CMR
INSTITUTE OF
TECHNOLOGY

USN | | | | | | | | | | |

Internal Assessment Test - II

Sub:	File structures							Code:	10IS63
Date:	/ 0 / 2017	Duration:	90 mins	Max Marks:	50	Sem:	VI	Branch:	ISE

Answer Any FIVE FULL Questions

		Marks	OBE	
			CO	RBT
1	Explain K-Way merging algorithm in files with an example of k=8 ?	[10]	CO 4	L2
2	Discuss the methods used to avoid collision in hashing.	[10]	CO 6	L2
3	Discuss about Hashing Functions in detail with necessary examples.	[10]	CO 6	L1
4	What is an index? How to create a simple Index ?Illustrate with an example and the implementation class for INDEX?	[10]	CO 3	L1
5	Apply Data compression over the files for saving space? Explain.	[10]	CO 2	L1
6	Define multiple key ? List the Operations involved with secondary keys?	[10]	CO 3	L3
7	Define reclaiming space? explain how is achieved on fixed and variable length records.	[10]	CO 2	L1

CMR
INSTITUTE OF
TECHNOLOGY

Internal Assessment Test 2 – may 2017

SCHEME & SOLUTION

Sub:	File Structures		Code:	10IS63
		Sem: VI	Branch:	ISE

1. Explain K-Way merging algorithm in files with an example of k=8 ? [10]

k-way merge 2

A merge of order k.

order of a merge

The number of input lists being merged.

- If the distribution phase creates *k* runs, a single k-way merge can be used to produce the final sorted file.
- A significant amount of seeking is used by a k-way merge, assuming the input runs are on the same disk.

k-way merge

A merge of order k.

order of a merge

The number of input lists being merged.

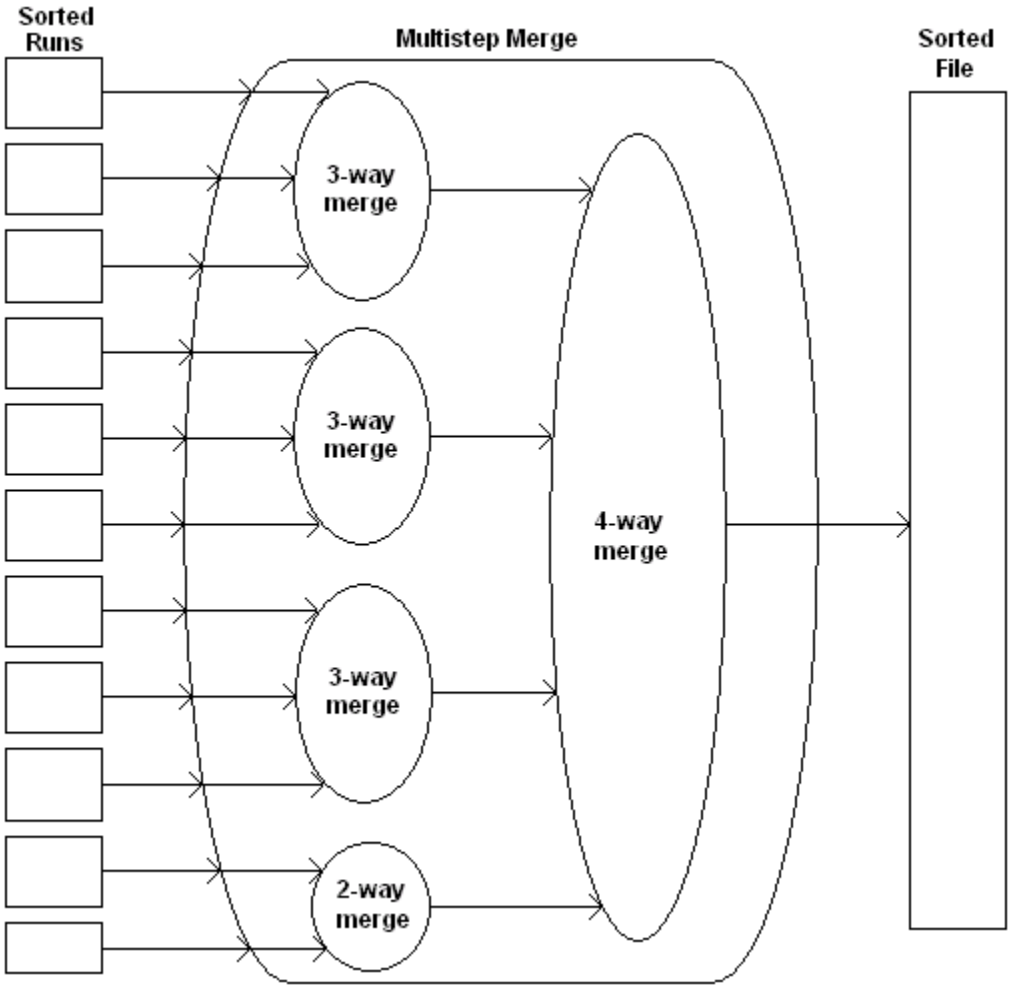
- If the distribution phase creates *k* runs, a single k-way merge can be used to produce the final sorted file.
- A significant amount of seeking is used by a k-way merge, assuming the input runs are on the same disk.

