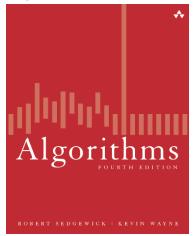


# Algorithms Stanford



 $\label{eq:week1} \label{eq:week1}$  Introduction, Union find and analyse of Algorithms.

Introduction

topic	data structures and algorithms			
data types	stack, queue, bag, union-find, priority queue	l T		
sorting	quicksort, mergesort, heapsort	part 1		
searching	BST, red-black BST, hash table			
graphs	BFS, DFS, Prim, Kruskal, Dijkstra	lτ		
strings	radix sorts, tries, KMP, regexps, data compression	part 2		
advanced	B-tree, suffix array, maxflow			



1. Fun and profit.



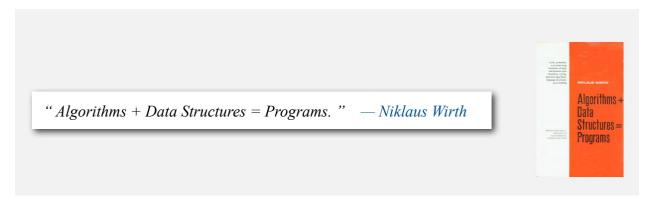
- 2. Power Al
- 3. Better code



4. Brain stimulation

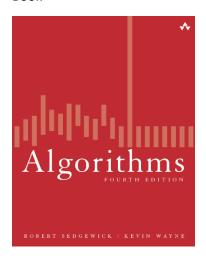
(Maths build brain muscles)

#### 5. Become best programmer.



#### Extra Materials

#### Book



# Book web page



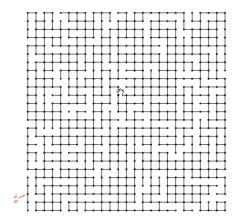
the number of components, thus, we arriculate the following AFT.

# public class UF UF(int N) void union(int p, int q) int find(int p) boolean connected(int p, int q) int count() int count() int count() int count() int count(int p) int intialize N sites with integer names (0 to N-1) add connection between p and q component identifier for p (0 to N-1) return true if p and q are in the same component number of components

Union-find API

#### Problem

The problem is when the big variety of object



Applications involved in in connectivity problem

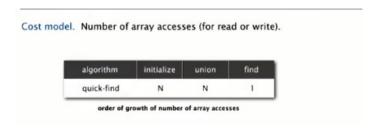
- 1. Pixels in digital photos
- 2. Computers in a networks
- 3. Friends in social network
- 4. Transistors in a computer chip
- 5. Elements in mathematical set
- 6. Variables in Fortran program
- 7. Metallic seites in composite system

#### Quick Find

It is very simple to implement

## Analyze

So the algorithm is very slow



To Union all elements it will take to run Union Operation N times what will be O( what is super slow.

And again quadratic algorithm is to slow to be preform. So we need to find better way.

# Quadratic algorithms do not scale

# Rough standard (for now).

- 109 operations per second.
- · 109 words of main memory.
- · Touch all words in approximately 1 second.

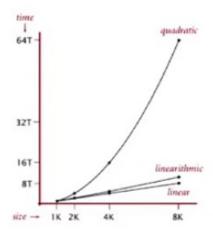
# a truism (roughly) since 1950!

# Ex. Huge problem for quick-find.

- 109 union commands on 109 objects.
- · Quick-find takes more than 1018 operations.
- · 30+ years of computer time!

# Quadratic algorithms don't scale with technology.

- · New computer may be 10x as fast.
- But, has 10x as much memory ⇒
  want to solve a problem that is 10x as big.
- · With quadratic algorithm, takes 10x as long!



#### QuickUnion

# Data structure. • Integer array id[] of Size N. • Interpretation: id[i] is parent of i. • Root of i is id[id[id[...id[i]...]]]. • Root of i is id[id[id[...id[i]...]]]. • Interpretation: id[i] is parent of i. • Root of i is id[id[id[...id[i]...]]]. • Interpretation: id[i] is parent of i. • Root of i is id[id[id[...id[i]...]]]. • Interpretation: id[i] is parent of i. • Root of i is id[id[id[id[...id[i]...]]].

