

Advanced Database Technology
Anna Pagh and Rasmus Pagh
IT University of Copenhagen
Spring 2005

Index structures

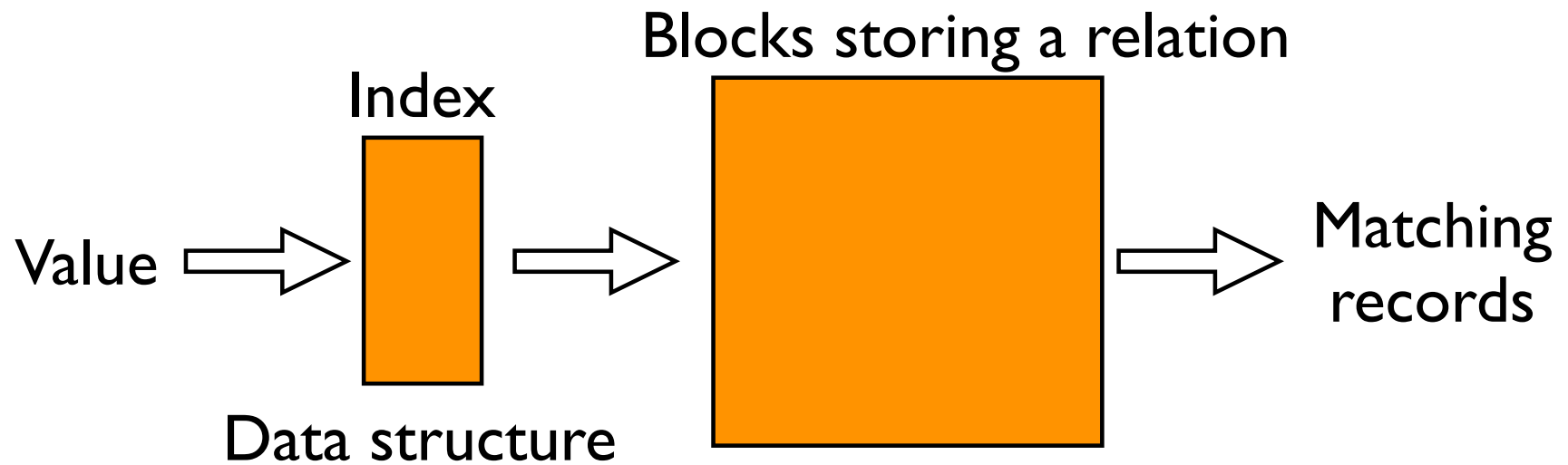
February 14, 2005

Based on Chapter 13.1-13.2 in G UW



IT University
of Copenhagen

Why indexes?



Goal of an index:

Look at as few blocks as possible to find the matching record(s)



Sequential files

- Store relation in sorted order according to search key.
- Binary search in logarithmic time (I/Os) in the number of blocks used by the relation.



Dense index

- For each record store the key and a pointer to the record in the sequential file.
- Why? Uses less space, hence less time to search. Time (I/Os) logarithmic in number of blocks used by the index.
- Can also be used as secondary index, i.e. with another order of records.



Sparse index

- Store first value in each block in the sequential file and a pointer to the block.
- Uses even less space than dense index, but the block has to be searched, even for unsuccessful searches.
- Time (I/Os) logarithmic in the number of blocks used by the index.



Multiple levels of indexes

- If an index is small enough it can be stored in internal memory. Only one I/O is used.
- If the index is too large, an index of the index can be used.
- Generalize, and you have a B-tree. The top level index has size equal to one block. (Topic of next lecture).



Updates

- In a dense index there are pointers to all records, hence the index is updated after every insertion/deletion.
- A sparse index is sometimes updated after insertions/deletions, e.g. when an index record is deleted.
- Overflow blocks and tombstones can be used both in the sequential file and in the index.



Primary and secondary indexes

- In a primary index records are stored in order decided by the index, e.g. sequentially.
- A relation can have at most one primary index. (Often on primary key.)
- A secondary index can not take advantage of any specific order, hence it has to be dense.
- Secondary index can have a second, sparse level.



Other indexes

- B-tree indexes (next week)
- Hash indexes (next week)
- Indexes for text data
- Indexes for geometric data
- and more... (not covered in the course)