Multistep Merging

multistep merge [8]

A merge which is carried out in two or more stages, with the output of one stage being the input to the next stage.

- A multistep merge increases the number of times each record will be read and written.
- Using a multistep merge can decrease the number of seeks, and reduce the overall merge time.



Sorting Files on Tape

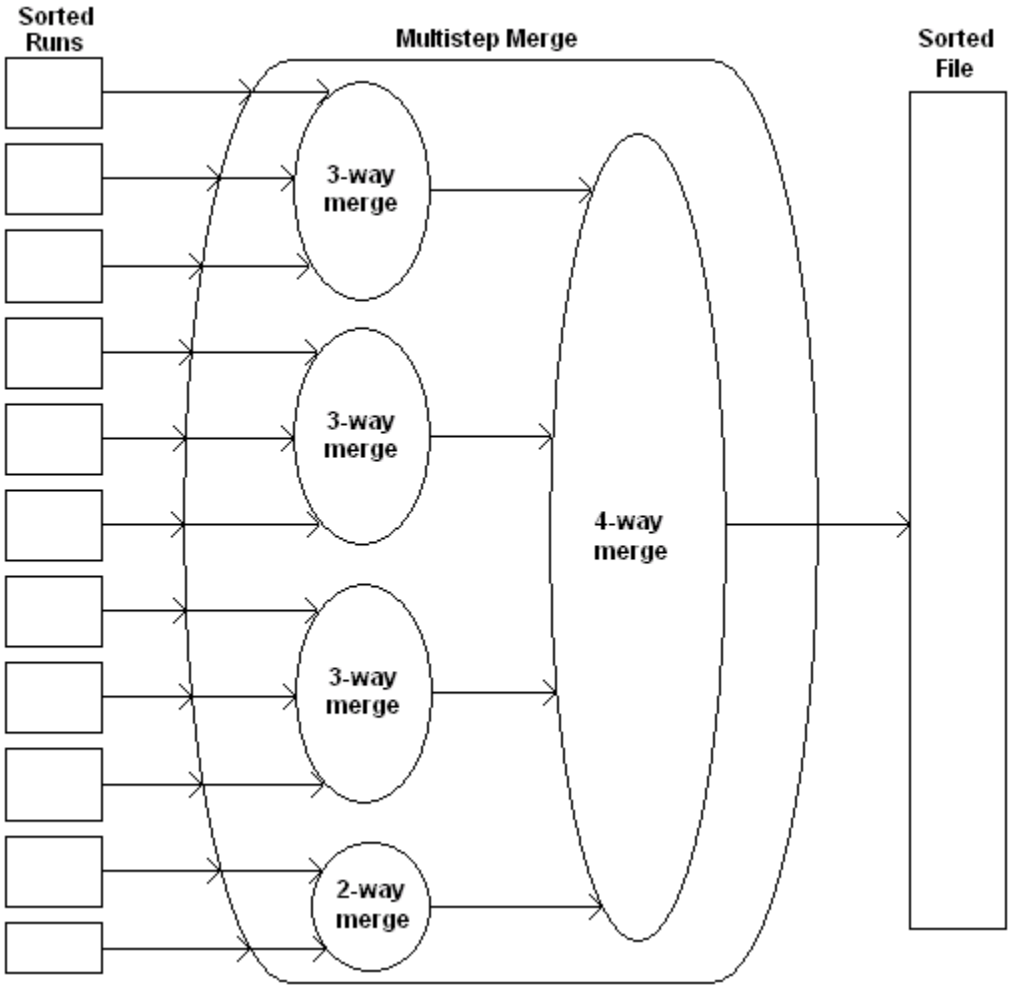
- When sorting with tape, multiple runs are placed in a single file.

balanced merge

A multistep merge which uses the same number of output files as input files.