**Implementation** 

```
9 references | O changes | O authors, O changes
public override int Find(int p) {
    while (Id[p]!=p)
        p = Id[p];
    return p;
}

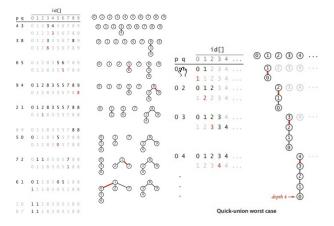
3 references | O changes | O authors, O changes
public override void Union(int p, int q) {
    int pRoot = Find(p);
    int qRoot = Find(q);

    Id[pRoot] = qRoot;

UnionCount--;
}
```

## Analyze

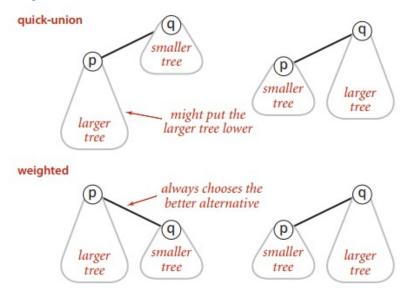
So in perfect world we have those trees that speed up over find operation.



## We

But some times we will have it linear tree. So it would be to slow to use Quick Union

## Weighted Quick Union



Weighted quick-union

In that way our tree can maximum can be

Analyze

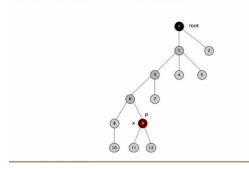
lg N

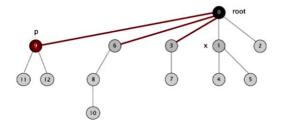
algorithm	initialize	union	connected	
quick-find	N	N	1	
quick-union	N	N †	N	
weighted QU	N	Ig N <sup>†</sup>	lg N	

† includes cost of finding roots

## Path Compression

Quick union with path compression. Just after computing the root of  $\rho$ , set the id of each examined node to point to that root.





# 1 Line of Code!!!!

```
private int root(int i)
{
   while (i != id[i])
   {
      id[i] = id[id[i]];
      i = id[i];
   }
   return i;
}
```

```
public override int Find(int p){
    while (Id[p] != p){
        Id[p] = Id[Id[p]];
        p = Id[p];
    }

    return p;
}
```

# Analyze

N	lg* N			
1	0			
2	1			
4	2			
16	3			
65536	4			
265536	5			
iterate log	function			

# Path Compression is close to be linear.

# Analyze

	order of growth for N sites (worst case)				
algorithm	constructor	union		find	
quick-find	N	N		1	
quick-union	N	tree height	₹~?	tree height	
weighted quick-union	N	$\lg N$		$\lg N$	
weighted quick-union with path compresson	N	very, very nearly, but not quite 1 (amortize (see EXERCISE 1.5.13)			
impossible	N	1		1	

Application

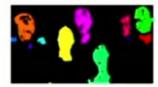
- Percolation.
- · Games (Go, Hex).

# ✓ Dynamic connectivity.

- · Least common ancestor.
- · Equivalence of finite state automata.
- · Hoshen-Kopelman algorithm in physics.
- · Hinley-Milner polymorphic type inference.
- · Kruskal's minimum spanning tree algorithm.
- · Compiling equivalence statements in Fortran.
- · Morphological attribute openings and closings.
- Matlab's bwlabel() function in image processing.

