

Better Programming

Chuncheng Zhang

March 29, 2021

Abstract

Every one has *computer*. It computes at a very high speed. Basically, you put your data in *objects*, and generate *functions* to do the computation. The *languages* are the communication between you and the computer. And the *algorithms* are the tools of how to do the computation precisely and quickly. A good *programming* is about all of the above, you have to manage them, like the arm controls the hand.

Contents

1	Computer	2
1.1	Input & Output	2
1.2	Memory	2
1.3	Disk	3
2	Objects	3
2.1	Data Organization	4
2.1.1	Chain List	5
2.1.2	Hash Table	6
2.1.3	Conclusion	8
2.2	Data Structure	9
2.2.1	Stack	9
2.2.2	Heap	10
3	Functions	10
4	Languages	10
5	Algorithms	10
A	Computation Complexity	10

1 Computer

The computer is PC or laptop in real-world. But in programming, the computer can be a very *abstraction* concept. Four components are necessary:

- Input
Used for user input
- Output
Used for output to screen or speaker
- Memory
Where the variables stay during computing
- Disk
Where the files stay forever

1.1 Input & Output

The input and output are the interfaces between the computer and users. Input means message from user to computer, and output means the reverse. They are applied in different syntax in different language. Take *Python* and *JavaScript* for examples.

Listing 1: InputOutput.py

```
1 '''
2 File: InputOutput.py
3 Aim: Example of input and output in Python
4 '''
5
6 # Input
7 inp = input('Input:')
8
9 # Output
10 print(f'You just inputted: {inp}')
```

Listing 2: InputOutput.js

```
1 /*
2 File: InputOutput.js
3 Aim: Example of input and output in JavaScript
4 */
5
6 // Input
7 const inp = prompt("Input:");
8
9 // Output
10 console.log("You just inputted:", inp);
```

1.2 Memory

When you practice programming, all the variables, functions and objects in your code is in the memory. In another word, the computation equals to the operation to the memory. More about memory can be found in the section of Objects. In current stage, all you need to know is everything you program is in the memory.

1.3 Disk

When computing is finished, users may have their stuff to be stored forever. The disk is where to put them.

It should be noticed that it can be very different between the things in memory and their storage in disk. For example, an article is structured as a characters array in the memory. However, it is stored as a highly compressed binary series in the disk. Although the difference between the two formats, they are the same article in fact.

At the viewpoint of the memory, there are fine programs to save the data to and read the data from the disk. In ideal condition, the two-way process is *transparent* to the user, which it frees the users to think about the conversion, thus the users can focus on the object in memory during computation.

2 Objects

The object is an overall calling to the things of interest. It can be a number, character, string, list or set. Moreover, it can even be a collection of them.

When you are thinking about a object, you are actually summarizing its features in the mind. But the computer works in the real world. That is a large separation.

In the abstraction level, the object is the summary of features.

And in the concrete level, the object is an instance of features.

The gap causes several problems.

Basically, a number has two features, the value and the precision. In abstraction level, $100/3$ is an existing number. However, in the real world, the computer can not represent it with infinite precision. As a result, every time you put your hand to it in the program, it shows as the given precision, and the output value can be different.

Listing 3: Precision.js

```
1  /*
2  File: Precision.js
3  Aim: Example of a number
4  */
5
6  // The number of 100 / 3
7  const a = 100 / 3;
8
9  // The int precision
10 // 33
11 console.log(parseInt(a));
12
13 // The float precision
14 // 33.333333333333336
15 console.log(parseFloat(a));
```

Fortunately, the *float* precision is far beyond to meet the standard of daily usage. Evenly, in modern programming language, like JavaScript and Python, the precision can be intelligent assigned to the variables

without causing problems in most cases, which is largely convenience to the users.

The effect of precision is only a small piece of the difference between the object in your mind and its instance in your computer. Keep it in your mind when practicing the programming, and it will save you from lots of unnecessary troubles.