A merge which is carried out in two or more stages, with the output of one stage being the input to the next stage.

- A multistep merge increases the number of times each record will be read and written.
- Using a multistep merge can decrease the number of seeks, and reduce the overall merge time.



Sorting Files on Tape

- When sorting with tape, multiple runs are placed in a single file.

balanced merge

A multistep merge which uses the same number of output files as input files.

2. Discuss the methods used to avoid collision in hashing. [10]

Collisions [2]

- synonyms
  Keys which hash to the same value.
- collision
  An attempt to store a record at an address which does not have sufficient room
- packing density
  The ratio of used space to allocated space.
- For simple hashing, the probability of a synonym is the same as the packing density.

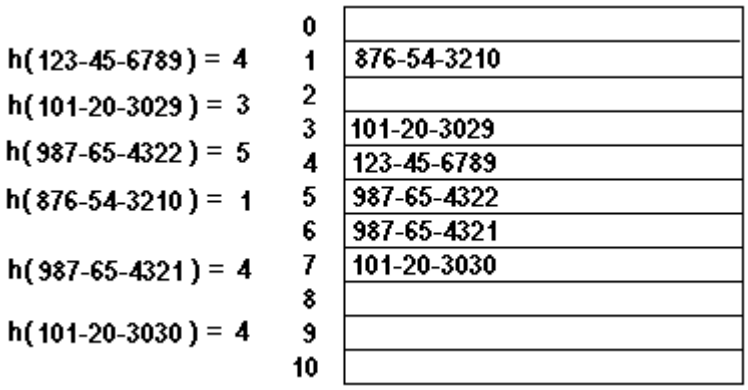
How Much Extra Memory Should be Used?

- Increasing memory (i.e., increasing the size of the hash table) will decrease collisions.

Collision Resolution by Progressive Overflow [8]

- progressive overflow

  A collision resolution technique which places overflow records at the first empty address after the home address
- With progressive overflow, a sequential search is performed beginning at the home address.
- The search is continued until the desired key or a blank record is found.
- Progressive overflow is also referred to as *linear probing*.



3. Discuss about Hashing Functions in detail with necessary examples. [10]

Modulus [5*2]

- Modulus - the key is divided by the size of the table, and the remainder is used as the hash function.
- Example:

  Key = 123-45-6789

  123456789 % 11 = 5

  h(123-45-6789) = 5
- Modulus functions work better when the divisor is a prime number, or at least not a composite of small numbers.

Fold and Add

- fold and add

  A hashing techique which separates a long key field into smaller parts which are added together.
- Example:

  Key = 123-45-6789

  123

  456

+789

1368

h(123-45-6789) = 1368

- Fold and add is used for long keys.

Mid-square

- mid-square

  hashing method which squares the key and the uses the middle digits of the result.

- Example:

  Key = 123-45-6789

  123456789^2 = 15241578750190521

  h(123-45-6789) = 8750

Combined methods

- Practical hashing functions often combine techniques.
- Example:
  Key = 123-45-6789
  123
  456
  +789
  1368

  1368 % 11 = 4
  h(123-45-6789) = 4
- For non-numeric keys, the key is simply treated as though it were a number, using its internal binary representation.
- Example:
  Key = "Kemp"
  "Kemp" = $4B656D70_{16} = 1264938352_{10}$

4. Discuss about Making deletions in records using hashing. [10]

Making Deletions [2]

- tombstone
  A marker placed on a deleted record.
- Using a tombstone avoids terminating a probe prematurely.
- Locations containing a tombstone can be be used for records being added.