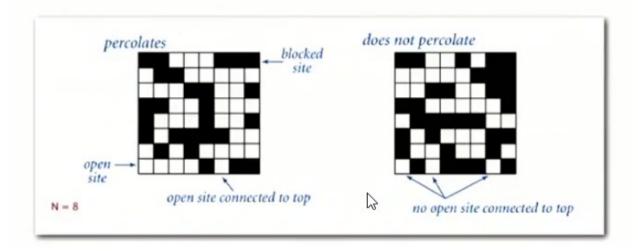






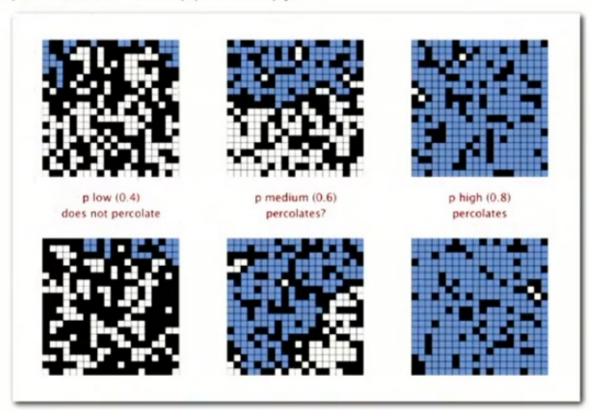
# A model for many physical systems:

- · N-by-N grid of sites.
- Each site is open with probability p (or blocked with probability 1-p).
- · System percolates iff top and bottom are connected by open sites.



# Likelihood of percolation

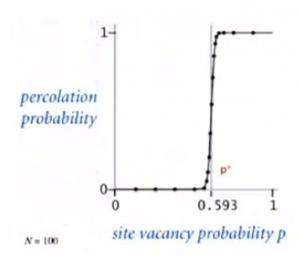
# Depends on site vacancy probability p.



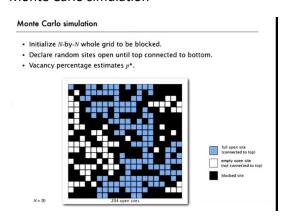
# Percolation phase transition

When N is large, theory guarantees a sharp threshold  $p^*$ .

- p > p\*: almost certainly percolates.
- p < p\*: almost certainly does not percolate.</li>
- Q. What is the value of  $p^*$ ?



#### Monte Carlo simulation

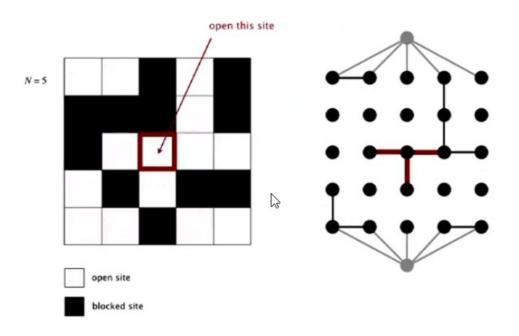


There is Monte Carlo simulation where we open block by block to see is there perculation or not

# Dynamic connectivity solution to estimate percolation threshold

- Q. How to model opening a new site?
- A. Connect newly opened site to all of its adjacent open sites.





Here is the way to represent the Percoalation problem using Quick Union way.

I can do it on my computer and run it infinite amount of times and get result for my self.

# Codility

# **codility** for Programmers

So codility have this lesson sequence that is used for TopTal but I would like to get max score to apply for toptal.

## Lesson 1: Iteration

For, while, foreach

Basically it talks about iterations.

#### Binary Gap

The task is basicly to made while cycle together with while operation.

So it is very basic and siple

60

## Lesson 2: Arrays

Reading material focus on basic arrays operations like element value chaning and array length determination. It focus on reversing the array

## CyclicRotation

The code is very basic just shift the array for count of elements.

```
public static int[] TaskSolution(int[] A, int K){
   int[] result = new int[A.Length];

for (int i = 0; i < A.Length; i++){
   K %= A.Length;
   int newIndex =K-i>0?A.Length-K+i:i-K;
   result[i] = A[newIndex];
}

return result;
}
```

#### Hidden trick

So the trick to the task was in one non describe detail

Number to shift K can be longer then array length

In such case for shifting our array we need to remove extra length by preforming

K%=A.Length;

And also you need to done task in O(N) time.

#### OddConcurrencyInArray

So all you need is mark number that alredy be counted.

And then check all the checked number and find one that wasn't dubble checked.

#### Hidden trick

The number can be to big so you need to go smart to mark it. You can't use as mutch space as you would like so you need to find way to count them on minimum space so I used trick with binary operations where each bit represent one number

And then using this technique I achived enough in total to get what I need.

```
1reference | 0 changes | 0 authors, 0 changes
public static bool NumberIsCheck(int a, int number)

{
    int shifted = 1 << number - 1;
    return (a & shifted) == shifted;
}

1reference | 0 changes | 0 authors, 0 changes
public static int MarkNumber( int a, int number)

{
    int shifted = 1 << number - 1;
    return a ^ shifted;
}
</pre>
```

## Lesson 3: Time complexity

Now it basically talking about the O(n)

About types I already well-known O(n) O() O(),O().

