

Training für Programmierwettbewerbe - Exam Questions

Hanno Barschel (174761)

1 Introduction

(1) Zeit Komplexität von drei Funktionen:

- **foo(.)** hat eine Zeitkomplexität von $\mathcal{O}(n)$ (wobei n die Länge des gegebenen Vektors ist) und eine Platzkomplexität von $\mathcal{O}(n)$.

Das folgt direkt daraus, dass wir genau 2 Vorschleifen die jeweils den gesamten Vektor durchgehen haben also $\mathcal{O}(2n) = \mathcal{O}(n)$.

- **print _ pairs(.)** hat eine Zeitkomplexität von $\mathcal{O}(n^2)$ und eine Platzkomplexität von $\mathcal{O}(n)$.

Für jedes Element des Vektors ruft die Funktion nochmal jedes Element des Vektors auf und printed beide als Paar aus. Somit kommt eine Zeitkomplexität von $\mathcal{O}(n^2)$ zustande.

- **print _ unordered _ pairs(.)** hat eine Laufzeitkomplexität von $\mathcal{O}(n^2)$ und eine Platzkomplexität von $\mathcal{O}(n)$.

Für jedes Element des Vektors rufen wir alle Folgeelemente auf. Das sind es $n - (i + 1)$ Aufrufe für Element i (i startet bei 0). Damit kommen wir auf eine Laufzeit von

$$\sum_{i=0}^{n-1} (n - (i + 1)) = n^2 - \sum_{i=1}^n i = n^2 - \frac{n \cdot (n + 1)}{2} = n^2 - \frac{n^2}{2} - \frac{1}{2} \in \mathcal{O}(n^2).$$

- **all _ fib(.)** hat eine Zeitkomplexität von $\mathcal{O}(n)$ und eine Platzkomplexität von $\mathcal{O}(n)$.

Die for-Schleife printed uns im Prinzip alle Fibonacci Zahlen bis n aus. Sobald wir eine Null in $memo[i]$ finden sind aber bereits $memo[i - 1]$ und $memo[i - 2]$ berechnet, somit müssen wir diese nur addieren und haben keine wirklich tiefe Rekursion.

(2) Algorithmen Beispiele:

- $\mathcal{O}(1)$: Für eine gegebene Hashtabelle und einen Index gibt der Algorithmus den Wert für den Index zurück.
- $\mathcal{O}(\log(n))$: In einer sortierten Liste ein gegebenes Element suchen.
- $\mathcal{O}(n)$: Der Algorithmus berechnet zu einem gegeben Array die Summe aller Elemente und gibt diese zurück.
- $\mathcal{O}(n \cdot \log(n))$: Ein guter Sortier Algorithmus (Merge Sort).
- $\mathcal{O}(n^2)$: Quicksort (worst-case Analyse).

- (3) Ersteinmal natürlich ordentlich zuhören wenn die Frage gestellt ist. Dabei sollten wir uns alles merken und auch auf Besonderheiten achten. Dann wählen wir ein geeignetes Beispiel. Dabei sollten wir immer im Hintergrund behalten, dass oft Spezialfälle besonders zu betrachten sind. Anschließend schreiben wir einen brute force Algorithmus auf und betrachten dessen Zeit sowie Platzkomplexität. Jetzt optimieren wir diesen noch indem wir das Ganze möglichst einfach machen, überflüssige und doppelte Arbeit vermeiden. Mit diesen Betrachtungen verbessern wir den Brutforce Algorithmus schließlich noch.
- (4) Gegeben sei eine Liste mit Natürlichen Zahlen. Gesucht sei die maximale Differenz zweier Werte im Array wobei der vordere (steht weiter vorne im Array) Wert kleiner als der hintere sein muss. Wir iterieren mittels zweier indexe, *tmp_min* und *tmp* die am Anfang beide auf 0 gesetzt werden. Gleichzeitig haben wir die Variable *max_diff* = 0 in der die bisher gefundene maximale Differenz gespeichert wird. Nun beginnen wir mittels einer for-Schleife mit Variablen *tmp* über den Array zu iterieren: Ist der Wert im Array an Stelle *tmp* kleiner oder gleich als der an der Stelle *tmp_min* so setzen wir *tmp_min* auf *tmp*. Ist *tmp* größer so berechnen wir die Differenz zwischen *tmp* und *tmp_min*. Ist diese größer als *max_diff* so ersetzen wir diese, ansonsten inkrementieren wir *tmp* einfach weiter.
- (5) Ich würde es mit einer Hashmap checken: Wir testen erst einmal ob die Strings die gleiche Länge haben (wenn nicht dann return 0). Danach erstellen wir 2 Hashmaps (für jedes Wort eine). Anschließend gehen wir mittels einer for-Schleife beide strings durch und für jeden Buchstaben inkrementieren wir den zugehörigen Hashwert. Abschließend vergleichen wir beide Hashmaps.

