1. **Tổ chức nội dung cho 1 video:**

– Các khía cạnh nào cần được quan tâm khi truy nhập video ?

– Làm thế nào để lưu trữ các đối tượng 1 cách hiệu quả ?

– Ngôn ngữ truy vấn ?

– Làm thế nào để trích chọn được nội dung một cách tự động ?

*Với 1 khung hình (frame) nào đó có:*

– tập các đối tượng và hoạt động liên quan

– Mỗi đối tượng/hoạt động có nhưng thuộc tính nhất định

+ Con người: tên, tuổi

+ Thuyết giảng: AI trình bày về CÁI GÌ ?

Ví dụ: video 1 giờ học

– Con người:

giáo viên

sinh viên: người đặt câu hỏi hoặc thảo luận với nhau

– Hoạt động:

Thuyết giảng: GV trình bày 1 vấn đề gì đó

SV đặt câu hỏi/ trả lời câu hỏi, …

* Ta cần biểu diễn các đối tượng và mối liên hệ giữa chúng

1. **Biểu diễn đối tượng trong 1 video**

**Thuộc tính: (pname, values)**

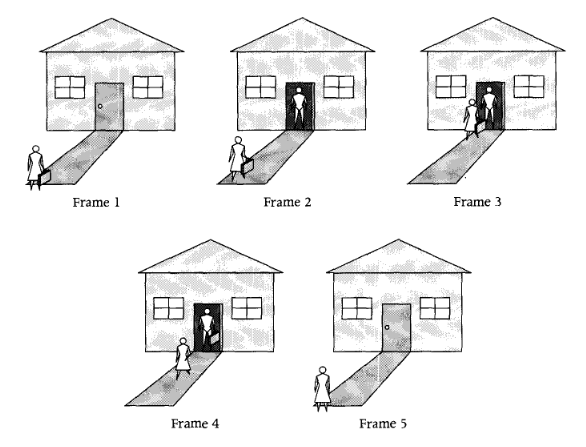
Vd: (height, R+), (primarycolor, {red, green, blue}), ..

**Sơ đồ đối tượng: (fd, fi)**

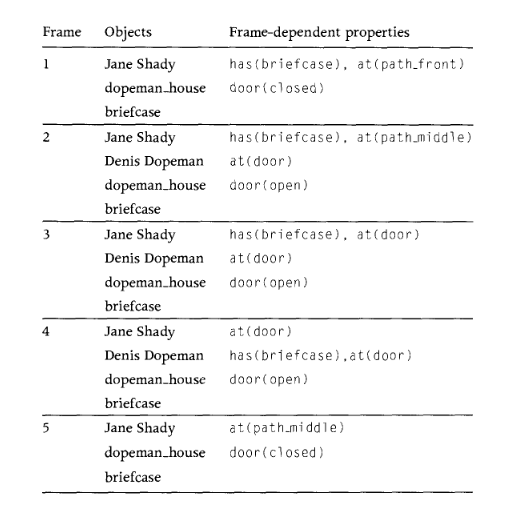
– fd: tập các thuộc tính biến đổi theo mỗi khung hình (framedependent)

– fi: tập các thuộc tính không phụ thuộc vào khung hình (frame-independent)

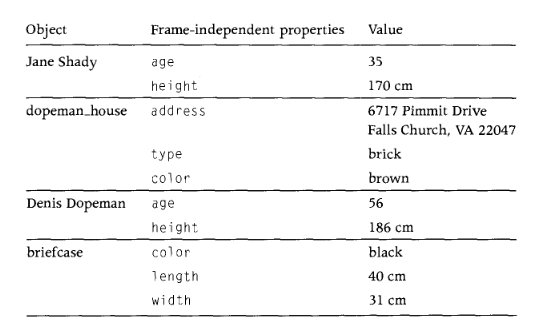
Ví dụ: Xét chuỗi khung hình sau:



Thuộc tính frame-dependent



Thuộc tính frame-independent



1. **Biểu diễn hành động trong 1 video**

Sơ đồ (Activity schema - ACT\_SCH):

*{(pnamei, valuesi), i = 1..n}*

– if tồn tại (pname, value1) , (pname, value2) ➔ value1 = value2

Ví dụ:

