

1: Dữ liệu đồ họa \_(?)\_ sẽ được xử lý giống ảnh số

A. dựa trên điểm (point-based graphic)

B. dựa trên vector (vector-based graphic)

C. dựa trên đường (line-based graphic)

D. dựa trên pixel (pixel-based graphic)

2: Phương tiện (media) có thể là

A. Tất cả đều đúng

B. Dữ liệu chữ và số (text)

C. Hình ảnh (image)

D. Âm thanh (audio)

3: Phân đoạn (segmentation) là chia một hình ảnh

A. thành các vùng hoặc các đối tượng trong ảnh

B. thành các vùng cùng kênh xám

C. thành các phần bằng nhau

D. thành các vùng đồng nhất

4: Giả sử cho hình như bên dưới (hình gốc và GLCM), có 4 mức xám (0 - 3), giá trị tại ô ?? là gì biết rằng hướng  $a = 0^\circ$  từ trái qua phải, khoảng cách  $d=2$

	0	1	2	3
0	0	1	0	2
1	0	2	1	1
2	3	1	0	0
3	0	0	2	3

Ảnh gốc

	0	1	2	3
0	??			
1				
2				
3				

GLCM

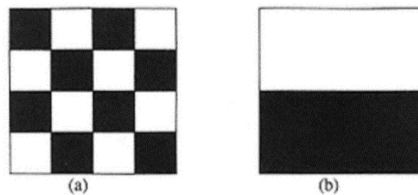
A. 2

B. 1

C. 3

D. 0

5: Hình minh họa bên dưới



Là vấn đề nào đối với phân bố màu

A. Các màu tương tự nhưng thuộc các bin khác nhau

B. Hai ảnh khác nhau có cùng phân bố màu

C. Ảnh hưởng của nền (background)

D. Không tính đến sự gần giống nhau của các màu

6: Với một tập dữ liệu gồm N tài liệu (documents), M từ (token) và bảng tần suất xuất hiện, mỗi văn bản sẽ được biểu diễn bởi 01 vector chỉ

A. sự tương tự với các văn bản khác và có kích thước là M

B. tần suất xuất hiện của các từ trong văn bản và có kích thước là N

C. tần suất xuất hiện của các từ trong văn bản và có kích thước là M

D. sự tương tự với các văn bản khác và có kích thước là N

7: Nếu trong phân đoạn bằng ngưỡng ta dùng n ngưỡng thì số vùng khác nhau sẽ được phân đoạn là