return(hashmap_string1 == hashmap_string2).

2 Arrays and Bits

- (1) Für geordnete Mengen, wo jeder Eintrag einfach nur angibt ob das Element in der Menge ist oder nicht, können wir einfach die Bitoperationen anwenden um beispielsweise leicht die Anzahl an Elementen herausfinden die in beiden Mengen enthalten sind (durch den & Operator und danach rufen wir die count-Funktion auf). Zusätzlich können wir neue Mengen erstellen die...
- ...alle Elemente enthalten die mindestens in einer der beiden Listen enthalten sind.
 - ...alle Elemente enthalten die nur in genau einer der beiden Listen enthalten sind.

Auch können wir sehr leicht das Komplement, Vereinigung, Durchschnitt und Differenz bilden. Natürlich geht noch viel mehr. Bitoperationen haben den Vorteil, dass sie deutlich schneller sind, als über einen Array zu iterieren.

- (2) Negative vorzeichenbehaftete Integer haben ein Vorzeichenbit, dass auf 1 gesetzt ist. Ein logischer Rechts shift würde jetzt einfach alle bits nach rechts schieben und in die höchste Stelle eine 1. Somit wäre das Resultat dann plötzlich positiv.

Ein arithmetischer Rechtsshift lässt das erste Bit auf 1 und ersetzt die Bits nach dem Vorzeichenbit mit 1 (so viele wie geschoben wird). Damit bleibt die Zahl negativ und die Operation entspricht (etwa) einer Division mit 2, was unserer Vorstellung von einem Rechtsshift entspricht.

- (3) $(n \& (n - 1)) == 0$ testet ob die bitdarstellung von n und $n - 1$ 0 ergeben. Das ist der Fall, falls $n = 0$, $n = 1$ ist ($\&0$ ergibt immer 0). Angenommen n ist eine 2er Potenz (also an k .ter Stelle eine 1 im bitstring). Dann ist in $n-1$ genau jede 0 die vor der k .ten Stelle kommt auf 1 gesetzt und die 1 in der k .ten Stelle wird zu 0. In den höheren Stellen ändert sich nichts. Somit ergibt für diesen Fall die Formel auch 0.

Angenommen n ist nicht durch 2 teilbar. In diesem Fall werden in $n-1$ alle Nullen hinter der ersten 1 in n auf 1 gesetzt und die erste 1 auf 0. Nach der Annahme, dass n keine 2er Potenz ist, existiert jetzt ein höheres bit, dass sowohl in n als auch in $n-1$ auf 1 gesetzt ist. Somit ergibt die Formel nicht 0.

- (4) Wir wollen die Anzahl der Einserbits in der Variablen `sum` speichern.

Wir gehen in einem for loop von 0 bis `sizeof(integer)` über jeden einzelnen bit indem wir die gegebene Zahl n immer um 1 shiften und dann $n \& 1$ auf unsere Summe addieren. $n \& 1$ ist genau gleich dem letzten bit und so ergibt der Algorithmus die geforderte Summe.

- (5) Zum reversen aller Bits in einem Integer benutzen wir einfach den `__builtin_bswap16` Operator aus der Vorlesung. Dieser flippt jedes einzelne Bit.

- (6) Die Quicksortpartitionierungsfunktion gibt für eine übergebene Liste, eine Liste aus wo das Pivotelement bereits an der richtigen Stelle steht. Nun gehen wir die Liste durch und testen für jedes Element ob es größer als das Pivotelement ist. Ist das der Fall so tauschen wir es nach hinten.