EXAMPLE & EXPLANATION [8]

	0	
h(123-45-6789) = 4	1	876-54-3210
h(101-20-3029) = 3	2	
h(987-65-4322) = 5	3	101-20-3029
h(876-54-3210) = 1	4	123-45-6789
	5	**
h(987-65-4321) = 4	6	987-65-4321
h(101-20-3030) = 4	7	101-20-3030
	8	
	9	
	10	

5. Define Extendible hashing. Explain in detail with steps and example [10]

How Extendible Hashing Works [2]

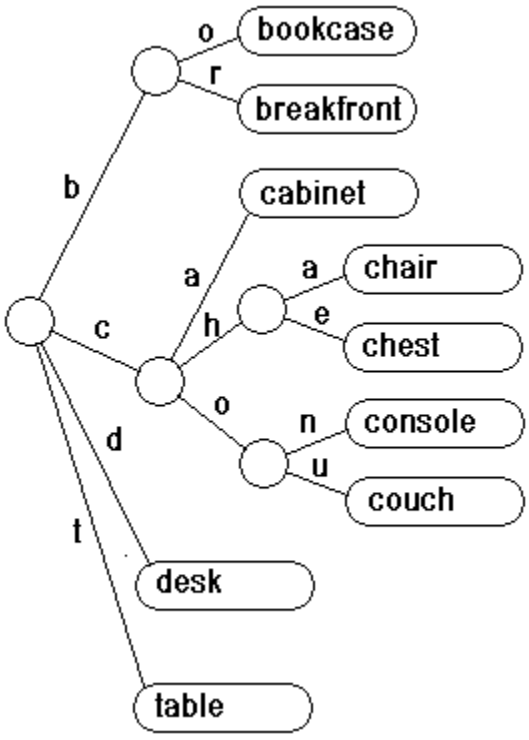
Tries

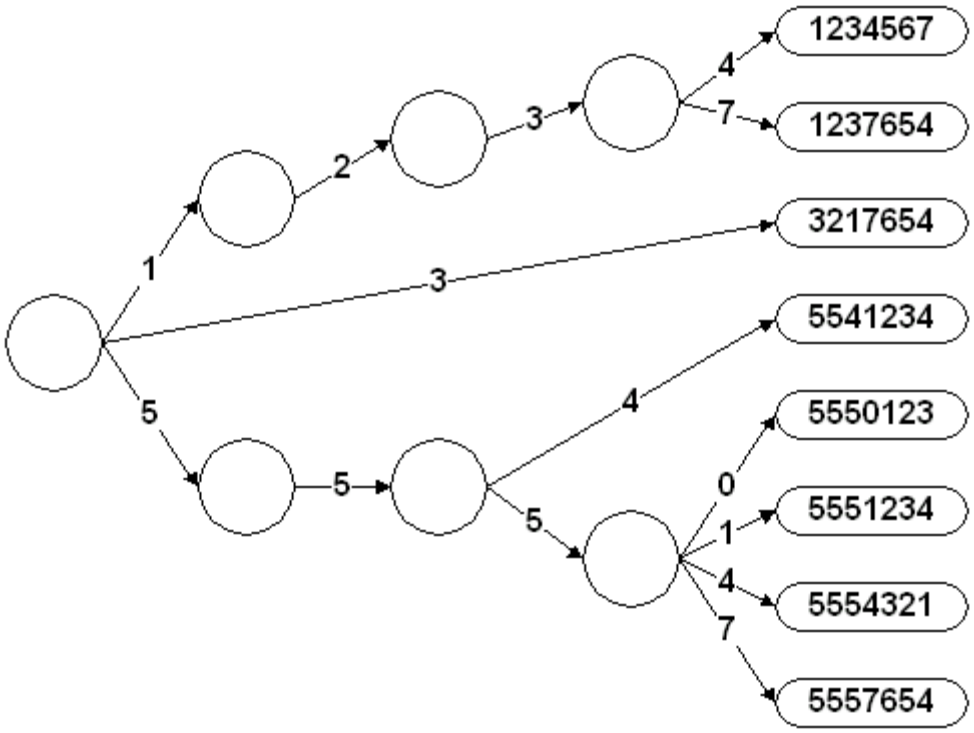
- trie

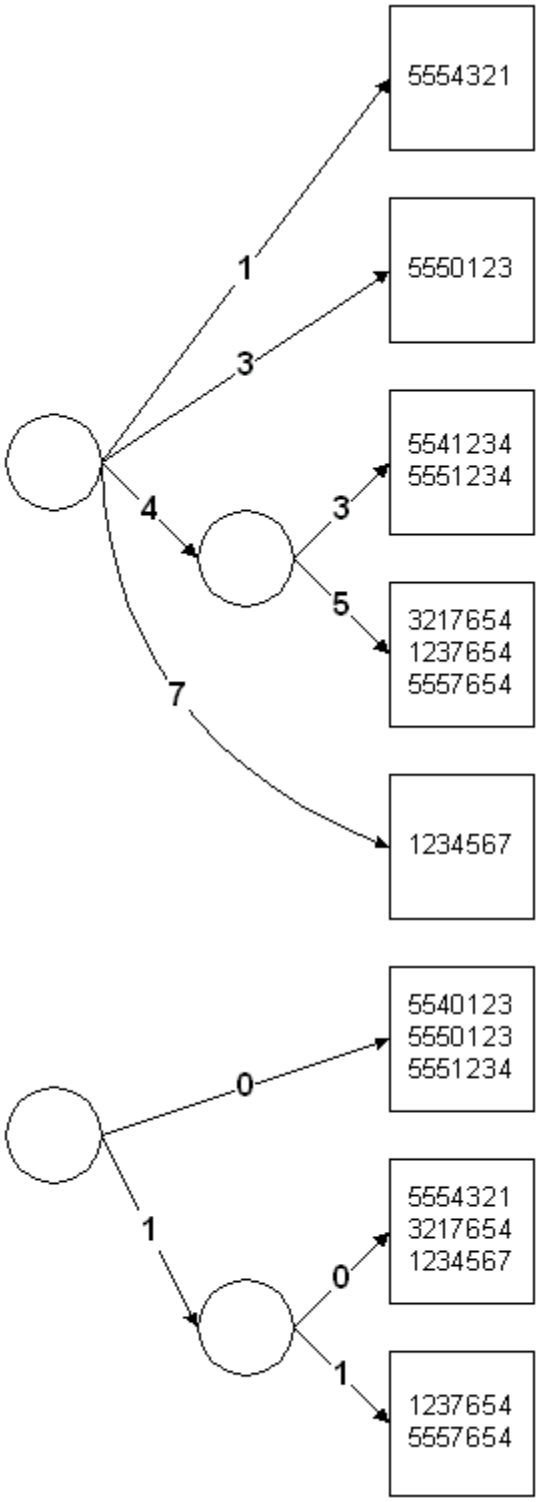
A search tree in which the child of each node is determined by subsequent charaters of the key.

