## 🔹 2.1 Truy vấn loại 1: Textual-KIS

* **Trả lời**: <Tên video>, <Frame Idx>
* **Ví dụ thực tế với data bạn có**:
  + Câu hỏi: "Tìm cảnh người đang mở laptop."
  + BTC cho đáp án: L21\_V002, frame từ **462 → 486**.
  + Bạn trả lời: L21\_V002, 480.
* **Chấm điểm**:
  + Vì 480 ∈ [462, 486] và đúng video L21\_V002 → **R-Score = 1**.
  + Nếu bạn trả 500 → ngoài khoảng → **R-Score = 0**.

## 🔹 2.2 Truy vấn loại 2: Visual Q&A

* **Trả lời**: <Tên video>, <Frame Idx>, <Answer>
* **Ví dụ thực tế**:
  + Câu hỏi: "Ở phút 16s, người phụ nữ mặc áo đỏ đang cầm ly màu gì?"
  + BTC: L21\_V002, frame [486–527], Answer = màu xanh.
  + Bạn trả lời: L21\_V002, 486, màu xanh.
* **Chấm điểm**:
  + Đúng video + đúng frame + đúng đáp án text → **R-Score = 1**.
  + Nếu đúng frame nhưng sai màu → **R-Score = 0**.

## 🔹 2.3 Truy vấn loại 3: Temporal-alignment (TRAKE)

* **Trả lời**: <Tên video>, <Frame\_1>, <Frame\_2>, …, <Frame\_N>
* **Ví dụ thực tế**:
  + BTC yêu cầu tìm 4 khoảnh khắc:
    1. Giơ tay (frame 462)
    2. Vẫy tay (frame 486)
    3. Ngồi xuống (frame 629)
    4. Đứng lên (frame 680)  
       (Epsilon = ±5 frame)
  + Bạn trả lời: L21\_V002, 463, 487, 628, 700