ExchangeObject = {(Giver, Person), (Receiver, Person),(Item, Thing)}

– Person: tập tất cả nhân vật trong cảnh

+ Jane Shady, Denis Dopeman

– Thing: tập tất cả các vật có thể trao đổi:

+briefcase

**1 hành động: thể hiện của hành động**

– AcID: activity-id

– Với moi (pname, values) thuộc ACT\_SCH➔ pname = v, v thuộc values

*Ví dụ*

ExchangeObject: Sơ đồ: {(Giver, Person), (Receiver, Person),(Item, Thing)}

1 hành động: Giver = Jane Shady, Receiver = Denis

Dopeman, Item = Briefcase

Lecturing Sơ đồ: {(Lecturer, Person), (Topic, String)}

1 hành động: Lecturer = Prof. Felix, Topic: Quadtree

Questioning Sơ đồ: {(Questioner, Person), (Questionee, Person),

(Question, String), (Answer, String)}

1 hành động:

Questioner = Maria

Questionee = Prof. Felix

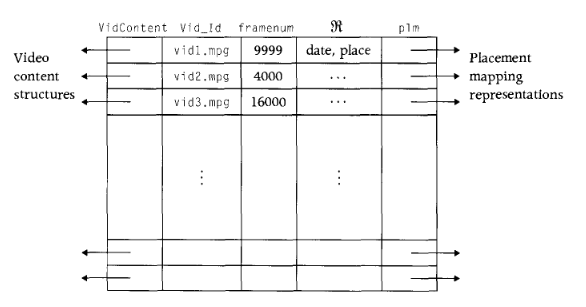
Question = Có bao nhiêu nút trong cây tứ phân

Answer = Nhiều nhất là 4

1. **Biểu diễn video**

+ Video V:  
– framenum(v): số khung hình trong video  
– content(v) = {OBJ, ACT, λ}: nội dung  
+ OBJ = {oid1, oid2,… , oidn}: các đối tượng  
+ ACT = {AcID1, AcID2,… , AcIDk}: tập các sự kiện, hành động  
+ λ: {1, 2, …, framenum(v)} → 2 mũ (OBJ∪ACT)

1. **Biểu diễn thư viện video đơn giản**  
   + VidLib = (Vd\_Id, VidContent, framenum, plm, ℜ)  
   – Vd\_id: tên của video  
   – VidContent: nội dung của video  
   – framenum: số khung hình trong video  
   – Plm (placement maping): chỉ rõ các địa chỉ các phần khác  
   nhau của video  
   – ℜ: tập các quan hệ về video để lưu các DL có cấu trúc:  
   date, author,



1. **Truy vấn DL video**

**+ Các kiểu truy vấn:**– Tìm kiếm các đoạn video trong CSDL thỏa mãn ĐK nào đó  
– Tìm kiếm các đối tượng: cho video v & đoạn [s, e] (startframe,endframe) của v  
➔các đối tượng xuất hiện trong tất cả các khung hình hoặc 1 số khung hình nằm trong [s, e]  
– Tìm kiếm các hành động: cho 1 video + đoạn (s, e) của v  
➔các sự kiện xảy ra trong tất cả các khung hình hoặc 1 số khung hình nằm trong [s, e]  
– Truy vấn dựa trên thuộc tính: tìm kiếm các video hoặc các đoạn video trong đó các đối tượng/sự kiện thỏa mãn ĐK nào đó.

**Các Functions truy vấn video:**

|  |  |
| --- | --- |
| FindVideoWithObject(o) | {(v, s, e)}: đoạn trong video v có chứa object o ở tất cả các frame thuộc [s, e] |
| FindVideoWithActivity(a) | {(v, s, e)}: đoạn trong video v có chứa sự kiện a xảy ra |
| FindVideoWithActivityandProp (a, p, z) | {(v, s, e)}: đoạn trong video v có chứa sự kiện a với thuộc tính p = z xảy ra |
| FindVideoWithObjectandProp (o, p, z) | {(v, s, e)}: đoạn trong video v có chứa object o với z là giá trị của thuộc tính p |
| FindObjectsInVideo(v, s, e) | các đối tượng xuất hiện trong đoạn [s, e] ∈v |
| FindActivitiesAndPropInVideo (v, s, e) | Các sự kiện và giá trị thuộc tính tương ứng trong đoạn [s, e] ∈v |
| FindActivitiesInVideo (v, s, e) | các sự kiện xảy ra trong đoạn [s, e] ∈v |
| FindObjectsAndPropInVideo (v, s, e) | Các đối tượng và giá trị thuộc tính tương ứng trong đoạn [s, e] ∈v |

