**Bài thực hành số 7: Phụ thuộc hàm và chuẩn hóa dữ liệu**

**1. Ôn tập kiến thức:**

* **Phụ thuộc hàm (Functional Dependency - FD):**
  + Khái niệm: X -> Y (X xác định Y) nghĩa là với mỗi giá trị của X, chỉ có duy nhất một giá trị của Y tương ứng.
  + Các luật suy diễn Armstrong:
    - Tính phản xạ: Nếu Y ⊆ X thì X -> Y
    - Tính tăng: Nếu X -> Y thì XZ -> YZ
    - Tính bắc cầu: Nếu X -> Y và Y -> Z thì X -> Z
    - Các luật suy diễn khác:
      * Phân tích: Nếu X -> YZ thì X -> Y và X -> Z
      * Hợp: Nếu X -> Y và X -> Z thì X -> YZ
      * Giả tạo: Nếu X -> Y và WY -> Z thì WX -> Z
* **Các dạng chuẩn hóa:**
  + Dạng chuẩn thứ nhất (1NF): Mọi thuộc tính đều có giá trị nguyên tử (atomic).
  + Dạng chuẩn thứ hai (2NF):
    - Ở dạng 1NF.
    - Mọi thuộc tính không khóa (non-key attribute) đều phụ thuộc đầy đủ vào khóa chính.
  + Dạng chuẩn thứ ba (3NF):
    - Ở dạng 2NF.
    - Mọi thuộc tính không khóa đều không phụ thuộc bắc cầu vào khóa chính.
  + Dạng chuẩn Boyce-Codd (BCNF):
    - Mọi phụ thuộc hàm X -> Y đều thỏa mãn X là siêu khóa (superkey).

**2. Ví dụ minh họa:**

* **Quan hệ R(A, B, C, D, E) và tập phụ thuộc hàm F:**

F = { A -> B, B -> C, CD -> E }

* **Xác định khóa chính:**
  + Tìm bao đóng của từng thuộc tính:
    - A+ = {A, B, C}
    - B+ = {B, C}
    - C+ = {C}
    - D+ = {D}
    - E+ = {E}
  + Tìm bao đóng của các tổ hợp thuộc tính cho đến khi tìm được siêu khóa:
    - AB+ = {A, B, C}
    - AC+ = {A, B, C}
    - AD+ = {A, B, C, D}
    - AE+ = {A, B, C, E}
    - ...
    - ACD+ = {A, B, C, D, E} (siêu khóa)
  + Khóa chính tối thiểu: AD
* **Kiểm tra dạng chuẩn và phân rã (nếu cần):**
  + 1NF: Thỏa mãn vì mọi thuộc tính đều có giá trị nguyên tử.
  + 2NF: Vi phạm vì C phụ thuộc một phần vào khóa chính AD (C chỉ phụ thuộc vào A).
  + Phân rã thành 2NF:
    - R1(A, B, C) với F1 = { A -> B, B -> C }
    - R2(A, D, E) với F2 = { AD -> E }
  + 3NF: R1 và R2 đều thỏa mãn 3NF.
  + BCNF: R1 và R2 đều thỏa mãn BCNF.

**3. Hướng dẫn từng bước:**

1. **Xác định phụ thuộc hàm:** Liệt kê tất cả các phụ thuộc hàm tồn tại trong quan hệ dựa trên ngữ nghĩa của dữ liệu.
2. **Tìm bao đóng và khóa chính:** Sử dụng các luật suy diễn Armstrong để tìm bao đóng của từng thuộc tính và các tổ hợp thuộc tính, từ đó xác định khóa chính (siêu khóa tối thiểu).
3. **Kiểm tra dạng chuẩn:** Kiểm tra xem quan hệ có thỏa mãn các điều kiện của từng dạng chuẩn hay không.
4. **Phân rã (nếu cần):** Nếu quan hệ không thỏa mãn dạng chuẩn mong muốn, thực hiện phân rã quan hệ thành các quan hệ nhỏ hơn để đạt được dạng chuẩn đó. Đảm bảo phân rã bảo toàn phụ thuộc và không mất mát thông tin.

**4. Bài tập thực hành:**

* **Cung cấp cho sinh viên một số quan hệ và tập phụ thuộc hàm**, yêu cầu:
  + Xác định khóa chính.
  + Kiểm tra dạng chuẩn.
  + Phân rã quan hệ để đạt được dạng chuẩn mong muốn (nếu cần).
  + Giải thích các bước thực hiện và lý do phân rã.
* **Yêu cầu sinh viên thảo luận về các vấn đề có thể xảy ra** trong quá trình chuẩn hóa và cách giải quyết.
* **Khuyến khích sinh viên tìm hiểu thêm** về các dạng chuẩn khác và các kỹ thuật chuẩn hóa nâng cao.

**5. Đánh giá:**

* Đánh giá dựa trên sự chính xác, đầy đủ và hiệu quả của quá trình xác định phụ thuộc hàm, khóa chính, kiểm tra dạng chuẩn và phân rã quan hệ.
* Đánh giá khả năng giải thích và bảo vệ các bước thực hiện của sinh viên.
* Đánh giá sự tham gia và đóng góp của sinh viên trong thảo luận nhóm.

**Lưu ý:**

* Giảng viên có thể điều chỉnh nội dung và yêu cầu của bài thực hành tùy theo trình độ và mục tiêu của lớp học.
* Nên cung cấp cho sinh viên tài liệu tham khảo về phụ thuộc hàm và chuẩn hóa.
* Khuyến khích sinh viên sử dụng các công cụ trực quan hoặc bảng tính để hỗ trợ quá trình tìm bao đóng và phân rã quan hệ.
* Bài thực hành này giúp sinh viên hiểu rõ tầm quan trọng của chuẩn hóa trong thiết kế cơ sở dữ liệu, từ đó tránh được các vấn đề về dư thừa và bất nhất dữ liệu.