

خوارزمية البحث الثنائي

في علم الحاسوب، خوارزمية البحث الثنائي (Binary search algorithm) : والمعروفة أيضاً باسم البحث المحدد بفاصل المنتصف أو البحث اللوغاريتمي أو القطع الثنائي أو البحث بالتصنيف هي خوارزمية بحث تجد موضع القيمة المستهدفة داخل مصفوفة مرتبة. يقارن البحث الثنائي القيمة المستهدفة بالعنصر المتوسط من المصفوفة. إذا لم تكن متساوية، يتم التخلص من النصف الذي لا يمكن للهدف أن يكون فيه ويستمر البحث في النصف المتبقي؛ تتكرر العملية مرة أخرى مع أخذ العنصر المتوسط للمقارنة بالقيمة المستهدفة، حتى العثور على القيمة المستهدفة. إذا كانت نتيجة البحث أن النصف المتبقي فارغ من العناصر، فهذا يعني أن القيمة المُستهدفة غير موجودة في المصفوفة.



في أسوأ حالة يعمل البحث الثنائي في وقت لوغاريتمي، مُنجزاً $O(\log n)$ مقارنة، حيث n هو عدد العناصر في المصفوفة، و O هو تمثيل O الكبرى، و \log هو اللوغاريتم.

البحث الثنائي أسرع من البحث الخطي من خلال المصفوفات الصغيرة. مع ذلك، يجب ترتيب المصفوفة أولاً حتى تتمكن من تطبيق البحث الثنائي. هناك هياكل بيانات متخصصة مصممة للبحث السريع، مثل جداول التجزئة، والتي يمكن البحث عنها بكفاءة أكبر من البحث الثنائي. ومع ذلك، يمكن استخدام البحث الثنائي لحل مجموعة أكبر من المشاكل، مثل العثور على العنصر التالي الأصغر أو التالي الأكبر في المصفوفة بالنسبة للهدف حتى إذا لم يكن موجوداً في المصفوفة.

ثمة تطبيقات متنوعة للبحث الثنائي؛ على وجه الخصوص، يسرّع التتالي الجزئي عمليات البحث الثنائية عن قيمة واحدة في مصفوفات متعددة. وهو قادر على حل عدد من مشكلات البحث في الهندسة الحسابية وفي العديد من المجالات الأخرى بكفاءة عالية.

يوسع البحث الأساسي البحث الثنائي ليشمل القوائم غير المحدودة، وتعتمد بنى بيانات شجرة البحث الثنائية والشجرة الثنائية على البحث الثنائي أيضاً.

خوارزمية العمل:

الإجراءات:

يعمل البحث الثنائي على المصفوفات المرتبة بهدف إيجاد موقع قيمة ما تُسمى القيمة المُستهدفة أو القيمة الهدف. تبدأ العملية بمقارنة قيمة العنصر الموجود في منتصف المصفوفة بالقيمة المُستهدفة، فإذا كانت هذه القيمة مُطابقة لقيمة لعنصر، تُرجع مرتبة العنصر في المصفوفة، وينتهي البحث. أمّا إذا كانت القيمة المُستهدفة أقل من قيمة العنصر، فإن البحث سيستمر في النصف السفلي من المصفوفة. وأمّا إذا كانت القيمة المُستهدفة أصغر من قيمة العنصر، فإن البحث سيستمر في النصف العلوي من المصفوفة. يُحدد العنصر الموجود في منتصف نصف المصفوفة، وتعاد العملية السابقة حتى الوصول إلى القيمة المُستهدفة أو انتهاء البحث دون ذلك، ويُقال عندها أن البحث غير ناجح أو أنه فشل في إيجاد القيمة المُستهدفة في المصفوفة. في كل تكرار، تستبعد الخوارزمية النصف الذي لا يمكن أن توجد فيه القيمة المُستهدفة.

الإجراء الرئيس

إذا كانت A مصفوفةً تحتوي على n عنصر أو سجل هي :

$A_{\{0\}}, A_{\{1\}}, A_{\{2\}}, A_{\{n-1\}}$ مرتبة بالشكل التالي:

$$A_{\{0\}} = A_{\{1\}} = A_{\{2\}} = \dots = A_{\{n-1\}}$$

وإذا كانت القيمة المُستهدفة هي T ، فبالإمكان العثور على مرتبة القيمة المُستهدفة T في المصفوفة A ، أي إنجاز بحثٍ ثنائي، باستخدام الخوارزمية التالية:

1- L و R هما متحولان، اضبط قيمة L إلى 0 وقيمة R إلى n-1 .