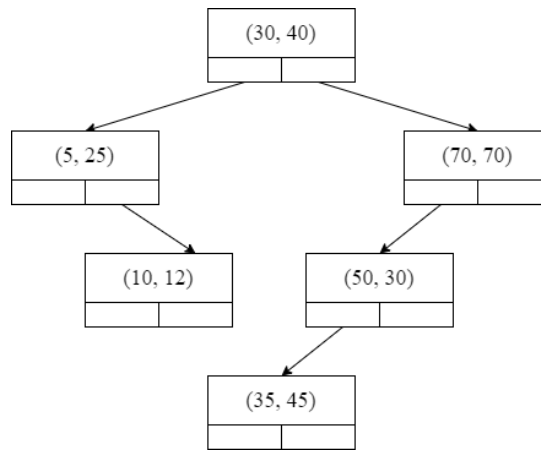
A. 2

B. n

C. n - 1

D. n + 1

8: Cây k-D trong hình sai ở nút nào



**A.** (10, 12)

**B.** (35, 45)

**C.** (5, 25)

**D.** (50, 40)

**9:** Với một bảng các từ trong file đảo như bên dưới:

Term1: D1, 0.3; D2, 0.5; D5: 0.1

Term2: D1, 0.1; D3, 0.6; D4: 0.6

Truy vấn Term1 NOT Term2 sẽ trả về kết quả

**A.** D2, D5

**B.** D5, D2

**C.** D1, D2, D5

**D.** D2, D1, D5

**10:** Theo Tamura, đặc điểm nào không mô tả kết cấu

**A.** Tính giống đường

**B.** Tính trật tự

**C.** Tính vô hướng

**D.** Tính thô ráp

**11:** k-D tree dành lưu trữ

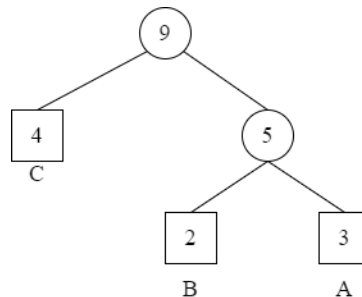
**A.** dữ liệu vùng đa chiều

**B.** dữ liệu điểm và vùng

**C.** dữ liệu đường đa chiều

**D.** dữ liệu điểm đa chiều

**12:** Trong kỹ thuật nén Huffman coding, dựa vào cây bên dưới cho biết mã hóa của các ký tự A, B, C



**A.** A: 11, B: 01, C: 0

**B.** A: 01, B: 00, C: 1

**C.** A: 11, B: 10, C: 0

**D.** A: 00, B: 01, C: 1

**13:** Mô hình LSI

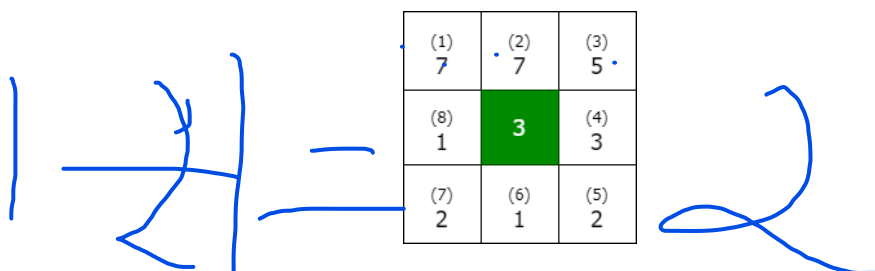
**A.** là một biến thể của mô hình mạng suy luận

**B.** là một biến thể của mô hình vector

**C.** là một biến thể của mô hình Boolean

**D.** là một biến thể của mô hình xác suất

**14:** Giả sử cho ảnh và nhân (seed) được chọn như hình.



Sử dụng 8 kết nối để tìm láng giềng. Tiêu chuẩn tương tự xác định theo công thức  $|f(\text{seed}) - f(\text{neighbors})| \leq 2$ . Vậy các điểm sẽ được gộp vào vùng của nhân gồm những điểm nào

A. (4), (5), (7)

**B. (3), (4), (5), (6), (7), (8)**

C. (4), (5), (6), (7), (8)

D. Tất cả các điểm xung quanh

15: Một ngưỡng cho một vùng của ảnh gọi là ngưỡng

A. hỗn hợp

B. toàn cục

C. thích nghi

**D. cục bộ**

16: Ma trận đồng xuất hiện mức xám (GLCM) xác định

A. tần suất hai giá trị pixel xuất hiện trong một vùng đồng nhất và khoảng cách giữa 2 pixel đó có giá trị nằm trong khoảng xác định

B. tần suất hai giá trị pixel xuất hiện trong một vùng đồng nhất

C. Tất cả đều sai

**D. tần suất hai giá trị pixel xuất hiện trong một môi quan hệ không gian xác định và cách nhau một khoảng cách nhất định**

17: Trong kiến trúc tổng thể của hệ thống truy vấn thông tin (information retrieval), giai đoạn nào tiếp theo sau giai đoạn tìm kiếm (searching)

A. lập chỉ mục (indexing)

**B. xếp hạng (ranking)**

C. biến đổi truy vấn (query operations)

D. phản hồi người dùng (user feedback)

18: Phương tiện có thể được phân loại thành

A. Phương tiện riêng (private media) và phương tiện chung (public media)

B. Phương tiện tĩnh (static media), phương tiện động (dynamic media) và phương tiện hỗn hợp (mixed media)

C. Tất cả đều sai

**D. Phương tiện tĩnh (static media) và phương tiện động (dynamic media)**

19: Stop list là

A. danh sách các mạo từ, giới từ

**B. danh sách các từ không giúp phân biệt các tài liệu trong một tập dữ liệu được xem xét**

C. danh sách các từ xuất hiện ít trong tập dữ liệu

D. danh sách các từ được sử dụng nhiều trong toàn tập dữ liệu

20: Giả sử có các điểm dữ liệu Banja (19, 45), Derventa (40, 40). Cây tứ phân dạng điểm (point quadtree) được khởi tạo với nút gốc là Banja (19, 45)

