

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN ĐỀ THI CUỐI KỲ 2 (2023 – 2024)
KHOA HỆ THỐNG THÔNG TIN Môn thi: Cơ sở dữ liệu
Thời gian: 90 phút

MÃ ĐỀ: 02

(Sinh viên không được sử dụng tài liệu – Đề thi gồm có 02 trang)

Câu 1: Cho lược đồ cơ sở dữ liệu Quản lý mô hình ngôn ngữ lớn (LLM) như sau:

DUAN (MaDA, TenDA, NgayBD, NgayKT, TrangThai, NgPhuTrach)

Tân từ: Lưu trữ thông tin các dự án huấn luyện mô hình ngôn ngữ lớn, bao gồm: mã dự án (MaDA), tên dự án (TenDA), ngày bắt đầu (NgayBD), ngày kết thúc (NgayKT), trạng thái hiện tại (TrangThai) của dự án và họ tên người phụ trách (NgPhuTrach) dự án.

MOHINH (MaMH, TenMH, LoaiMH, MaDA, KichThuoc, CauTruc, NgayTao)

Tân từ: Lưu trữ thông tin các mô hình ngôn ngữ lớn thuộc dự án. Các chi tiết được lưu trữ bao gồm: mã mô hình (MaMH), tên mô hình (TenMH), loại mô hình (LoaiMH), mô hình thuộc dự án nào (MaDA), số lượng tham số/kích thước (KichThuoc) của mô hình theo đơn vị tính là tỷ, cấu trúc (CauTruc) và ngày tạo (NgayTao) mô hình.

BODULIEU (MaBDL, TenBDL, LoaiDL, DungLuong, ChatLuong)

Tân từ: Quản lý thông tin về các bộ dữ liệu dùng để huấn luyện mô hình. Các thông tin được lưu trữ gồm: mã bộ dữ liệu (MaBDL), tên bộ dữ liệu (TenBDL), loại dữ liệu (LoaiDL), dung lượng (DungLuong) có đơn vị tính là Terabyte và chất lượng (ChatLuong) của bộ dữ liệu.

HUANLUYEN (MaMH, MaBDL, NgayHL, NhiemVu, DanhGia)

Tân từ: Quản lý thông tin huấn luyện các mô hình ngôn ngữ lớn. Các chi tiết được quản lý bao gồm: mã mô hình (MaMH), mã bộ dữ liệu (MaBDL), ngày huấn luyện (NgayHL), loại nhiệm vụ (NhiemVu) và chỉ số đánh giá hoàn thành nhiệm vụ (DanhGia) với đơn vị là phần trăm.

TAINGUYEN (MaPBTN, SoGPU, LoaiGPU, DungLgSSD, MaDA, NgayCap)

Tân từ: Lưu trữ thông tin các tài nguyên được phân bổ cho dự án, bao gồm: mã phân bổ tài nguyên (MaPBTN), số lượng GPU (SoGPU), loại GPU (LoaiGPU), dung lượng ổ cứng SSD (DungLgSSD) tính theo đơn vị Terabyte, mã dự án được cấp tài nguyên (MaDA) và ngày cấp (NgayCap).

Lưu ý: thuộc tính được gạch chân là khóa chính của quan hệ.

1. Hãy phát biểu chặt chẽ ràng buộc toàn vẹn sau (bao gồm: bối cảnh, nội dung, bảng tầm ảnh hưởng)^[G2]: (1,5 điểm)

“Những dự án có thời gian hoạt động nhỏ hơn 60 ngày chỉ được phân bổ từ 50 GPU trở xuống”

Lưu ý: không được sửa thuộc tính khóa chính.

2. Thực hiện các câu truy vấn sau bằng ngôn ngữ SQL^[G2]: (6 điểm)

- Tìm thông tin các mô hình (TenMH, CauTruc) thuộc loại *Transformer* được huấn luyện trên những bộ dữ liệu thuộc loại *Code* có chất lượng *trung bình*. (1 điểm)
- Liệt kê những dự án (TenDa, TrangThai, NgPhuTrach) được phân bổ GPU thuộc loại *NVIDIA DGX H200*. (1 điểm)

- c. Trong số các dự án có trạng thái *Tạm dừng*, tìm dự án có số lượng nhiệm vụ *Dịch thuật* được đánh giá thấp hơn 50% nhiều nhất. (1 điểm)
- d. Tìm mô hình (*TenMH*, *LoaiMH*) được tạo từ năm 2023 đã huấn luyện trên tất cả các bộ dữ liệu thuộc loại *Văn bản*. (1 điểm)
- e. Liệt kê thông tin tất cả mô hình (*MaMH*, *CauTruc*) thuộc loại *BERT* và nếu có hãy liệt kê các thông tin về bộ dữ liệu (*TenBDL*) thuộc loại *Văn bản* cùng chỉ số đánh giá hoàn thành nhiệm vụ (*DanhGia*) huấn luyện của mô hình đó. (1 điểm)
- f. Tìm thông tin tài nguyên (*LoaiGPU*) chỉ được dùng để huấn luyện nhiệm vụ *Trả lời câu hỏi*. (1 điểm)

Câu 2: ^[G2] (2,5 điểm) Cho lược đồ quan hệ R (A, B, C, D, E, G, H, I) và tập phụ thuộc hàm như sau:

$$F = \{f_1: EI \rightarrow AC, f_2: HI \rightarrow BG, f_3: EH \rightarrow A, f_4: GI \rightarrow EH\}$$

1. $HI \rightarrow BC$ có là phụ thuộc hàm thành viên của F^+ không? Giải thích. (0,75 điểm)
2. Tìm tất cả các khóa của lược đồ quan hệ trên. Giải thích. (1 điểm)
3. Lược đồ quan hệ trên có đạt dạng chuẩn 2 không? Giải thích. (0,75 điểm)

Ghi chú: ^[G2, G3] là chuẩn đầu ra môn học, xem chi tiết trong Đề cương môn Cơ sở dữ liệu.

~ Hết ~

Duyệt đề Khoa/Bộ môn

Giảng viên ra đề

P. Trường Khoa

KK

Nguyễn Thị Kim Phụng