The good example case in the end in task where you need count

Arethmetical progression 1+2+3+4+...+N instead of using

You can do it very bad:

```
3.7: Slow solution — time complexity O(n^2).

1 def slow_solution(n):
2    result = 0
3    for i in xrange(n):
4         for j in xrange(i + 1):
5         result += 1
6    return result
```

Bad:

# 3.8: Fast solution — time complexity O(n).

```
1 def fast_solution(n):
2    result = 0
3    for i in xrange(n):
4         result += (i + 1)
5    return result
```

Or basicly remember that it was in math and it's basic arithmetical progression and you can easly apply to get result.

# 3.9: Model solution — time complexity O(1).

```
1 def model_solution(n):
2     result = n * (n + 1) // 2
3     return result
```

# FrogJmp

So the solution is just count the count of jumps in arithmetical way instead of any iteration.

#### Hidden Trick

So ther is alwas can be situation when you have or not have extra jump

But the calculation is based on divisionso you need manully preoform opperaiton if there is some rest after devision then add one extra jump because frog need to jump one more time toget throw that distance.

## PermMissingElement

The task was simply to find missing element using the lowest possible algorithm complexity.

#### Hidden trick

- 1. Using algebraial progression
- 2. Using longs to get totalsum of elemetns.

#### TypeEquilibrium

So the task was kind of tricky. So the solution was to made sum of all elemnts.

#### Hidden trick

- 1. Use longs for calculation sum
- 2. Pay attention to the minun numbers
- 3. Attention on ittration of two slices it counts to much elemnts but last shouldn't be counted

```
for (int i = 0; i < A.Length-1; I++){
    left += A[i];
    right -= A[i];
    long possibleMin = Math.Abs(left - right);
    if (possibleMin < result)
        result = possibleMin;</pre>
```

So if I have the three number

We should have 2 slices

$$P[0] = |100 - -95| = 195$$

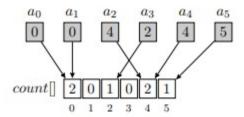
$$P[1] = |105 - 100| = 205$$

But if we let got all ittration of array what 3 then we will have one extra elemtn

$$P[2] = [5 -0] = 5$$

That would be mistake and it would not be seen in positive number but in negative it would be issaly missanderstand.

Lesson 4 : Counting Elements



So here we have system where each element is counted and count index stores the count of elements with such number But for such operation we need first of all know the max element.

And there are were interesting example with swaping one element

Where obvious solution is O() but my the smart moves we come to the length O(n+m).

Very smart and interesting example.

## FrogRiverOne

So the trick was in arithmetical progression Using counting technique in mix.

#### Hidden trick

- 1. A lot of test
- 2. Using longs for total sum
- 3. Mixed technique of counting.

#### Perm Check

So the task was just detect is there any missing element or not.

It was really basic if you think about it but. There is one major thing I should get noted about.

#### Hidden trick

1. Mix counting technique

- N is an integer within the range [1..100,000]
- each element of array A is an integer within the range [1..1,000,000,000]
- 2. IN the task description is clue that our integer can be bigger than our array. So we just check is array max bigger then array length.
- 3. Our our main integer could be smaller then aray length.

## Missing integer

So I need to find minimal positive inateger but ther could minus number just minus number and just queue of numbers.

#### Hidden trick

Unit testing. Unit testing for ever evething.

## Max Counters

It's easy you just do old trick

#### Hidden trick

It's sounds like you should made O(N\*M) by description, but you should actually made O(N+M)

## Lesson 5: Prefix sum

So we have been give the some array and we create new array containing the prefix sums of all possible combinations of first array.

