10 minutes of Algorithms

and making them faster

If this isn't nerdy enough

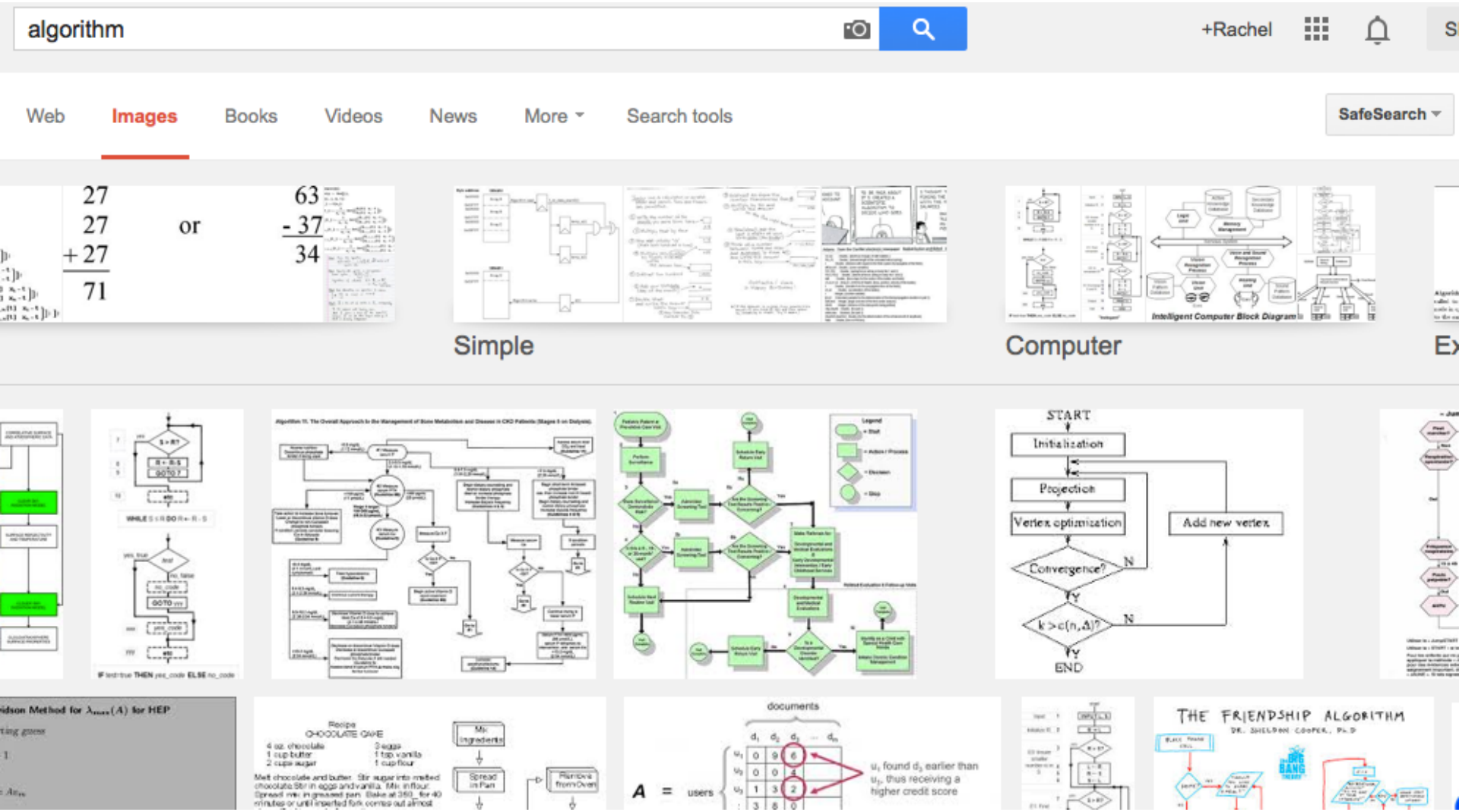
Here's a picture of me wearing 2 pairs of glasses at the same time

Algorithms?