2- إذا كانت المُتراجحة $L > R$ ، البحث غير ناجح، نهاية.

3- m هو متحول يُمثّل قيمة موقع العنصر المتوسط، اضبط قيمته إلى قيمة الجزء الصحيح المحسوب بالعلاقة $L + R/2$ ، أي أن m هو أكبر عدد صحيح أصغر أو يساوي القيمة $L + R/2$.

4- $A_{\{m\}} < T$ هو متحول يمثل قيمة العنصر الذي مرتبته m في المصفوفة A :

1- إذا كانت المُتراجحة $A_{\{m\}} < T$ مُحققة، اضبط قيمة L إلى m+1 ثم انتقل إلى الخطوة 2 .

2- إذا كانت المُتراجحة $A_{\{m\}} > T$ مُحققة، اضبط قيمة R إلى m-1 ، ثم انتقل إلى الخطوة 2 .

3- وإلا، فإن المساواة $A_{\{m\}} = T$ مُحققة، أرجع قيمة m ، نهاية.

يتتبع هذا الإجراء التكراري حدود البحث باستخدام المتغيرين L و R اللذين يُمثَّلان دائماً الحدين الأسفل والأعلى على الترتيب لجزء المصفوفة الذي تُطبَّق خوارزمية البحث الثنائي عليه. يمكن أيضاً التعبير عن الخوارزمية السابقة بالشفرة التقريبية التالية، حيث تبقى أسماء وأنواع المتغيرات كما هي أعلاه، ويُمثَّل floor دالة الجزء الصحيح، وتشير الكلمة unsuccessful إلى قيمة رقمية محددة تُعبّر عن فشل البحث :

```
function binary_search(A, n, T) is
```

```
    L := 0
```

```
    R := n - 1
```

```
    while L ≤ R do
```

```
        m := floor((L + R) / 2)
```

```
        if A[m] < T then
```

```
            L := m + 1
```

```
        else if A[m] > T then
```

```
            R := m - 1
```

```
        else:
```

```
            return m
```

```
    return unsuccessful
```

يمكن أيضاً أن تُستبدل دالة المتمم الصحيح الأعلى `ceil` بدالة الجزء الصحيح المُطبَّقة على القيمة $L+R/2$.

ولكن هذا قد يؤدي إلى إعطاء نتيجة مُغايرة إذا وجدت القيمة المُستهدَفة أكثر من مرة في المصفوفة.

تقييم الأداء:

تكون الحالة الأسوأ عندما يصل البحث إلى أعظم مستوى في الشجرة الثنائية، أما الحالة الفضلة فتحصل عندما تكون القيمة المُستهدَفة هي العنصر المتوسط في المصفوفة. من حيث عدد المقارنات، يمكن تحليل أداء البحث الثنائي من خلال استعراض سير الإجراء على شجرة ثنائية. تكون العقدة الجذر في الشجرة هي العنصر المتوسط في المصفوفة. بعد ذلك، يكون العنصر المتوسط في النصف الأسفل من المصفوفة هو العقدة الفرعية اليسرى للجذر، والعنصر المتوسط في النصف الأعلى من المصفوفة هو العقدة الفرعية اليمنى للجذر. ثم يُبنى ما تبقى من الشجرة الثنائية باتباع نفس الطريقة. عند تطبيق الإجراء، يبدأ البحث من العقدة الجذر، وتكون هي العقدة المنظورة، وتُقارن قيمتها مع القيمة المُستهدَفة، فإذا كانت القيمة المُستهدَفة أكبر من قيمة

العقدة المنظورة تصبح العقدة المنظورة التالية هي العقدة الفرعية اليمنى للجذر، وإذا كانت القيمة المُستهدفة أصغر من قيمة العقدة المنظورة فإن العقدة المنظورة التالية تصبح العقدة الفرعية اليسرى للجذر، وتكرر عملية المقالة والتحريك على فروع الشجرة الثنائية نحو اليمين أو اليسار حسب ما سبق، حتى تكون القيمة المُستهدفة مساوية للقيمة المنظورة، وعندها البحث يتوقف بنجاح.