Wenn wir Einträge nach gerade und ungerade sortieren wollen können wir ähnlich vorgehen nur, dass wir wir testen ob das Element gerade ist. Ist das der Fall so tauschen wir es nach hinten.

3 Arrays and Strings

- (1) Als erstes testen wir ob das erste Zeichen ein '-' ist, ist das der Fall so starten wir die for-Schleife Index 1 ansonsten bei Index 0. Wir initialisieren einen unsigned Integer mit 0 und starten die besagte for-Schleife. Diese geht von einem Index (entweder 0 oder 1) bis zur Länge des Strings alle string Teile durch. Für jeden Durchgang multiplizieren wir *result* mit 10 um alle bisherigen Zahlen um 1

nach links zu verschieben. Jeder Charakter wird mittels $s[i] - '0'$ zu einem Integer zwischen 0 und 9 konvertiert und auf *result* addiert.

- (2) Um auf die einzelnen Dezimalziffern unseres Integers zuzugreifen verwenden wir den modulo Operator. Solange die gegebene Nummer nicht 0 ist rechnen wir modulo 10, bekommen so die letzte Ziffer, wandeln diese mit *char(.) + '0'* in einen char um, der in Ascii genau der Ziffer entspricht und fügen diesen an unseren String an. Anschließend teilen wir die Nummer durch 10. Jetzt haben wir allerdings immer nur die kleinste Ziffer genommen also müssen wir noch den string reversen.
- (3) ASCII hat 128, was relativ wenig sind, aber für viele Texte reicht es und somit ist die Kodierung optimal, weil wir pro Buchstabe nur 8 bit brauchen (bzw 7 aber das wird meist dann auch in 8 gespeichert).

EXTENDED ASCII ist ähnlich wie ascii nur haben wir das Doppelte an Bits (256) also brauchen wir zur Speicherung genau 8 bits.

UTF-8 ist eine dynamische Kodierung, d.h. wir haben einen Präfix der uns angibt wie viele bytes wir für einen Charakter lesen müssen. Damit können wir sowohl normales Ascii umsetzen als auch Sonderzeichen in mehrere Bytes kodieren.

UTF-16 nutzt 16 oder 32 bits und hat somit auch dynamische Länge. Für normales Ascii bietet sich nicht an, weil wir dann immer 16 bits hätten wovon wir aber nur 7 benötigen.

UTF-32 nutzt 32 bit zur Speicherung und ist somit von fester Länge. Hier brauchen wir keinen Präfix, weil wir feste Länge haben, also nutzen wir unseren Platz besser aus. Andererseits wäre diese Kodierung eine Platzverschwendung für Texte mit Ascii Kodierung.

- (4) Um einen String so schnell wie möglich zu säubern nutze so viel wie möglich inplace Replacememt, wenn das Wort kürzer ist, und lösche am Ende redundante Lücken. Falls du Regex Ausdrücke verwendest versuche diese für mehr Geschwindigkeit und Effizienz hard zu coden. Nutze so wenig Fallunterscheidungen wie möglich.
- (5) Die regex Klasse macht viele Fallunterscheidungen und ist somit relativ langsam. **nicht inplace?**.
- (6) Run-length Encoding wird für Daten genommen, die viele sich wiederholende Charaktere haben. Bei normalen Strings ist das meistens nicht der Fall also eher nicht. Eher dort wo immer ähnliche Muster auftreten oder es Blöcke, die nur einen Buchstaben haben gibt.

4 LinkedLists und so

- (1) Sei eine verlinkte Liste und ein Element daraus gegeben, dass wir in konstanter Zeit löschen wollen.

Das geht indem wir die Daten aus dem Nachfolger in unseren Knoten schreiben und den pointer auf die übernächste Node (`node->next->next`) zeigen lassen. Das funktioniert nur wenn unser Knoten nicht das Ende der Liste ist. Ist das der Fall so setzen wir den Knoten einfach zu NULL.