"Any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values as output. An algorithm is thus a sequence of computational steps that transform the input into the output." [Cormen et al]

Translated from Math to English: a list of steps for performing a computation.

…are not very funny….



…but they are everywhere

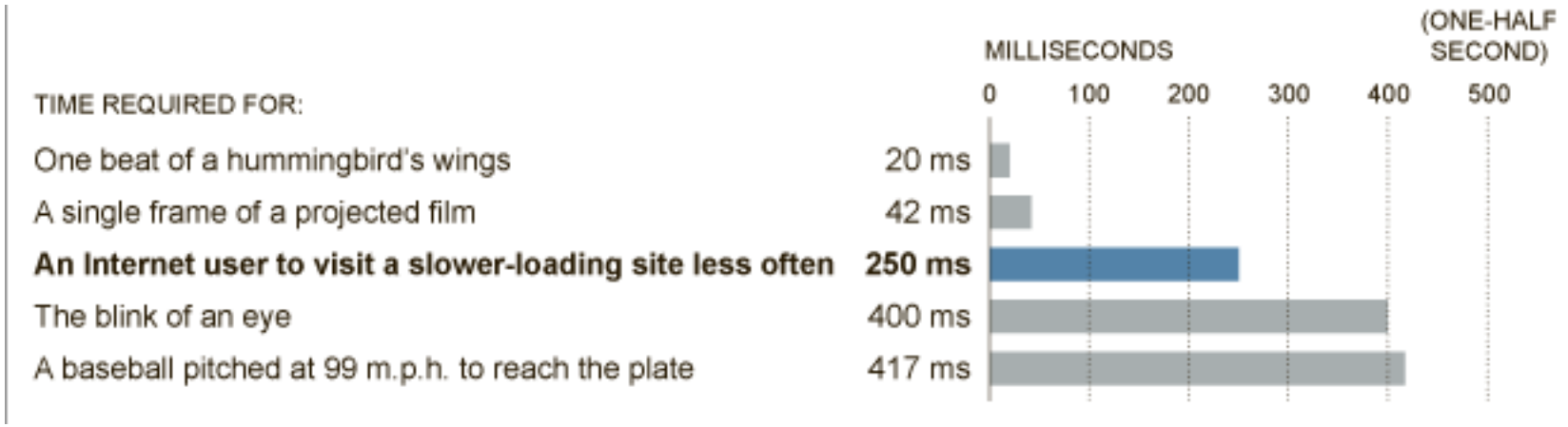
What is the most relevant page?

Who is likely to be a friend?

What product to advertise?

What movie to recommend?

It’s not enough to be right. You also have to be fast.



New York Times, “The Blink of an Eye? Oh Please”. Feb 29, 2012*

*before all their graphics were in D3

Sort a list from lowest to highest

1,4,3,2,7,9,5

Algorithm:

1. Reorder the list in every possible order
2. Check them until you find one that is correct

Sort a list from lowest to highest

1,4,3,2,7,9,5

Algorithm:

1. Reorder the list in every possible order
2. Check them until you find one that is correct

There are 5040 ways to order the 7 numbers.

Sort a list from lowest to highest

1,4,3,2,7,9,5

Algorithm:

1. Reorder the list in every possible order
2. Check them until you find one that is correct

There are 5040 ways to order the 7 numbers.

Computers are fast. This won't take long.

If the list contains 1000 numbers:

402387260077093773543702433923003985719374864210714632543799910429938512398629020592044208486969
404800479988610197196058631666872994808558901323829669944590997424504087073759918823627727188732
519779505950995276120874975462497043601418278094646496291056393887437886487337119181045825783647
849977012476632889835955735432513185323958463075557409114262417474349347553428646576611667797396
668820291207379143853719588249808126867838374559731746136085379534524221586593201928090878297308
431392844403281231558611036976801357304216168747609675871348312025478589320767169132448426236131
412508780208000261683151027341827977704784635868170164365024153691398281264810213092761244896359
928705114964975419909342221566832572080821333186116811553615836546984046708975602900950537616475
847728421889679646244945160765353408198901385442487984959953319101723355556602139450399736280750
137837615307127761926849034352625200015888535147331611702103968175921510907788019393178114194545
257223865541461062892187960223838971476088506276862967146674697562911234082439208160153780889893
964518263243671616762179168909779911903754031274622289988005195444414282012187361745992642956581
746628302955570299024324153181617210465832036786906117260158783520751516284225540265170483304226
143974286933061690897968482590125458327168226458066526769958652682272807075781391858178889652208
164348344825993266043367660176999612831860788386150279465955131156552036093988180612138558600301
435694527224206344631797460594682573103790084024432438465657245014402821885252470935190620929023
136493273497565513958720559654228749774011413346962715422845862377387538230483865688976461927383
814900140767310446640259899490222221765904339901886018566526485061799702356193897017860040811889
729918311021171229845901641921068884387121855646124960798722908519296819372388642614839657382291
123125024186649353143970137428531926649875337218940694281434118520158014123344828015051399694290
153483077644569099073152433278288269864602789864321139083506217095002597389863554277196742822248
757586765752344220207573630569498825087968928162753848863396909959826280956121450994871701244516
461260379029309120889086942028510640182154399457156805941872748998094254742173582401063677404595
741785160829230135358081840096996372524230560855903700624271243416909004153690105933983835777939
410970027753472000000000000000000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000

different arrangements are possible

A better way to sort?