**Ngôn ngữ truy vấn**

**SQL mở rộng:**

SELECT Vid:[s,e]

FROM Video <source>

WHERE term IN func\_call

VD:

SELECT Vid:[s,e]

FROM Video: VidLib1

WHERE (Vid, s, e) IN FindVideoWithObject(Denis Dopeman)

VD: Tìm tất cả đoạn video trong CSDL trong đó Jane trao 1 chiếc vali cho Denis

SELECT Vid:[s,e]

FROM Video: VidLib1

WHERE (Vid, s, e) IN FindVideoWithObject(Denis Dopeman)

AND (Vid, s, e) IN FindVideoWithObject(Jane Shady)

AND (Vid, s, e) IN FindVideoWithActivityandProp(ExchangeObject, Giver, Jane Shady)

AND (Vid, s, e) IN FindVideoWithActivityandProp(ExchangeObject, Receiver, Denis Dopeman)

AND (Vid, s, e) IN FindVideoWithActivityandProp(ExchangeObject, Item, Briefcase)

1. **Đánh chỉ mục cho nội dung video**

**Mục tiêu:**

– Hỗ trợ hiệu quả việc thực hiện các kiểu truy vấn (8 hàm)

– Biểu diễn gọn (compact):

+ Không thể lưu nội dung của tất cả các frame:

1 video 90’ >= 90 x 60x25 frames

* **Kỹ thuật FS-tree (Frame-segment tree)**

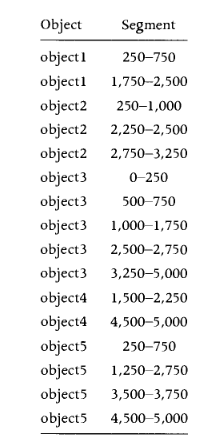
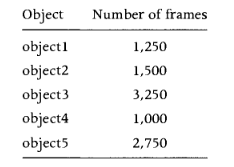
**Khái niệm:**

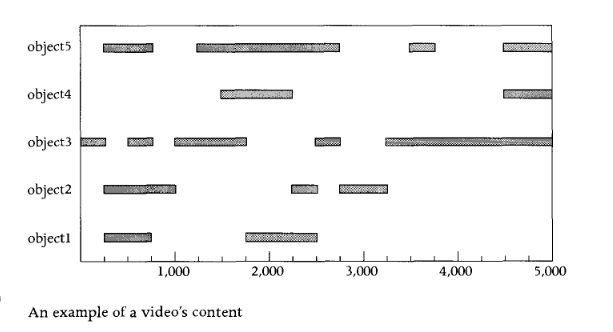
– Chuỗi khung hình: [i, j) = {k | i≤ k <j}  
– Thứ tự chuỗi khung hình: [i1, j1) ⊑ [i2, j2) ⇔ i1 < j1 ≤ i2 < j2  
fs1 = [8, 10), fs2 = [10, 15), fs3 = [11, 13)  
 fs1 ⊑ fs2, fs1 ⊑ fs3, fs2 ⋢ fs3  
– Tập có thứ tự chuỗi khung hình:  
+ X= {[i1, j1), [i2, j2), …, [ir, jr)}  
+ [i1, j1) ⊑ [i2, j2) ⊑ … ⊑ [ir, jr)  
+ VD: X= {[4, 9), [10, 11), …, [27, 41)}

– Tập chặt (*solid set*) chuỗi khung hình  
+ Là tập có thứ tự chuỗi khung hình  
+ Không tồn tại cặp chuỗi: [i1,i2) , [i2, i3)  
+ X = { [3, 5), [5, 7), [9, 11) }: not solid  
+ X = { [3, 5), [6, 7), [9, 11) }: solid  
– Biểu đồ kết hợp đoạn (*segment association map*) σv  
+ V = *(OBJ, ACT,* λ*)*+ σv(x) = A, x ∈ OBJ ∪ ACT, A: solid set  
 [s, e) ∈ A ➔ ∀f: s ≤ f < e, x ∈ λ*(f)* ∀f , ∀x ∈ OBJ ∪ ACT, nếu x ∈ λ*(f)* ➔ ∃ *[s, e)* ∈ A *: f* ∈[s, e)