- (2) Sei erneut eine einzeln verlinkte Liste gegeben. Um jetzt einen Pointer zu erstellen der beispielsweise genau auf die Mitte zeigt können wir folgendermaßen vorgehen:
Wir erstellen 2 Pointer. Der eine läuft immer ganz normal einen Schritt nach vorne, während der Zweite immer 2 Schritte macht. Das Ganze kapseln wir in eine while Schleife, die läuft solange der schnelle Pointer nicht am Ende ist. Erreicht der schnelle Pointer jetzt das Ende so ist der langsame Pointer in der Mitte. Das geht deutlich schneller und effektiver als die naive Methode erst mit einem Pointer die Länge bestimmen und dann nochmal von vorne bis zur Mitte laufen.
- (3) Statt eine extra Stack klasse zu implementieren benutzen wir einfach Vektoren und die dafür definierten Methoden *push_back()*, *empty()*, *pop_back* und *back()*.
- (4) Um eine Liste als queue zu verwenden benutzen wir die Methoden *emplace_back()*, die ein Element hinten an die Liste anhängt. Zusätzlich benutzen wir *front()* um das erste Element zu erhalten sowie *pop_front()* um das erste Element zu entfernen.
- (5) Ein `BinaryTree` ist einfach ein Baum, indem jeder Knoten maximal 2 Kinder hat. Ein `BinarySearchTree` hingegen ist ein `BinaryTree`, bei dem für jeden Knoten gilt: Das Maximum des linken Teilbaumes ist kleiner oder gleich als die Wurzel des Teilbaumes. Das Minimum des rechten Teilbaumes ist größer als die Wurzel des Teilbaumes.
- (6) Es gibt 3 verschiedene Möglichkeiten einen `BinaryTree` zu traversieren:
 - Preorder: Hier betrachten wir zuerst den jeweiligen Knoten wo wir sind, gehen dann erst nach Links und machen hier wieder das Selbe und anschließend noch nach rechts. Da wir erst die Knoten immer betrachten kann man mit diesem Algorithmus gut eine Kopie des Baumes machen.
 - Inorder: Wir gehen erst nach Links, wenn im linken Teilbaum alles abgearbeitet ist betrachten wir unseren Knoten und gehen anschließend noch in den rechten Teilbaum und wiederholen das Ganze. Da in einem `BinarySearchTree` links immer die kleinsten Werte stehen gibt uns dieser Algorithmus hierfür eine sortierte Liste mit allen Werten.
 - Postorder: Bei postorder gehen wir erst von einem Knoten erst in den linken Teilbaum, dann in den Rechten und schließlich betrachten wir die Wurzel des Teilbaumes. Diese Reihenfolge hat zur Folge, dass wir erst alles Kinder eines Knotens betrachten, bevor der Knoten selbst an der Reihe ist. Somit eignet sich dieser Algorithmus zum Löschen eines Teilbaumes.

- (7) Ein trie ist im Prinzip ein Baum wo jeder Knoten beliebig viele Kinder haben kann. Die Wurzel ist natürlich ausgezeichnet. In jedem Knoten steht ein Charakter und somit repräsentiert jeder Pfad von der Wurzel zum Blatt ein Wort.

Ein radix Tree ist ähnlich aufgebaut, nur sind an den Knoten statt einzelner Buchstaben, Strings gegeben (Können auch einzelne Buchstaben sein). Somit haben wir einfach nur einen Speicher effizienteren trie.

