Foreign keys.
joining Relations
Excel as Relational model.

Video1: Vector space model:-

→ vector model
→ concepts of similarity function and similarity search

→ Recognize that many document and image search enegine use vector models and similarity search.

* Document vector:

d1: "new york times"
d2: "new york post"
d3: "los angeles times"

tf

term frequency matrix

	angles	los	new	post	times	york
d1	0	0	1	0	1	1
d2	0	0	1	1	0	1
d3	1	1	0	0	1	0

penalty factor

idf

Inverse docunt freqct

$\log_2\left(\frac{3}{1}\right)$ $\log_2\left(\frac{3}{1}\right)$ $\log_2\left(\frac{3}{2}\right)$ $\log_2\left(\frac{3}{1}\right)$ $\log_2\left(\frac{3}{2}\right)$ $\log_2\left(\frac{3}{2}\right)$

$= 1.584$ $= 1.583$

tf-idf matrix

Term	angles	los	new	post	times	york	Length
d1	0	0	0.584	0	0.584	0.584	1.011
d2	0	0	0.584	1.584	0	0.584	1.784
d3	1.584	1.584	0	0	0.584	0	2.316

length of d1 $= \sqrt{0.584^2 + 0.584^2 + 0.584^2} = 1.011$

8.00 Searching in vector space :-

query ⟹ **q : new new york**

9.00 • Max frequency of a term ("new") = 2

10.00 • Create the query vector

Q [0 0 $(\frac{2}{2})$ * 0.584 = 0.584 0 $(\frac{1}{2})$ * 0.584 = 0.292]

1.00 length (q) = 0.652

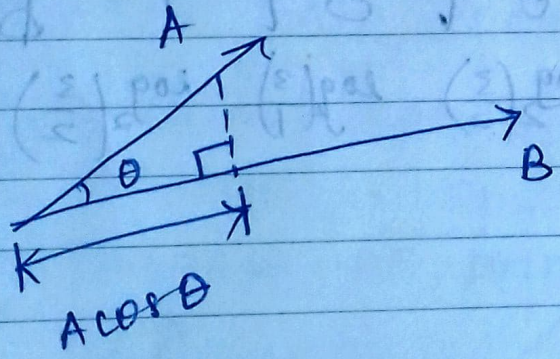
2.00 → A similarity fun^n bt^n two vectors is a measure of how far they are apart.

Similarity function :-

⟹ Many possible functions

⟹ Cosine distance :

$$\cos\theta = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$



Query Term Weighting
→ Every query term may optionally be associated with a weighting term.

Video2:

Graph Data model :-
Social Network

→ Identify graph data in practical problems.
→ Describe path operations, neighborhood operations, and connectivity operations in graph data.
→ paths
→ Optimal path operations
i) Find the shortest path btn two nodes.
ii) Find the optimal round-trip path that must include some specific nodes.
iii) Find "best compromise" paths btn two nodes.
→ Pareto - optimality

communities :- A subgraph of a graph that has many more edges within the subgraph compared to edges to node outside the subgraph.
• Operations :- Dense subgraph finding
- Optimization of clusteredness.

Anamalous Neigborhoods :- Near star, Near Clique
Heavy vicity, predominant Edge

Connectivity operations :-
Connectedness :- Every node is reachable from each node in the undirected version of the graph.

29
Wednesday

July 2015
Week 31
Day 210 • 155
Date 29 • 07 • 2015

Video 3 :-

Other Data Models :-

→ How arrays can serve as a data model

→ Explain why images can be modeled as vector arrays

→ Specify a set of operations on scalar and vector arrays.

* Array as a Data Model :

- Array → Indexed relation
- Table representation.

→ No. of columns = no. of dim + 1

→ No. of tuples = size of dim 1 x

Operations on Array of Vectors

- dim(A) - number of dimensions of A
- size(A, dim) - size of a specific dimension.
- A(i,j) - value of the element at the (i,j)-th cell
- A(i,j)[k] - value of the k-th element of the cell at A(i,j)
- length(A(i,j)) :- vector-length of the vector at the (i,j)-th cell
- distance(A(i,j), A(k,l), f) :- Vector distance bt^n the values of two cells given the dist. funct. f.