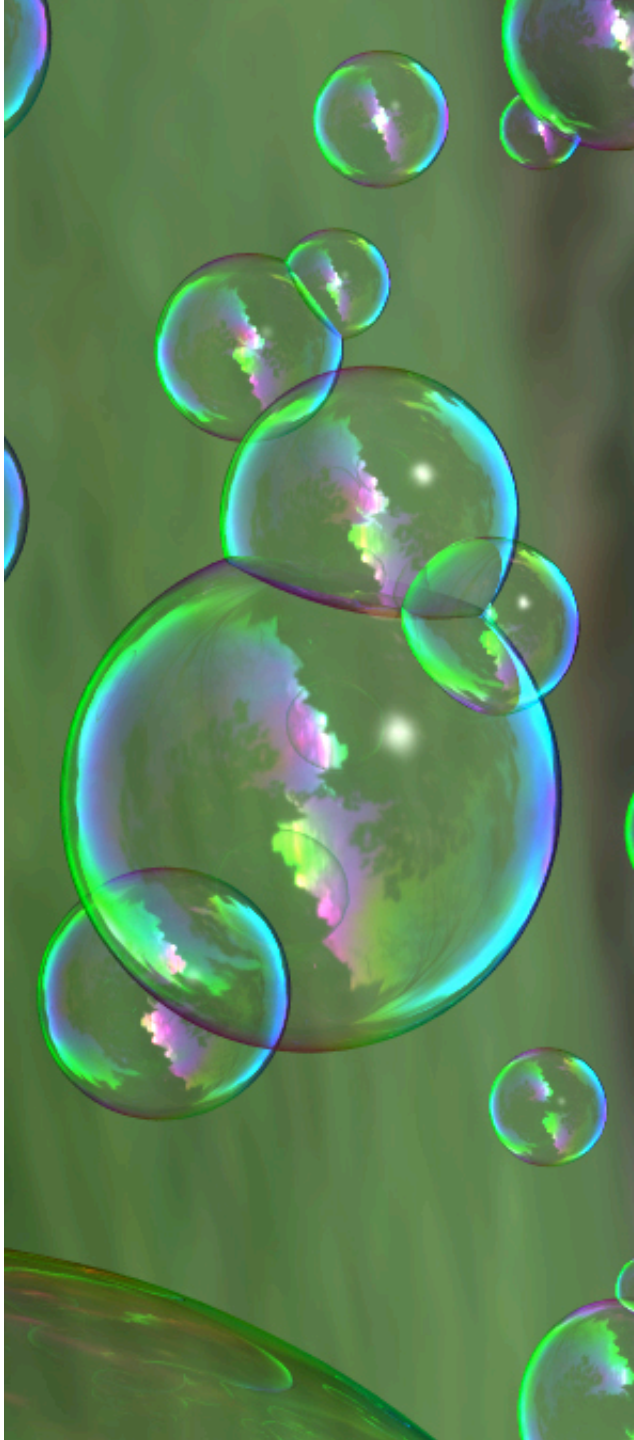
1,4,3,2,7,9,5

Algorithm:

1. Starting with the second number, check whether it's smaller than the number in front of it
2. If it is, swap them
3. Keep swapping until you hit a smaller number
4. Move on to the next number you haven't checked

Example

Start: 1,4,3,2,7,9,5
Compare 1 and 4: 1,4,3,2,7,9,5
Compare 4 and 3: 1,4,3,2,7,9,5
Switch: 1,3,4,2,7,9,5
Compare 1 and 3: 1,3,4,2,7,9,5
Compare 2 and 4: 1,3,4,2,7,9,5
Switch 2 and 4: 1,3,2,4,7,9,5
Compare 2 and 3: 1,3,2,4,7,9,5
Switch 2 and 3: 1,2,3,4,7,9,5
Compare 1 and 2: 1,2,3,4,7,9,5
etc



Worst case?

- You have to compare and move every number
- For 7 items, this requires 49 moves
- For 1000 items, it requires 1,000,000 moves

1,000,000 is a lot less than

402387260077093773543702433923003985719374864210714632543799910429938512398629020592044208486969
404800479988610197196058631666872994808558901323829669944590997424504087073759918823627727188732
519779505950995276120874975462497043601418278094646496291056393887437886487337119181045825783647
849977012476632889835955735432513185323958463075557409114262417474349347553428646576611667797396
668820291207379143853719588249808126867838374559731746136085379534524221586593201928090878297308
431392844403281231558611036976801357304216168747609675871348312025478589320767169132448426236131
412508780208000261683151027341827977704784635868170164365024153691398281264810213092761244896359
928705114964975419909342221566832572080821333186116811553615836546984046708975602900950537616475
847728421889679646244945160765353408198901385442487984959953319101723355556602139450399736280750
137837615307127761926849034352625200015888535147331611702103968175921510907788019393178114194545
257223865541461062892187960223838971476088506276862967146674697562911234082439208160153780889893
964518263243671616762179168909779911903754031274622289988005195444414282012187361745992642956581
746628302955570299024324153181617210465832036786906117260158783520751516284225540265170483304226
143974286933061690897968482590125458327168226458066526769958652682272807075781391858178889652208
164348344825993266043367660176999612831860788386150279465955131156552036093988180612138558600301
435694527224206344631797460594682573103790084024432438465657245014402821885252470935190620929023
136493273497565513958720559654228749774011413346962715422845862377387538230483865688976461927383
814900140767310446640259899490222221765904339901886018566526485061799702356193897017860040811889
729918311021171229845901641921068884387121855646124960798722908519296819372388642614839657382291
123125024186649353143970137428531926649875337218940694281434118520158014123344828015051399694290
153483077644569099073152433278288269864602789864321139083506217095002597389863554277196742822248
757586765752344220207573630569498825087968928162753848863396909959826280956121450994871701244516
461260379029309120889086942028510640182154399457156805941872748998094254742173582401063677404595
741785160829230135358081840096996372524230560855903700624271243416909004153690105933983835777939
410970027753472000000000000000000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000

Takeaway

- “Algorithm” refers to the method of solving, not the solution
- There are multiple algorithms for solving any problem
- Bad ones won’t scale. They’ll crash your website when you have enough data
- Or fail to run in your lifetime (really!)
- Clever algorithms return the same solution, they just do it faster

Implementations of bubble sort and permutation sort (and also these slides) at https://github.com/rshorey/cfp_sorting_algorithms