2.1 Data Organization

Before you do anything with your computer, you need to understand the organization of data, since it may be too large to look at and place them one-by-one. Moreover, the data can be arbitrary on the using end. A good program is designed to deal with different values and sizes in the given data structure. As a result, the first thing of programming is in two folds:

- Decide how to organize the data;
- Estimate how long the operation will take at typical situation.

Does it look like two things? Yes, but actually No. I believe they are from one thing with the name of *Data Structure*. Before actually enter into the field of data structure, the data organization is easier to understand.

Think of a simple program to score the kids with their raw scores.

Listing 4: Score.js

```
1  /*
2  File: Score.js
3  Aim: Example of the scoring object
4  */
5
6  // We think the obj is the collection of scores.
7  // It records the raw scores and knows how to convert them
   into discrete scores.
8
9  // The raw scores
10 let obj = [79, 54, 80, 90];
11
12 // The scoring thresholds
13 obj.thresholds = {
14     A: 90,
15     B: 80,
16     C: 70,
17     D: 60,
18     E: 0,
19 };
20
21 // Scoring the raw scores
22 obj.score = (e, i, t) => {
23     for (let s in t.thresholds) {
24         if (e >= t.thresholds[s]) {
25             return [i, s];
26         }
27     }
28 };
29
```

```

30 // See what we got
31 // [ [ 0, 'C' ], [ 1, 'E' ], [ 2, 'B' ], [ 3, 'A' ] ]
32 console.log(obj.map(obj.score));

```

The script can be used to generate discrete scores based on the raw scores. As it writes, one can tell the discrete score based on the raw score value using the threshold table. It is quite short and simple. But it contains two classic data structures, the coder can solve more than 80% computation problems with them. They are *chain list* and *hash table*.

2.1.1 Chain List

A chain list contains countless nodes, the length can be very large. They are linked one after another like the nodes in a chain. It is almost the simplest data structure in the computer. The advantage is memory saving and easy to access to certain position, and the pitfall is it can be slow for query the certain value.

Listing 5: Array.js

```

1  /*
2  File: Array.js
3  Aim: Example of array.
4      Explain why it is good or bad.
5  */
6
7  // Generate array
8  // Assume it is very long, like 1,000,000 elements.
9  let arr = [4, 5, 70, 29, ..., ];
10
11 // Example of Good
12 // Access to the 7,364th element,
13 // it only requires ONE operation.
14 console.log("The 7,364th element is", arr[7363]);
15
16 // Example of Bad
17 // Tell if 365 is in the array, and where it first appears,
18 // it may require 1,000,000 operations to find an answer.
19 let found = false;
20 for (let i = 0; i < arr.length; i++) {
21     if (arr[i] === 365) {
22         console.log("Found 365 in the array at", i)
23         found = true;
24         return
25     }
26 }
27 if (!found) {
28     console.log("The 365 is not in the array")
29 }

```

The code describes the thing. It can be time consuming to find out if certain value exists. Commonly, it will find it after too many operations or report it can not be found in the list. At the worst condition, the computer has to check every node in the chain list to find the given value. As a result, the expectation of the time consuming is

$$\mathcal{O}(n) \tag{1}$$

where n refers the length of the chain list. It is a common notion of time consuming estimation in the field of *computational complexity*.

Computational complexity! As we know, Computation is the process of calculating something by mathematical or logical methods. Execution of an algorithm is a well-known example of computation. In order to perform any sort of computation, we will need energy. This energy can be directly related to the resources that we spend to perform computation.

Data from: Wikipedia

It refers the time consuming can be estimated by n

$$\mathcal{E}(t) \propto f(n) + c$$

where $f(\cdot)$ refers a linear function, and c is constant independent with n . As a result, when the n is too large, the program may use up countless computation times.

No one wants to wait almost forever to just find out some value. The question is how to speed up the process? The hash table is one solution.

2.1.2 Hash Table

The thresholds in Listing 4 is a hash table. As it writes, the elements in the hash table has two factors:

- Key: The unique key to access the record;
- Value: The value belonging to the key.

The hash table is used for quickly access to the value by certain key.

