằng ngôn ngữ SQL mở rộng:

- Đưa ra các đoạn video có bài giảng về chủ đề "Khí hậu" được giảng dạy bởi GS.T.V.Trung.
- Cho biết các đoạn video có sự tham gia của GS. L.V.Lang trong các hoạt động Giảng Bài hoặc Hỏi Đáp.
- Đưa ra các đối tượng có tham dự sự kiện bài giảng của GS.T.V.Trung và là người hỏi trong 1 sự kiện hỏi đáp nào đó.

**Bài 4:** Cho tập văn bản với tiêu đề tương ứng

D1	<i>Không làm phức tạp tình hình Biển Đông</i>
D2	<i>Miễn visa cho công dân 5 nước Tây Âu</i>
D3	<i>Chủ tịch HĐND TPHCM "đổi chất" Bộ trưởng Thăng về thu phí xe máy</i>
D4	<i>Mỹ chỉ trích Trung Quốc đưa vũ khí đến Biển Đông</i>
D5	<i>Vũ khí Trung Quốc được phát hiện trên đảo nhân tạo Biển Đông</i>

- Loại bỏ các từ dừng (theo quan điểm cá nhân) và biểu diễn mỗi văn bản bằng 1 tập các từ
- Giả sử file đảo (inverted file) được sử dụng để đánh chỉ mục cho cơ sở dữ liệu này, sinh viên hãy đưa ra thông tin tương ứng với các từ « Biển Đông », « vũ khí » được lưu trong file này. *Lưu ý : có thể có các đáp án khác nhau, sinh viên cần chỉ rõ lựa chọn của mình.*
- Chọn 1 cách để biểu diễn văn bản D1, D4, D5. Cho biết văn bản nào «gần» với D5 nhất. Giải thích

**Bài 5:** Đặc trưng cục bộ được tính trên một số ảnh cần tổ chức lưu trữ dưới dạng cây kD-trees. Đặc trưng có số chiều là 3. Giá trị của các đặc trưng và thứ tự dữ liệu thêm vào cây như sau:

F1(120, 135, 5), F2(100,100, 5), F3(90, 70, 80), F4 (130, 120, 10), F5(140, 100, 90), F6(154, 100, 5), F7(80, 90, 100), F8(70, 80, 20), F9(60, 130, 40), F10(150, 140, 15)

- Biểu diễn cây kD cho dữ liệu trên
- Loại bỏ nút F1, xây dựng lại cây.

**Nội dung bài thi**

(Được sử dụng tài liệu cá nhân, trừ các thiết bị số)

**Bài 1:** Trong 1 hệ thống MoD (Movie-On-Demand), có 1 đĩa điều khiển và 5 đĩa từ được dùng để lưu dữ liệu (D0, D1, D2, D3, D4), sử dụng phương pháp lưu trữ RAID-5 với 2 cluster. Mỗi cluster đều gồm 5 đĩa. Cluster 1 có đĩa kiểm tra (parity disk) nằm ở đĩa số  $(i \bmod 5)$ . Cluster 2 có đĩa kiểm tra (parity disk) nằm ở đĩa số  $((i+2) \bmod 5)$ . Trong mỗi cluster, kỹ thuật striping được bắt đầu ở đĩa kế tiếp đĩa kiểm tra. Trong đó,  $i$  là số cuối của mã số sinh viên

- a) Minh họa cách lưu trữ dữ liệu trên hệ thống này cho 3 movie M1 (8 blocks), M2 (7 blocks), M3 (5 blocks). M1 đặt trên cluster 1, M2 và M3 đặt trên cluster 2.
- b) Giả sử 1 đoạn dữ liệu ở D4 bị hỏng, dữ liệu ở các bit tương ứng ở các đĩa khác như sau. Hãy khôi phục lại dữ liệu ở đĩa D4, chỉ rõ cách xác định.

D0	D1	D2	D3	D4
1	0	1	0	?
1	0	1	1	?
0	1	1	0	?
1	0	1	1	?
1	1	0	1	?