	$a_0$	$a_1$	$a_2$	 $a_{n-1}$
$p_0 = 0$	$p_1 = a_0$	$p_2 = a_0 + a_1$	$p_3 = a_0 + a_1 + a_2$	 $p_n = a_0 + a_1 + \ldots + a_{n-1}$

It allow us to count the

## 5.1: Counting prefix sums — O(n)

```
def prefix_sums(A):
    n = len(A)
    P = [0] * (n + 1)
    for k in xrange(1, n + 1):
        P[k] = P[k - 1] + A[k - 1]
    return P
```

#### **5.2:** Total of one slice — O(1).

```
def count_total(P, x, y):
    return P[y + 1] - P[x]
```

#### Mushroom picker

So there is this exercise where you need to count how many mushrooms can pic the mushroom picker.

About the guy who pick the mushrooms.



The exercise is kind of complex so I would like to sort of rewrite it.

## **Passing Cars**

Simple task where we just need to use the prefix sum technique.

## Genomic range query

More fun task But the basic is the same all you need to do is made multiple prefix sums.

```
// A, C, G and T have impact factors of 1, 2, 3 and 4,
int[] prefixA = new int[S.Length + 1];
int[] prefixC = new int[S.Length + 1];
int[] prefixG = new int[S.Length + 1];

for (int i = 1; i < S.Length+1; i++){
    prefixA[i] = prefixA[i - 1] + (S[i-1] == 'A' ? 1 : 0);
    prefixC[i] = prefixC[i - 1] + (S[i-1] == 'C' ? 1 : 0);
    prefixG[i] = prefixG[i - 1] + (S[i-1] == 'G' ? 1 : 0);
}</pre>
```

I would say not that hard.

# Minimal average of two slice

Task is hard or sort of and from the first attempt I get only 60%

Candidate Report: Anonymous  Test Name:				
SUMMARY TIMELINE				
Test Score	Tasks in Test			
60 out of 100 points			Time Spent 10	Task Score
60%	MinAvgTwoSlice Submitted in: C#		59 min	60%
TASKS DETAILS				
1. MinAvgTwoSlice Find the minimal average of any slice containing at least two elements.	Task Score	Correctness 60%	Performa	60%

Than I found that guy who has kind of hobby of solving codility and Hacker rank tasks.



Prove by mathematical that any sequence can be made out of sub-sequence of 2 or 3 elements.

And even mathematical prove that:

Problem. See https://codility.com/programmers/task/min\_avg\_two\_slice/

Martin Kysel claimed that we only need to check slices of size 2 and 3 as the min average should be in the sub-slices.<sup>1</sup> In this note, I just want to give a formal proof for this observation.

Without loss of generality, consider a non-empty array  $A = A_1, A_2, \ldots, A_N$  and an interger 1 < i < N. We consider the average of A and its two sub-arrays  $A_1, \ldots, A_i$  and  $A_{i+1}, \ldots, A_N$ . Let

$$B = \frac{A_1 + \ldots + A_N}{N} \qquad C = \frac{A_1 + \ldots + A_i}{i} \qquad D = \frac{A_{i+1} + \ldots + A_N}{N - i}.$$

We claim that either  $C \leq B$  or  $D \leq B$  (\*). Indeed, assume the opposite, that is, B < C and B < D. We have that:

$$B < C \Leftrightarrow \frac{A_1 + \ldots + A_N}{N} < \frac{A_1 + \ldots + A_i}{i}$$

$$\Leftrightarrow i \times (A_1 + \ldots + A_N) < N \times (A_1 + \ldots + A_i)$$

$$\Leftrightarrow i \times (A_{i+1} + \ldots + A_N) < (N - i) \times (A_1 + \ldots + A_i)$$

$$\Leftrightarrow \frac{A_{i+1} + \ldots + A_N}{N - i} < \frac{A_1 + \ldots + A_i}{i}$$

$$\Leftrightarrow D < C$$

Similarly,  $B < D \Leftrightarrow C < D$ .

Then B < C and B < D means D < C and C < D, a contradiction! Therefore, claim  $(\star)$  holds. What follows is simple, since a slice of size bigger than 3 is composed of sub-slices of size 2 or 3.