Banja (19, 45)			
NW	SW	NE	SE

A. NW

B. NE

C. SW

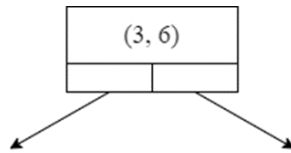
**D. SE**

21: Ý tưởng của phát triển vùng (region growing) bắt đầu với một điểm ảnh và lựa chọn các điểm ảnh để gộp lại theo tiêu chuẩn xác định. Phát triển này

A. Sai

**B. Đúng**

22: Giả sử các điểm dữ liệu trong không gian 2 chiều như sau: (3, 6), (17, 5), (13, 15), (6, 2), (2, 7). Khởi tạo cây với nút gốc là (3, 6) như hình. Nút (17, 5) sẽ được thêm vào



A. Nhánh bên trái của (3, 6) vì giá trị xval của nút (17, 5) > nút (3, 6)

B. Nhánh bên phải của (3, 6) vì giá trị yval của nút (17, 5) < nút (3, 6)

C. Nhánh bên phải của (3, 6) vì giá trị xval của nút (17, 5) > nút (3, 6)

D. Nhánh bên trái của (3, 6) vì giá trị yval của nút (17, 5) < nút (3, 6)

23: Kết cấu có thể được phân thành hai nhóm chính gồm

A. kết cấu đa hướng và kết cấu đơn hướng

B. kết cấu đơn và kết cấu phức

C. kết cấu có cấu trúc và kết cấu ngẫu nhiên

D. kết cấu kênh xám và kết cấu kênh màu

24: Kiến trúc MIRS (hệ thống truy vấn thông tin đa phương tiện) gồm giao diện người dùng (user interface), quản lý truyền thông (communication manager), quản lý lưu trữ (storage manager) và

A. Công cụ lập chỉ mục và tìm kiếm (Indexing and search engine) và Xếp hạng dữ liệu (Ranking engine)

B. Trích xuất đặc trưng (feature extractor) và Công cụ lập chỉ mục và tìm kiếm (Indexing and search engine)

C. Trích xuất đặc trưng (feature extractor) và Xếp hạng dữ liệu (Ranking engine)

D. Công cụ lập chỉ mục và tìm kiếm (Indexing and search engine) và Trình bày dữ liệu (Representation)

25: Trong mô hình Boolean, các phép logic được sử dụng

A. Tất cả đều đúng

B. AND, OR, NOT

C. AND, OR, IF... THEN D. AND, OR, XOR

26: Sử dụng Huffman coding để mã hóa chuỗi m n n n m. Chuỗi trên được mã

A. 0 1 1 1 0

B. Tất cả đều sai

C. 1 0 0 0 1

D. 0 0 1 0 1 0 1 0

27: Phát biểu nào sau đây KHÔNG ĐÚNG với Histogram

A. cách tính dựa trên giá trị điểm ảnh tương ứng với mỗi bin

B. thể hiện tần suất xuất hiện các mẫu màu

C. chỉ có thể được áp dụng cho ảnh mức xám và ảnh nhị phân

D. thể hiện sự phân bố màu sắc trên ảnh

28: Trong biểu diễn văn bản, thông số  $idf$

A. được xác định theo công thức  $idf_i = \log(\frac{df_i}{N})$

B. xác định mức độ quan trọng của mỗi từ trong một tài liệu (document)

C. của một từ có giá trị càng lớn khi nó càng xuất hiện nhiều trong văn bản

D. xác định mức độ quan trọng của mỗi từ trong tập dữ liệu văn bản đang xem xét

29: Trong truy vấn dựa trên nội dung, truy xuất theo đặc trưng ban đầu như màu sắc, kết cấu, hình dạng, vị trí không gian của các thành phần trong hình thuộc cấp độ truy vấn nào