في أسوأ حالة، يكرر البحث الثنائي حلقة المقارنة $\log_2(n)+1$ مرة، حيث $\log_2(n)+1$ هو رمز دالة القسم الصحيح، وتعطي أكبر عدد صحيح يكون أصغر أو مساوٍ للقيمة التي تمرر لها، أي إذا مررنا 2.5 مثلاً، فإنها ستعيد القيمة 2. أما $\log_2(n)+1$ فهي دالة اللوغاريتم الثنائي. ويعني الوصول إلى الحالة الأكثر سوءاً بلوغ أعظم مستوى للشجرة. في شجرة بحث ثنائي عدد عناصرها n هناك دائماً $\log_2(n)+1$ مستوى. ويمكن أيضاً الوصول إلى الحالة الأكثر سوءاً عندما تكون القيمة المستهدفة غير موجودة في المصفوفة.

في أفضل الأحوال، تكون القيمة المستهدفة هي العنصر المتوسط للمصفوفة، وعندها ترجع مرتبة العنصر المتوسط بعد تكرار الإجراء مرة واحدة فقط.

من حيث عدد مرات تكرار الإجراء، ليس هناك أي خوارزمية بحث أخرى، تعتمد على مقارنة العناصر فقط، لها قيمة أداء وسطية أفضل من خوارزمية البحث الثنائي، وكذا الأمر في الحالتين الفضلى والأكثر سوءاً. تحتوي شجرة المقارنة التي تمثل البحث الثنائي على أقل عدد ممكن من المستويات، فجميع المستويات فيها، ما خلا المستوى الأخير، تكون ممتلئة بالعناصر بسعتها القصوى، وذلك لأن تقسيم المصفوفة إلى نصفين متساويين، أو قريبين من التساوي، تضمن أن تكون أحجام المصفوفات الفرعية الناتجة متساوية أو شبه متساوية وبالتالي تكون الشجرة الثنائية التي تمثل المصفوفة متوازنة بأفضل صورة ممكنة. بناء على ما سبق، يكون عدد العناصر المُستبعدة في كل تكرار أعظم ما يُمكن، ويقلل ذلك عدد مرات التكرار الوسطي اللازمة للعثور على القيمة المستهدفة، وكذا الأمر في الحالة الأكثر سوءاً. قد تعمل بعض خوارزميات البحث الأخرى التي تعتمد على المقارنة أسرع من خوارزمية البحث الثنائي من أجل بعض القيم المستهدفة، ولكن متوسط أداء الخوارزمية، من أجل عناصر المصفوفة جميعها يكون ذات قيمة أكبر من تلك التي تحققها خوارزمية البحث الثنائي.

اشتقاق الحالة المتوسطة

يعتمد متوسط عدد التكرارات التي يقوم بها البحث الثنائي على احتمال وجود كل عنصر يجري البحث عنه. تختلف الحالة المتوسطة لعمليات البحث الناجحة عن تلك المرتبطة بعمليات البحث غير الناجحة. تُعرّف الحالة المتوسطة لعمليات البحث الناجحة بأنها عدد التكرارات المطلوبة للبحث في كل عنصر في المصفوفة مرة واحدة بالضبط، مقسومةً على عدد العناصر n فيها. أما الحالة المتوسطة لعمليات البحث غير الناجحة فهي عدد التكرارات المطلوبة للبحث عن عنصر ما في كل المجالات الفاصلة مرة واحدة بالضبط، مقسومةً على عدد الفواصل، أي $n+1$

فيما يأتي، من أجل عمليات البحث الناجحة، سيُفترض أن البحث الثنائي يحصل بصورة متكافئة عن كل عنصر في المصفوفة، أمّا من أجل عمليات البحث غير الناجحة، فسيفترض أن البحث الثنائي يحصل بصورة متكافئة على الفواصل بين العناصر غير الموجودة في المصفوفة وعلى الفواصل بين العناصر غير الموجودة وتلك الموجودة في المصفوفة.