- An alphabetic (radix 26) trie potentially has one child node for each letter of the alphabet.
- A decimal (radix 10) trie has up to 10 children for each note.
- The trie can be shortened by the use of buckets.
- The bucket distribution can be balanced by the use of hashing.

Key	Hash
5554321	100111001
5550123	10111010
5541234	100111100
5551234	1011110
3217654	100111101
1237654	10011011
5557654	101110011
1234567	1101001



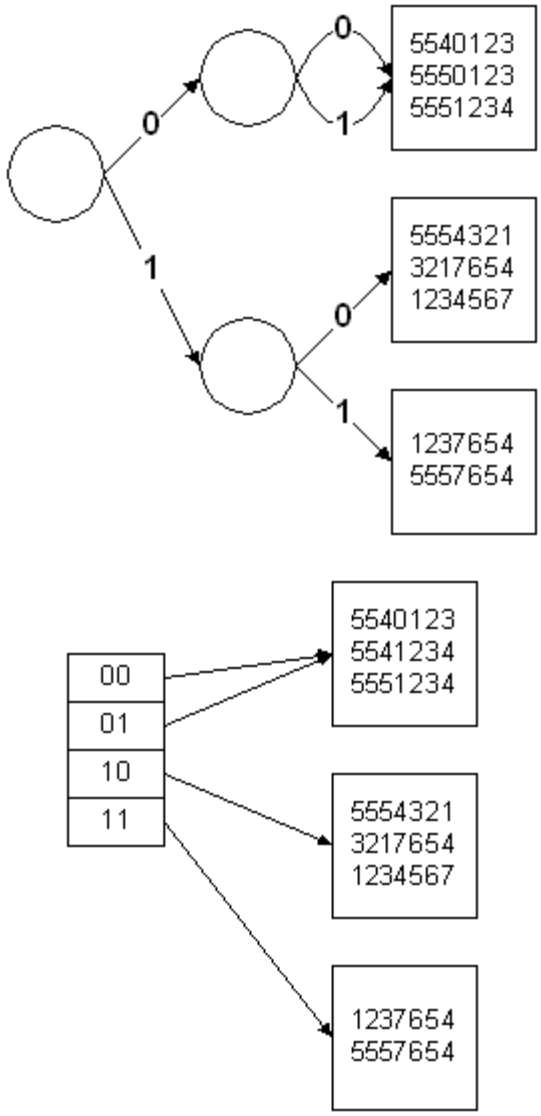

1
2
3
4
7
1234567
1237654
3
3217654
5
5
4
5541234
5
0
1
4
7
5550123
5551234
5554321
5557654