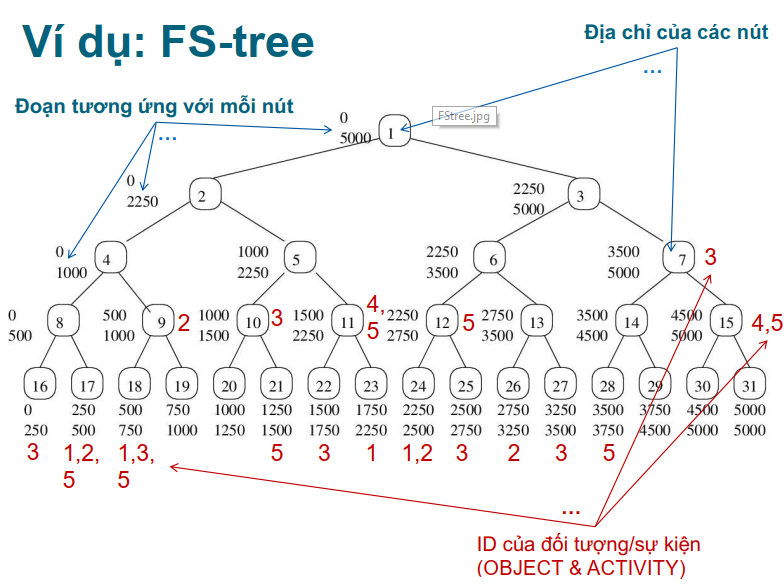
FS tree có các thành phần:  
– OBJECTARRAY (các đối tượng): danh sách con trỏ có thứ tự trỏ đến các nút trên FS-tree chứa frame có các đốitượng đó  
– ACTIVITYARRAY (các sự kiện): danh sách con trỏ có thứ tự trỏ đến các nút trên FS-tree chứa frame có các sự kiệnđó  
– FS-tree: cây nhị phân, được xây dựng dựa trên bảng phân đoạn

Ví dụ:

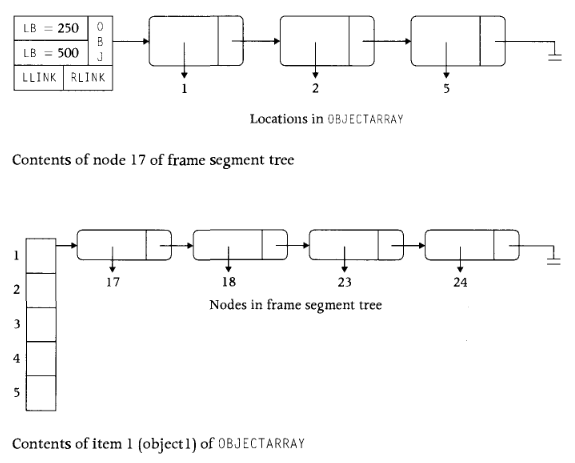




– FS-tree: cây nhị phân, được xây dựng dựa trên bảng phân đoạn (segment table):  
+ Nếu số các đoạn (z) trong bảng phân đoạn ≠ 2r:  
 thêm phần tử qz+1, …,q2^r: qz+j = qz + j,q2^r = framenum(v) + 1,  
 với r: số nhỏ nhất sao cho: 2r > z & 2r < framenum(v)  
+ Mỗi nút: biểu diễn chuỗi khung hình [x, y)  
+ Nút lá: mức r, theo thứ tự từ trái qua, các nút lá sẽ biểu diễn:  
[z1, z2), [z2, z3), [z3, z4), …  
+ Nút N có 2 con biểu diễn [p1, p2) & [p2, p3) thì N biểu diễn [p1, p3)