A. Cấp độ 3

B. Cấp độ 2

C. Cấp độ 1

D. Cấp độ 4

30: Phân đoạn bằng ngưỡng thường áp dụng cho ảnh nào

A. Ảnh mức xám

B. Ảnh màu (RGB)

C. Ảnh trắng đen

D. Ảnh bất kỳ

- 31: Nếu cơ sở dữ liệu ảnh có nội dung đa dạng và không đồng nhất thì CSDL đó
- ☒ A. CSDL chung (general database)
  - B. CSDL cụ thể (specific database)
  - C. CSDL toàn cục (global database)
  - D. CSDL cục bộ (local database)
- 32: Trong quá trình xử lý văn bản, bảng tần suất xuất hiện (frequency table) sẽ được lập
- A. sau khi thực hiện kỹ thuật stop list và stemming
  - ☒ B. trước khi thực hiện kỹ thuật stemming và sau stop list
  - C. trước khi thực hiện kỹ thuật stop list và sau stemming
  - D. trước khi thực hiện stop list và stemming
- 33: Trong kỹ thuật nén Run-length Coding, đoạn văn bản **abcddeetttt** sẽ được mã hóa thành
- A. @a1@b1@c1@d3@e2@t4
  - B. abc@d3#@e2#@t4#
  - C. #a1#b1#c1#d3#e2#t4
  - ☒ D. abc@d3@e2@t4
- 34: Trong quá trình xử lý dữ liệu âm thanh, quá trình chuyển đổi từ tín hiệu tuần tự sang tín hiệu số ADC (Analog-to-Digital Converter) gồm các giai đoạn chính:
- A. Lấy mẫu, lọc nhiễu, lượng tử hóa
  - ☒ B. Lấy mẫu, lượng tử hóa, mã hóa
  - C. Thu âm, lọc dữ liệu, mã hóa
  - D. Tất cả đều sai
- 35: Tìm các vùng đồng nhất trong ảnh là cách tiếp cận nào trong phân đoạn ảnh
- A. Tiếp cận biên
  - B. Tiếp cận điểm
  - C. Tiếp cận kết hợp biên - vùng
  - ☒ D. Tiếp cận vùng
- 36: Trong các kiểu dữ liệu phương tiện, dữ liệu nào thường có kích thước nhỏ hơn các kiểu còn lại
- A. Dữ liệu dạng video
  - B. Dữ liệu dạng hình ảnh (image)
  - C. Dữ liệu dạng âm thanh (audio)
  - ☒ D. Dữ liệu dạng văn bản (text)
- 37: Kiểu truy vấn hình ảnh dựa trên nội dung có các truy vấn bằng ví dụ,
- ☒ A. tất cả đều đúng
  - B. bằng màu sắc và bằng phác thảo (sketch)
  - C. bằng bộ phận và bằng phác thảo (sketch)
  - D. bằng bộ phận và bằng hình mẫu
- 38: Bước nào sau đây không thuộc các bước chính của truy vấn dựa trên nội dung
- A. Lựa chọn thước đo độ tương tự
  - ☒ B. Thu thập dữ liệu hình ảnh mẫu
  - C. Đánh chỉ mục
  - D. Trích chọn đặc trưng
- 39: Đặc điểm nào KHÔNG phải là đặc điểm của dữ liệu đa phương tiện
- ☒ A. Cấu trúc ngữ nghĩa rõ ràng để máy tính có thể "hiểu" nội dung
  - B. Giàu thông tin
  - C. Nhiều ứng dụng yêu cầu biểu diễn nhiều kiểu dữ liệu đồng thời
  - D. Kích thước dữ liệu lớn
- 40: Đặc trưng kết cấu (texture)
- A. Tập các điểm được sắp xếp theo một hướng nhất định
  - B. Tất cả đều sai
  - ☒ C. Tập các điểm liên thông với nhau có độ xám gần giống nhau
  - D. Tập các mẫu được sắp xếp theo quy tắc sắp xếp nào đó