#### What is sort of awesome.

#### CountDiv

So the task was to count the div between number find the amount of number devidebla by K between A and B. The solution is so simple that I would even not write the description.

```
Breferences | O changes | O authors, O changes
public static int TaskSolution(int A, int B, int K){
    int result = (B / K) - (A / K);

    if (A % K == 0)
        result++;
        I
    return result;
}
```

#### Lesson 6: Sorting

So whole idea of the lesson goes around our opportunite to sort things and we have this idea of using counter of elemts as sorting method.

many, just count the number of distinct pairs in adjacent cens.

#### 

In the example task we need to find count of unique values in the array N.

#### And what w

#### Distinct

Basically the same task as it was in the beginning but array can be empty so just notice that detail but otherwise normal sorting

#### MaxProductOfThree

Task was to find max multiplication of 3 elements in the array the hidden trick was that there could be minus number. If the biggest are two lest first numbers

So it can be A[0]\*A[1]\*A[N] or A[N]\*A[N-1]\*A[N-2] if the biggest are positive numbers.

Easy task.

## Triangle

So the thing was to find triangle. The solution was easy just sort it. And look for 3 elements one after other. If there such sequence then we have triangle one trick in task description

- N is an integer within the range [0..100,000];
- each element of array A is an integer within the range [-2,147,483,648, 2,147,483,647]

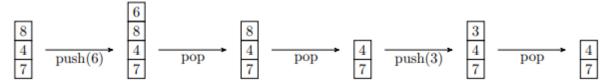
The sum of 2 sides can be sum of 2 int. MaxValue what is more then int so we need to use long.

But meanwhile it is simple task.

# Lesson 7: Stack and Queues

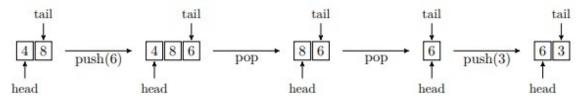
```
Stack<int> stack = new Stack<int>();
Queue<int> queue = new Queue<int>();
```

## Stack



## FIFO(First In First Out)

#### Queue

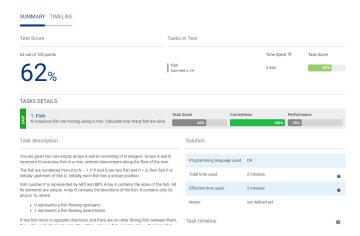


## LIFO (Last In First Out

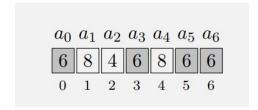
## Fish

I found solution with Stack but currently it is 62%

I should look for other solution. Tomrrow or maybe next time.

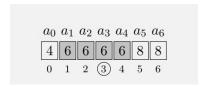


# Lesson 8: Leader

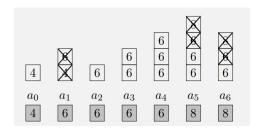


Basically saying we looking for element that occurs more then times. In given sequence A.

And there is basically two solutions:



1. Is simply sorting all elements and then choosing central one Because in order to be leader who should occur more then more then times



2. Smarter one using stack. We just add each element to stack and check if we have same element on the top of our stack the same we plan to currently add to the stack .

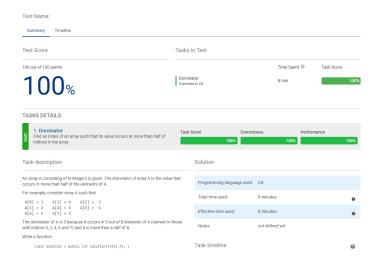
But in such cause it's not guaranty to be the leade it's only nown to be candidate for leader role.

So we need to check is value occurs more then time.

If yes then we have our leader else we need to return -1.

And after we recicle throw the array to check is our leader correct.

#### **Dominator**



It was simply the copy of the explanation task but with one small difference. It returns us the index instead of number itself .

Very easy task. Just perfect to get back

#### EquiLeader

I believe I found solution for the Equla Leader it is sort of obvious we just use 2 stacks instead of one.



# Lesson 9: Max Slice

```
\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ \hline 5 & -7 & 3 & 5 & -2 & 4 & -1 \end{bmatrix}
```

So for given sequency we have covered in grey the maximum sum.

