



Data Structures and Algorithms (10)

Instructor: Ming Zhang

Textbook Authors: Ming Zhang, Tengjiao Wang and Haiyan Zhao

Higher Education Press, 2008.6 (the "Eleventh Five-Year" national planning textbook)

<https://courses.edx.org/courses/PekingX/04830050x/2T2014/>



Chapter 10 Search

- 10.1 Search in a linear list
- 10.2 Search in a set
- 10.3 Search in a hash table
- Summary



Basis Concepts

- Search

The process of finding a record with its key value equal to a given value in a set of records, or the records whose keys meet some specific criteria.

- The efficiency of search is very important
 - Especially for **big data**
 - Need **special storage processing** for data



Methods of Improving Search Efficiency

- Sorting

- Indexing

- Hashing

- When hashing is not suitable for disk-oriented applications, we can use B trees.

- Take much time
- Organize the data into a table
- preprocessing (finished before search)
- Get the position of records in the table according to the key values
- disadvantages :
- Make the most of auxiliary index information
- Unsuitable for range searches
- Generally, duplicate keys are not allowed
- Sacrifice space
- To improve search efficiency



Average Search Length (ASL)

- Comparison of keys: main operation of search
- **Average Search Length**
 - Average number of comparisons during search
 - The time metric for evaluating search algorithms

$$ASL = \sum_{i=1}^n P_i C_i$$

■ P_i is probability of searching the i -th element

■ C_i is the number of comparisons needed to find the i -th element



Other Metrics for Evaluating Search Algorithms

- Considerations when evaluating search algorithms
 - The storage needed
 - Implementation difficulties
 - ...



Thinking

- Assume that a linear list is (a, b, c), and the probabilities of searching a, b, c are 0.4, 0.1, 0.5 respectively
 - What is the ASL of sequential search algorithms? (which means how many times of comparisons of key values are needed to find the specific element on the average)



Data Structures and Algorithms

Thanks

the National Elaborate Course (Only available for IPs in China)

<http://www.jpk.pku.edu.cn/pkujpk/course/sjjg/>

Ming Zhang, Tengjiao Wang and Haiyan Zhao

Higher Education Press, 2008.6 (awarded as the "Eleventh Five-Year" national planning textbook)