Cấu trúc 1 node trong FS-Tree



**CSDL Video (Video Library)**  
VidLib = {v1, v2, ..., vn}  
+ Mỗi vi, có 1 FS-tree: fst(vi)+ INTOBJECTARRAY = {(vi, oij, ptrij)}:  
 (v, o, ptr) ∈ INTOBJECTARRAY  
 iff (o, ptr) ∈ OBJECTARRAY của video v  
+ INTACTIVITYARRAY = {(vi, aij, ptrij)}:  
 (v, a, ptr) ∈ INTACTIVITYARRAY  
 iff (a, ptr) ∈ ACTIVITYARRAY của video v

**Tìm kiếm trên FS tree**

**+ FindVideoWithObject(o):**

SELECT Video\_id

FROM INTOBJECTARRAY

WHERE OBJ = o

**+ FindVideoWithActivity(a ):** Tương tự như trên

**+ FindVideoWithActivityAndProp(a, p, z):**

SELECT Video\_id

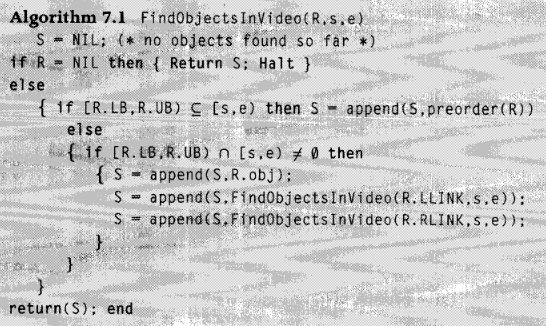
FROM INTACTIVITYARRAY t

WHERE ACT = a AND t.p = z

**+ FindVideoWithObjectandProp(o,p,z):** tương tự như trên

**+ FindObjectsInVideo( v , s ,e):**

Giải thuật

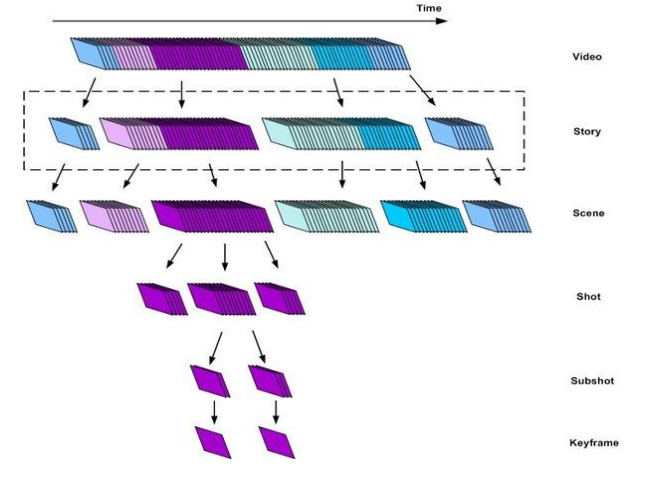


preorder (R ) : duyệt toàn bộ cây con R, lấy ra tất cả các object được gán ở mỗi nút

**Làm thế nào để lấy được nội dung của video ?  
Các đối tượng và các sự kiện**

Cấu trúc video

1 video v = fn(… (f2 (f1(S0, S1, t1), S2, t2) …, Sn, tn)– Si: shot: cảnh quay liên tục được tạo bởi 1 camera, đơnvị nhỏ nhất tạo nên nội dung video  
– fi (Si-1 ,Si , ti): phép toán kết hợp (composition operators)  
shot Si-1 và Si thành 1 đoạn video độ dài ti + Shot concatenation: cut (chuyển tức thì sang cảnh khác)  
 + Spatial composition: translate (dịch chuyển 1 shot trên 1 shot khác)  
 + Chromatic composition: fades in (hiện dần) / fades out (mất dần)/ dissolve (chuyển dần từ cảnh này sang cảnh khác)



**Một vài thuật ngữ**

- Scene:  
 + tập các shot liên tục có liên hệ ngữ nghĩa nhất định,thường là tập các cảnh quay ở cùng 1 địa điểm, thời gian  
- Keyframe:  
 + khung hình biểu diễn tốt nhất nội dung 1 cảnh quay (shot)  
 + thường sử dụng để đánh chỉ mục (indexing) hoặc hiển thị các video (browsing)  
- Video segmentation (phân đoạn video):+ Phân chia video thành các « đoạn đồng nhất » (homogeneous segment)  
 + Xác định Si, fi, ti, i = 1..n