5 Heaps, Sorting und Hash Tabellen

- (1) Ein Heap ist ein Binärer Baum, der vollständig ist (shape property). Dadurch kann man ihn als Array speichern. Zusätzlich gilt die heap property, also, dass der Wert im Elternknoten mindestens so groß ist wie in den Kindern (Max Heap).
- (2) Wir benutzen einen Min Heap. Hierfür gehen wir den Daten Stream Elementweise durch. beinhaltet der Heap jetzt schon k Elemente und ist unser Wort kürzer als die Wurzel, so kommt das Wort nicht in Frage und wir gehen weiter. Ansonsten pushen wir das Element in die Wurzel und falls die Länge jetzt größer als k ist, so löschen wir die Wurzel (ist das kürzeste Element nach der Definition von min Heap). Am Ende lesen wir des Heap aus.
- (3) Auch hierfür benutzen wir einen Min Heap. Wir beginnen damit für jeden Datenvektor einen Verweis auf die Liste in den Min Heap zu pushen. Ist der Min Heap fertig aufgebaut, so popen wir das top Element (Kleinstes nach Definition von min Heap). Aber das Element ist ja nur ein Verweis auf eine längere Liste. Insofern schauen wir jetzt noch ob danach noch Elemente kommen. Ist das der Fall so pushen wir den Verweis auf das nächste Element wieder in den Min Heap. Das machen wir solange bis der min Heap leer ist.
- (4) Sorting Networks sind aus kleinen Bauelementen (die im Prinzip 2 gegebene Zahlen entweder vertauschen oder nicht) aufgebaut und mit ihnen kann man Sortieren hart codieren. Für kleine Array Länge ist das schneller als normale Sortieralgorithmen.
- (5) Sortierte Arrays sind praktischer, weil man nicht immer den ganzen Array durchsuchen muss um ein Element zu finden. In einem sortierten Array steht dieses Element an einer ganz bestimmten Stelle. Für viele Mengenoperationen (bspw. Schnitt oder Vereinigung) muss man ständig Elemente suchen und mit sortierten Arrays braucht man jetzt nur die Arrays Elementweise durchgehen und vergleichen.
- (6) Gegeben sei eine Hash Funktion die eine Menge von Wörtern auf eine Hash Tabelle abbildet. Hat diese Funktion keine Kollisionen so ist es eine perfekte Hash Funktion. Hat der Hash zusätzlich in seiner Hash Tabelle keinen ungenutzten Platz, so heißt er minimal. Ein minimal perfekter Hash ist somit ein Hash der keine Kollisionen bildet und jeden Platz in der Hashtabelle ausnutzt.

6 Algorithmisches Design

- (1) Backtracking ist ein Verfahren, das ein Problem löst indem es einzelne mögliche Züge simuliert und, falls es einen Widerspruch gibt, zurücksetzt. Für höhere Effektivität benötigen wir ein gutes pruning Verfahren.
- (2) Wir beginnen damit eine Queen zu platzieren. Jetzt schauen wir wie viele mögliche Queens wir in die nächste Reihe setzen können (mittels Rekursion). Das machen wir solange bis jede Reihe eine Queen hat oder ein Fehler auftritt (in diesem Fall setzen wir das Brett um einen Zug zurück).
- (3) Um das 15 Puzzle zu lösen müssen wir auf jeden Fall eine Liste mit allen bisher besuchten States mitführen, um keine Loops zu kreieren (Das Board schicken wir in eine Hash Funktion um nicht das aufwendige Board speichern zu müssen.).

Wenn wir jetzt alle Kandidaten für den nächsten Zug suchen schauen wir für alle möglichen Züge ob der Board State schon erreicht wurde. Ist das der Fall so ignorieren wir diesen Kandidaten. Wenn nicht, so hängen wir den jetzigen Hash an und hängen den Kandidaten an den Rückgabvektor an.

Um effektiv zu prunen ist es außerdem hilfreich eine Abstandsfunktion einzuführen, die abschätzt wie viele Züge wir mindestens noch brauchen um auf die perfekte Lösung zu kommen.

- (4) Meist führt man schon sehr viel aus indem man das Problem in Subprobleme zerlegt. Auf der ersten Ebene ist eine Parallelisierung nicht trivial, da wir ja das ganze Problem erst aufteilen müssen.

Beispielsweise müssen wir für quicksort erst die gesamte Liste nach dem Pivot teilen. Hier ist eine Parallelisierung nicht ganz trivial. Danach können wir einfach für jede entstandene Hälfte einen Thread starten.