Turning the Tries into a Directory [4]

- extendible hashing

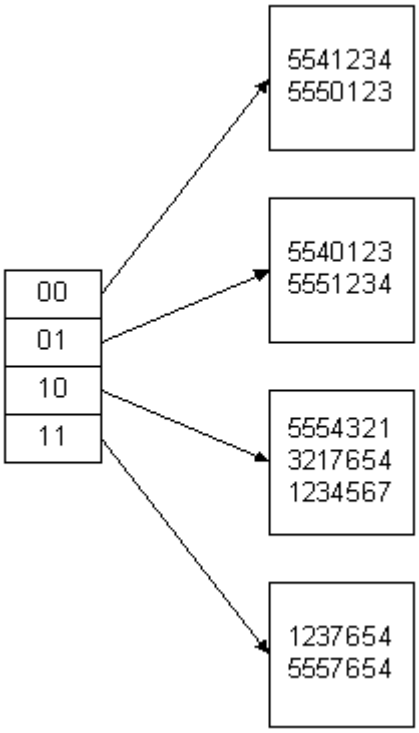
  An application of hashing that works well with files that over time undergo substantial changes in size.

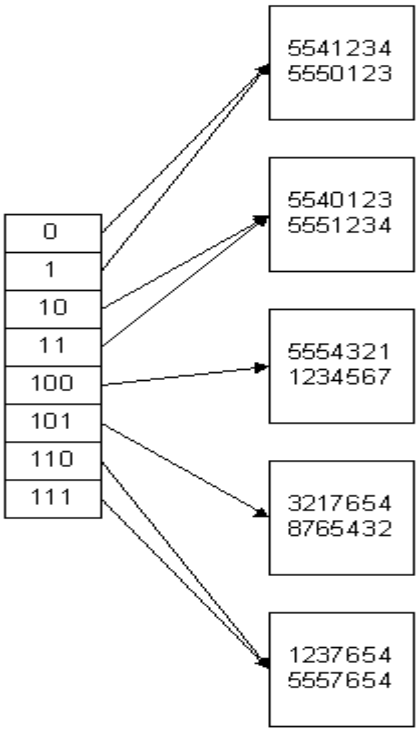


Splitting to Handle Overflow [4]

- splitting

  Creation of a new node when a node overflows, with the partial distribution of the contents of the overflowing node to the new node.





6. What is an index? How to create a simple Index ?Illustrate with an example and the

implementation class for INDEX? [10]

simple index [2]

An index in which the entries are a key ordered linear list.

- Simple indexing can be useful when the entire index can be held in memory.
- Changes (additions and deletions) require both the index and the data file to be changed.
- Updates affect the index if the key field is changed, or if the record is moved.
- An update which moves a record can be handled as a deletion followed by an addition.

What is an Index? [4]

index

A structure containing a set of entries, each consisting of a key field and a reference field, which is used to locate records in a data file.

key field [4]

The part of an index which contains keys.

reference field

The part of an index which contains information to locate records.

- An index imposes order on a file without rearranging the file.
- Indexing works by indirection.

7. What is multiple key ? List the Operations involved with secondary keys? [10]

secondary key [2]

A search key other than the primary key.

secondary index [3]

An index built on a secondary key.

- Secondary indexes can be built on any field of the data file, or on combinations of fields.
- Secondary indexes will typically have multiple locations for a single key.
- Changes to the data may now affect multiple indexes.

1. CREATE 2. Append 3.load 4.delete 5.search 6.modify

- The reference field of a secondary index can be a direct reference to the location of the entry in the data file.
- The reference field of a secondary index can also be an indirect reference to the location of the entry in the data file, through the primary key.
- Indirect secondary key references simplify updating of the file set.
- Indirect secondary key references increase access time.

Retrieval Using Combinations of Secondary Keys [5]

- The search for records by multiple keys can be done on multiple index, with the combination of index entries defining the records matching the key combination.
- If two keys are to be combined, a list of entries from each key index is retrieved.
- For an "or" combination of keys, the lists are merged.
- I.e., any entry found in either list matches the search.
- For an "and" combination of keys, the lists are matched.
- I.e., only entries found in both lists match the search.