**Bài 2:** Cho 1 ảnh đa mức xám được biểu diễn bởi ma trận dưới đây.

- a) Sử dụng phương pháp lấy ngưỡng để thực hiện phân đoạn ảnh thành 2 vùng. Giá trị ngưỡng chọn = 15.

0	1	2	1	1	0	0	0
4	0	0	0	0	10	10	10
3	2	1	20	20	40	40	35
1	1	1	20	20	40	40	35
20	20	20	20	35	1	1	2
20	20	35	35	35	35	4	1
20	20	35	35	35	35	10	10
0	0	0	0	35	10	10	10

- b) Sử dụng kết quả phân đoạn ở trên, hãy tính ma trận đồng xuất hiện với khoảng cách  $d = 1$  và hướng  $\beta = 0^0$  cho vùng ảnh của các điểm ảnh có giá trị mức xám lớn hơn ngưỡng.

- c) Khi các ảnh được biểu diễn bằng đặc trưng cục bộ, hãy chỉ ra các cách bạn có thể dùng để đối sánh 2 ảnh.

**Bài 3:**

Giả sử các video về giảng bài trực tuyến được lưu trữ, hoạt động **Giảng Bài và Hỏi Đáp** được phát hiện và biểu diễn bởi:

GiảngBàì: {(GiảngViên, Person), (ChủĐề, String) }

HỏiĐáp: {(NgườiHỏi, Person), (NgườiĐáp, Person), (CâuHỏi, String), (CâuTrảLời, String)}

Giả sử dữ liệu được quản lý/lưu trữ trong hệ CSDL ĐPT với cơ sở dữ liệu video: vdl2 xây dựng theo nguyên tắc đồng nhất. Hãy biểu diễn truy vấn sau bằng ngôn ngữ SQL mở rộng:

- Đưa ra các đoạn video có bài giảng về chủ đề "Khí hậu" được giảng dạy bởi GS.T.V.Trung.
- Cho biết các đoạn video có sự tham gia của GS. L.V.Lang trong các hoạt động Giảng Bài hoặc Hỏi Đáp.
- Đưa ra các đối tượng có tham dự sự kiện bài giảng của GS.T.V.Trung và là người hỏi trong 1 sự kiện hỏi đáp nào đó.

**Bài 4:** Cho tập văn bản với tiêu đề tương ứng

D1	<i>Không làm phức tạp tình hình Biển Đông</i>
D2	<i>Miễn visa cho công dân 5 nước Tây Âu</i>
D3	<i>Chủ tịch HĐND TPHCM "đổi chất" Bộ trưởng Thăng về thu phí xe máy</i>
D4	<i>Mỹ chỉ trích Trung Quốc đưa vũ khí đến Biển Đông</i>
D5	<i>Vũ khí Trung Quốc được phát hiện trên đảo nhân tạo Biển Đông</i>

- Loại bỏ các từ dừng (theo quan điểm cá nhân) và biểu diễn mỗi văn bản bằng 1 tập các từ
- Giả sử file đảo (inverted file) được sử dụng để đánh chỉ mục cho cơ sở dữ liệu này, sinh viên hãy đưa ra thông tin tương ứng với các từ « Biển Đông », « vũ khí » được lưu trong file này. *Lưu ý : có thể có các đáp án khác nhau, sinh viên cần chỉ rõ lựa chọn của mình.*
- Chọn 1 cách để biểu diễn văn bản D1, D4, D5. Cho biết văn bản nào «gần» với D5 nhất. Giải thích

**Bài 5:** Đặc trưng cục bộ được tính trên một số ảnh cần tổ chức lưu trữ dưới dạng cây kD-trees. Đặc trưng có số chiều là 3. Giá trị của các đặc trưng và thứ tự dữ liệu thêm vào cây như sau:

F2(100,100, 5), F1(120, 135, 5), F3(90, 70, 80), F4 (130, 120, 10), F5(140, 100, 90), F6(154, 100, 5), F7(80, 90, 100), F8(70, 80, 20), F9(60, 130, 40), F10(150, 140, 15)

- Biểu diễn cây kD cho dữ liệu trên
- Loại bỏ nút F2, xây dựng lại cây.