The idea of hash table is to use its hash value to identify the key and thus find its value. Since the key's location is computed rather than searched by exhaustively search ¹. The process can be expressed as

$$\begin{aligned} h &= \mathbb{H}(key) \\ i &= \mathbb{L}_n(h) \end{aligned}$$

where \mathbb{H} is hash function to compute the hash value ², and \mathbb{L}_n is projection function to compute its location i in the n sized hash table. A qualified hash system of $(\mathbb{H}, \mathbb{L}_n)$ can map the keys into the hash table evenly. That says the keys will be located to n locations with equal probability.

Using the hash method, the key-value pairs are separated evenly into the n length hash table. When query a key, the program only has to compute its hash value and find it from the location. Assume there m pairs in the hash table, the computation complexity is

$$\mathcal{O}\left(\frac{m}{n}\right) \quad (2)$$

The computation is because the exhaustively search process only has to deal with the list with length of $\frac{m}{n}$ in ideal condition ³. As a result, the hash table is commonly used for storage large data for arbitrary query.

Below is an example for explanation.

¹Exhaustively search: Including or considering all elements or aspects; fully comprehensive. Like search one-by-one in a chain list.

²Hash value is usually a string.

³Ideal refers the hash function is qualified.

Listing 6: HashTable.js

```

1  /*
2  File: HashTable.js
3  Aim: Example of hash table.
4  */
5
6  // The size of the table
7  const size = 7;
8
9  // The hash seed parameter
10 const I64BIT_TABLE = "0123456789abcdef".split("");
11
12 // Compute hash value
13 function hash(input) {
14     var hash = 5381;
15     var i = input.length - 1;
16
17     if (typeof input == "string") {
18         for (; i > -1; i--) hash += (hash << 5) + input.
            charCodeAt(i);
19     } else {
20         for (; i > -1; i--) hash += (hash << 5) + input[i];
21     }
22     var value = hash & 0x7fffffff;
23
24     var retValue = "";
25     do {
26         retValue += I64BIT_TABLE[value & 0xf];
27     } while ((value >>= 6));
28
29     return retValue;
30 }
31
32 // Convert hash value to idx of hash table
33 function hash2idx(h) {
34     const hex = "0x" + h;
35     return parseInt(hex) % size;
36 }
37
38 // Init empty hash table
39 let table = [];
40 for (let i = 0; i < size; i++) {
41     table[i] = undefined;
42 }
43
44 // The pairs to be restored in the hash table
45 const pairs = {
46     AliceBlue: "#F0F8FF",
47     AntiqueWhite: "#FAEBD7",
48     Aqua: "#00FFFF",
49     Aquamarine: "#7FFFD4",
50     Azure: "#F0FFFF",
51     Beige: "#F5F5DC",
52     Bisque: "#FFE4C4",

```

```

53     Black: "#000000",
54     BlanchedAlmond: "#FFEBCD",
55 };
56
57 // Insert the pairs in to the hash table
58 console.log("Inserting hash table");
59 for (const name in pairs) {
60     const h = hash(name);
61     const i = hash2idx(h);
62     console.log("\t", name, "->", h, "->", i);
63
64     if (table[i] === undefined) {
65         table[i] = [[name, pairs[name]]];
66     } else {
67         table[i].push([name, pairs[name]]);
68     }
69 }
70
71 // The generated hash table
72 console.log("Hash table is");
73 console.log(table);
74
75 /*
76 The output reads as following:
77 Inserting hash table
78     AliceBlue -> bc4bc1 -> 1
79     AntiqueWhite -> db6231 -> 4
80     Aqua -> d345c1 -> 2
81     Aquamarine -> 9c1c31 -> 4
82     Azure -> cf28f -> 1
83     Beige -> 18f7f -> 1
84     Bisque -> e42fb1 -> 2
85     Black -> 2931f -> 0
86     BlanchedAlmond -> 1b398 -> 2
87 Hash table is
88 [ [ [ 'Black', '#000000' ] ],
89   [ [ 'AliceBlue', '#F0F8FF' ],
90     [ 'Azure', '#F0FFFF' ],
91     [ 'Beige', '#F5F5DC' ] ],
92   [ [ 'Aqua', '#00FFFF' ],
93     [ 'Bisque', '#FFE4C4' ],
94     [ 'BlanchedAlmond', '#FFEBCD' ] ],
95   undefined,
96   [ [ 'AntiqueWhite', '#FAEBD7' ], [ 'Aquamarine', '#7FFFD4' ]
97     ],
98   undefined,
99   undefined ]
100 */

