



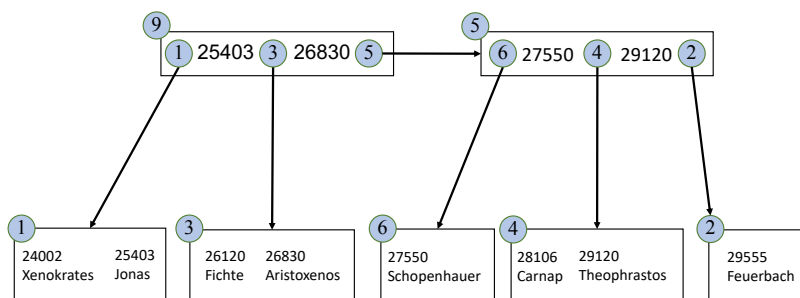
Exercise for *Database System Concepts for Non-Computer Scientist* im  
WiSe 19/20

Alexander van Renen (renen@in.tum.de)  
<http://db.in.tum.de/teaching/ws1920/DBSandere/?lang=en>

Sheet 11

Exercise 1

Consider the ISAM example from the lecture:



Determine how many pages have to be accessed for the following queries in the example ISAM.

- (a) `select * from Students;`
- (b) `select * from Students where studNr = 28106;`
- (c) `select * from Students where name = 'Carnap';`

**Solution:**

- (a) To access all tuples, we have to start at the “root” of the ISAM: Index page 9. From there we visit all children (page 1, 2) and then continue to the next index page (5). There we, again, visit all children (6, 4, 2) and are done, as there are no more index pages. In total we had to look at 7 pages.
- (b) To access a single tuple by its key (28106), we can make use of the fact that the ISAM is sorted. We, again, start at the “root” (page 9) and compare the key we are searching for (28106) to the separator keys in this index page (25403, 26830). The key is larger than all keys, therefore we have to move on to the next index page (5). By comparing with the separator keys in this index page, we determine that the key we are searching for (if it exists) must be in page 4 (because 28106 is between 27550 and 29120). We then check page 4 and find the “28106, Carnap” tuple.
- (c) To find a student tuple by name, we can do no better than using a full table scan. The ISAM only sorts the tuples by one attribute (studNr in this case).

## Exercise 2

Index structures are great to reduce the number of accessed pages. Let's assume a larger example with 1 million students that are all stored in an ISAM. The index pages of the ISAM can hold 2 separator keys (student numbers), as in the example. The data pages of the ISAM can hold 100 student tuples, unlike in the example where we only stored up to 2 student tuples on each data page. Calculate how many pages have to be accessed in the worst case for a scan query:

```
select * from Students;
```

and any primary key lookup / point query:

```
select * from Students where studNr = ???;
```

We first calculate the number of data/leaf pages/nodes in the ISAM and then the number of index pages. Using this information and the insights gained from **Exercise 1**, we can determine the result.

The number data pages can easily be calculated to be 10000 with 1 million students and 100 students fitting on one data page. For index pages we distinguish between the last index page (which can address 3 data pages) and regular index pages (which only address 2 data pages). There will be 1 "last" index page that takes care of 3 data pages and  $x$  regular index pages that need that need to handle  $10000 - 3$  data pages. With  $x = (10000 - 3) \div 2$  we get 4998.5. The number needs to be rounded up (because even if an index page is only half full, it still counts as a whole page). To get the total number of index pages we also need to add 1 to the number, as the last index page (which we treated separately) also counts as an index page. This leaves us at 5000 index pages and 10000 data pages.

**Full table scan:** As before we need to look at all pages:  $10000 + 5000 = 15000$

**Point query:** We are interested in the worst case, meaning the point lookup query that requires the most pages. This is the one where we iterate the index pages up until the very last one (5000), before we go to a data page (1). In total this makes 5001 accessed pages.