Interpolation search

การค้นหาโดยการประมาณช่วง

Interpolation search

เป็นการหาข้อมูลแบบเดาสุ่ม ที่ไม่ใช่การเดาสุ่มแบบไม่มีหลักการ ไม่ได้เริ่มจากตรงกลางของข้อมูล แต่เป็นการสุ่มตำแหน่งของข้อมูลที่ต้องการโดย ชุดข้อมูลที่ทำการค้นหาต้องมีการจัดเรียงไว้แล้วอย่างเรียบร้อยถูกต้อง ต้องรู้ค่ามากที่สุดและน้อยที่สุดของชุดข้อมูล รู้ขอบเขตของข้อมูล ซึ่งตำแหน่งที่เริ่มทำการค้นหาจะได้จากการคำนวณตามสูตร

ขั้นตอนวิธี

กำหนด

Find = ค่าที่ต้องการค้นหา

Low = 0

High = ขนาดของอาเรย์ที่ต้องการค้นหา − 1

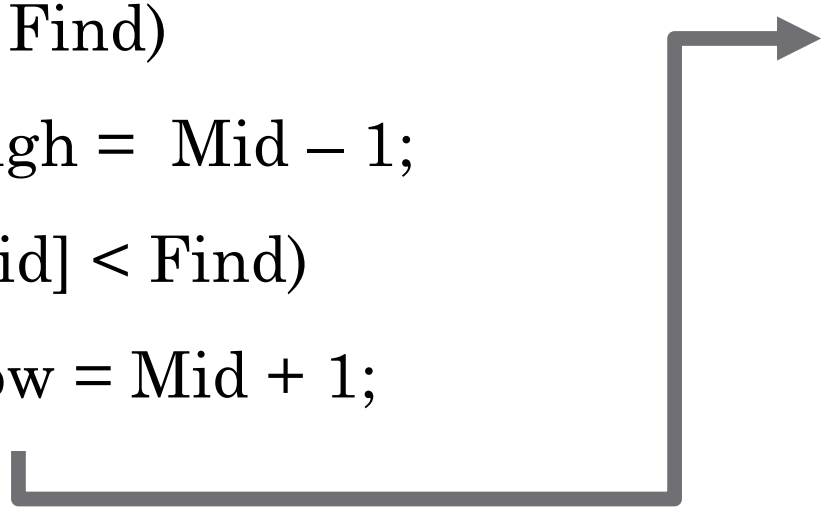
Mid = เป็นการคำนวณโดย

$$\text{Mid} = low + \frac{(Find - Array[Low]) * (High - low)}{Array[High] - Array[Low]}$$

ขั้นตอนวิธี

เช็คเงื่อนไข

```
While(Array[low] <= Find &&
Array[High] >= Find)
{
If(Array[Mid] > Find)
        return High =  Mid – 1;
Else if(Array[Mid] < Find)
        return Low = Mid + 1;
Else
        return Mid;
}
If(Find == array[low])
        return Low;
Else
        return -1;
```



โค้ดการทำงานเบื้องต้น

```
int Interpolation(int a[], int Size, int Find)
{
    int low = 0;
    int high = Size;
    int mid;

    //cout<<"high : "<<high<<endl;
    //mid = low +
    //                    (Find - a[low]) * (high - low) /
    //                   (a[high] - a[low]);
    //cout<<"mid : " <<mid<<endl;
    while(a[low] <= Find && a[high] >= Find)
    {
        mid = low + ((Find - a[low]) * (high - low)/
                                       (a[high] - a[low]));
        if(a[mid] > Find)
            high = mid - 1;
        else if(a[mid] < Find)
            low = mid;
        else
            return mid;
    }

    if(a[low] == Find)
        return low+1;
    else
        return -1;
}
```

ประสิทธิภาพการทำงาน

- การทำงานของการค้นหาโดยการประมาณช่วง จะมีประสิทธิภาพการทำงานเป็น $O(\log\log n)$
- ต่างจากประสิทธิภาพการทำงานของ binary search tree ที่มีประสิทธิภาพการทำงานเป็น $O(\log n)$
- เคสที่แย่ที่สุดคือ ข้อมูลที่ต้องการหามีการกระจายของข้อมูลไม่สม่ำเสมอ เช่น $(1,5,8,12,19,..)$ จะทำให้การค้นหาแบบ interpolation search กลายเป็นการค้นแบบ linear search ประสิทธิภาพการทำงานเป็น $O(n)$
- เคสที่ดีที่สุด คือเคสที่ข้อมูลมีการกระจายของข้อมูลสม่ำเสมอเช่น $(2,4,6,9,10,\ldots)$

การนำไปประยุกต์ใช้งาน

- การค้นหารายชื่อในโทรศัพท์มือถือ
- การค้นหาคำในพจนานุกรม
- การค้นหารายชื่อหนังสือในร้าน