```

2.1.3 Conclusion

The chain list or hash table are two common structures. The chain list is simple and memory save. It is commonly used for the circumstances that users are only concern to enumerate the values but not query certain value.

The hash table is more complex with the hash function. It is more suitable for the circumstances that it is necessary for query arbitrary keys. The speed up for the query is several times depending on the size of hash table. It shows the chosen of data structure may largely affect the performance of the program. As a result, the data structure research is to develop better structure for certain circumstances.

2.2 Data Structure

There are huge number of customized data structure. They can be classified into two main classes

- Stack: One dimensional ordered list;
- Heap: Other than stack, mostly likes a tree.

Since the data structure is goal driven, the combine of several structures is usually required. For example, the hash table in Listing 6 is a combination usage of stacks in two level. The outer table is a stack, and its inner elements are also stacks. Meanwhile, the two level stack structure is also an untypical heap. So, we will review the stack and heap respectively in the following, in practice, there is not clear separation between them.

2.2.1 Stack

Typically, the stack is a chain of elements with one dimension.

Definition We think the elements in the stack as the nodes

$$Node \rightarrow \{e|(e.value, *e.next)\} \quad (3)$$

where $e.value$ is the value of interest, $*e.next$ is the point pointing to the next node. As a result, one can always iterate over all the elements from the first to the second and so on until it reaches to the end.

Property The stack is a tube with the width only fit for one ball. The tube has only one open end at its head, the other end is sealed. For moving in, the ball is inserted and go forwardly until it reaches the latest. For moving out, the latest ball move out firstly, and then the next and so on. Thus, the stack is an ordered storage with first-in-last-out (FILO) manner.

Definition When first-in-first-out manner is required, we need to extend the elements in the stack.

$$Node \rightarrow \{e|(e.value, *e.before, *e.after)\} \quad (4)$$

where $*e.before$ / $*e.after$ is the point pointing to the element before / after to it.

Property Using the two-way pointing extension, we open the both ends of the tube. Then the two ends are separated, one for entrance, and the other for exit only. As a result, for moving out, we can start from the first ball, and to its next, and so on. It is the FIFO staking order.

Sorting If the order of the elements being inserted is not of interesting, we always want the stack to be well-ordered. In another word, we need the elements to be ascending or descending monotone. Huge amount sorting algorithms have been developed

The methods can use to sort the elements. Although the time costing may change in practice, the expectation can be estimated. The table shows the faster method costs $n \cdot \log(n)$ operations and the slower method

Table 1: Time Complexities of Sorting Algorithms

Algorithm	Average Running Time ¹
Insert Sort	$\mathcal{O}(n^2)$
Bubble Sort	$\mathcal{O}(n^2)$
Selection Sort	$\mathcal{O}(n^2)$
Quick Sort	$\mathcal{O}(n \cdot \log(n))$
Merge Sort	$\mathcal{O}(n \cdot \log(n))$
Heap Sort	$\mathcal{O}(n \cdot \log(n))$

1. The time is the expectation value, the value is only for estimating the performance under "normal" conditions

costs n^2 operations. And the difference between methods can be as large as $n/\log(n)$ times. If you look closer, you will find out the fact that The different sorting algorithms reach their best performance under different situations. So, choosing the qualified algorithm for your application is necessary for programming.

Another interesting thing in the table is the heap sorting. It sorts the elements by building a heap, another common data structure.

2.2.2 Heap

To be continued. **Definition**

Property

Example

3 Functions

4 Languages

5 Algorithms

A Computation Complexity