- (5) Weil wir bei Decrease and Conquer in jedem Schritt das Problem zwar aufteilen, aber danach nur eine Hälfte des Problems betrachten. Somit verkleinern wir das Problem eher. Bei Divide and Conquer teilen wir das Problem auf.
- (6) Dynamische Programmierung funktioniert ähnlich wie Divide and Conquer nur das wir hier bei der Lösung für ein Teilproblem suchen ob wir das Problem schon einmal gelöst haben (in beispielsweise einem Vektor lsg). Ist das der Fall so schreiben wir die Lösung einfach ab. Falls nicht so lösen wir das Problem und hängen die Lösung an einen Vektor (lsg) an. Somit verhindern wir, dass wir Dinge doppelt berechnen.
- (7) Dynamische Programmierung bottom up startet von allen Teilproblemen und löst die darüberliegenden Probleme mit diesen. Das ist meist leichter zu verstehen und wir brauchen für die Lösungen nur einen kleineren Cache (haben wir die unterste Ebene gelöst so brauchen wir die Werte meist nicht mehr, beispielsweise für die Fibonacci Zahlen brauchen wir 0 und 1 für die erste Ebene, aber danach nur noch 1, 2 bei richtiger Implementierung, also können wir 0,1 aus dem Cache löschen).

Bei top down starten wir mit der Aufgabe und zerlegen diese meist rekursiv in Teilprobleme. Diese lösen wir und speichern sie in einem Cache. Hier haben wir den großen Vorteil, dass wir pruning anwenden können, also das Abschneiden von Ästen die nicht mehr zur Lösung führen können.

- (8) Angenommen wir haben Gegenstände mit unterschiedlichen Werten und Gewichten. Wir wollen so viel Wert wie möglich wegtragen, aber können nur maximal ein bestimmtes Gewicht tragen. Ein greedy Algorithmus könnte jetzt beispielsweise sagen, dass wir, solange wir noch mehr tragen können, den wertvollsten Gegenstand, der leicht genug ist, einpacken. Das führt in manchen Szenarios aber nicht zur optimalen Lösung. Wir können das jetzt verändern indem wir das Verhältnis von Wert zu Gewicht betrachten.

7 Compilers

- (1) Um einen Syntax Baum aus einer Grammatik zu erstellen müssen wir zuerst einen Scanner schreiben, der den Input in Symbole verwandelt. Danach bauen wir einen Parser, der für die gegebene Grammatik kontrolliert ob die Sequenz von Symbolen, die uns unser Scanner gibt, den Regeln der Grammatik entspricht.

Daraus bauen wir jetzt einen Syntax Tree, indem wir, falls sich eine Operation auf 2 oder mehr Operanden bezieht, diese Operation in die Wurzel schreiben und die anderen in deren Blätter. Falls wir nur einen Operanden haben schreiben wir diesen einfach in das linke Kind des Operationsknoten. Somit bauen wir sukzessiv einen AST auf. Schlussendlich können wir den Syntax Tree noch optimieren.

- (2) Der Scanner liest den gegebenen Input lediglich und ordnet Zeichenketten daraus ihre "Klasse" (Symbole) zu. Jede Zeichenkette eines Types bekommt einen Identifier (Wert) und eine Beschreibung (Klasse).
- (3) Für die Regel

$$d = \{ 'a' \} ['b'] 'c'$$

ist der Parser folgendermaßen aufgebaut.

Am Anfang haben wir mindestens ein "a". Also checken wir ob wir ein "a" finden, und wenn das der Fall ist dann lesen wir mittels einer while Schleife solange bis wir einen anderen Buchstaben finden. Finden wir kein "a" am Anfang so werfen wir einen Fehler.

Danach schauen wir ob wir ein "b" finden. Entweder wir haben genau ein "b" oder keins. Anschließend kontrollieren wir, dass der nächste Buchstabe ein "c" ist und dann nichts mehr kommt. Ist eine der Bedingungen verletzt so werfen wir einen Fehler.

- (4) Wir können Teilbäume vereinfachen. Zum Beispiel falls eine Operation in der Wurzel steht können wir diese auf ihre Kinder anwenden und die Operation mit dem Ergebnis ersetzen.

Weiterhin können wir gleiche Teilbäume mit einer Variablen kodieren und durch diese ersetzen.

Außerdem können wir für komplexere Probleme Verzweigungen vorhersagen und gegebenenfalls eliminieren. Falls wir beispielsweise eine Multiplikation haben und in unserem ersten Knoten eine 0 steht, so wissen wir sofort, dass das Ergebnis auch 0 ist.