So In word there is sort of way of inserting next b

```
9.4: Maximal slice — O(n).

1 def golden_max_slice(A):
2 max_ending = max_slice = 0
3 for a in A:
4 max_ending = max(0, max_ending + a)
5 max_slice = max(max_slice, max_ending)
6 return max_slice
```

Okay so it's sort of interesting I get max slic by using this simplet implementation in the world.

But it's still workable.

Okay so I run out of time but the concept is with me.

# Lesson 10: Prime and composite number

Prime number – the one that can be only divided by itself and by one

Composite number – has more then two devisors.

2

3

4

5

6

7

8

9

10

11

12

13

14

White – prime numbers

Gray – composite numbers.

# Finding the number of divisors of n

This turns out to be an application of the multiplication principle for counting things.

For example, suppose we want to count (or find all of) the divisors of n = 144.

Begin by forming the prime factorization of 144:

$$144 = 2^4 \cdot 3^2$$

So any divisor of 144 must be a product of some number of 2's (between 0 and 4) and some number of 3's (between 0 and 2). So here's a table of the possibilities:

	<b>2</b> 0	2 <sup>1</sup>	2 <sup>2</sup>	<b>2</b> <sup>3</sup>	<b>2</b> <sup>4</sup>
<b>3</b> 0	1	2	4	8	16
31	3	6	12	24	48
32	9	18	36	72	144

From the table, it's easy to see that there are  $5 \times 3 = 15$  divisors of 144.

In general, if you have the prime factorization of the number n, then to calculate how many divisors it has, you take all the exponents in the factorization, add 1 to each, and then multiply these "exponents + 1"s together.

Algorithm

```
f

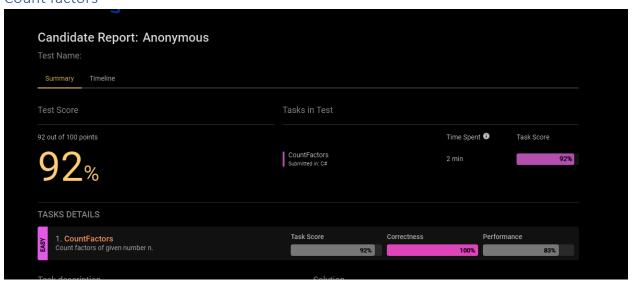
f references | 0 changes | 0 authors, 0 changes
public static int Solution(int number)
{
    int result = 0;
    int i = 1;
    while (i*i<number)
    {
        if (number % i == 0)
            result += 2;
        i++;
    }

    if (i * i == number)
        result++;

    return result;
}</pre>
```

To find all prime numbers till specific number we do this shit. I mean it's strange but still

# Count factors



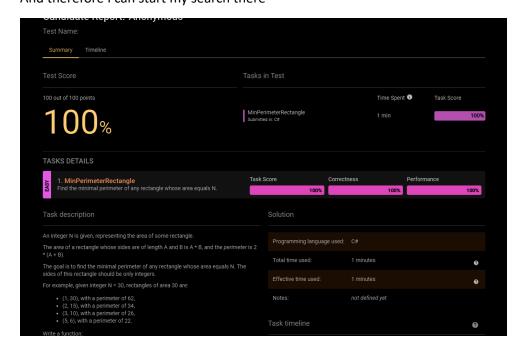
It was just rewriting the algorithm given in the book. And it's sort of sas that there should be other way of doing that.

# MinPeremeterRectangle

```
public static int Solution(int N)
{
    int i = (int) Math.Sqrt(N);
    while (i > 0)
    {
        if (N % i == 0)
        {
            int b = N / i;
            return (b + i) * 2;
        }
        i--;
    }
    return -1;
}
```

By the given formula I know that at least one of two multipliers should be less then sqrt(N)

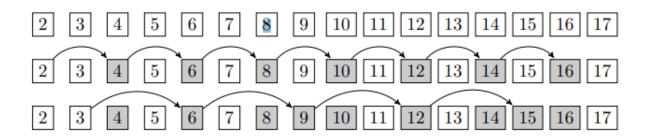
And therefore I can start my search there



By that solution I have the best performance garnty in many cases it will be almost linier.

Sieve of Eratosthenes

Lesson 11: Sieve of Eratosthenes



Lesson 11.1: Fractolization