Algorithmen & Datenstrukturen 2 Theoriefragen, Jänner 2011
von Moritz Resl & Markus Petschnigg, kein Anspruch auf Vollständigkeit und Richtigkeit

Textsuche

Textsuche: Beschreiben Sie die Problemstellung der Approximativen Zeichenkettensuche. Wie können Sie das naive Textsuche-Verfahren modifizieren, damit es auf dieses Problem anwendbar ist?
Bei der approximativen Zeichenkettensuche beschränkt sich die Suche auf ungefähr passende Textstellen. Um diesen Mechanismus bei der naiven Textsuche zu verwenden könnte zum Beispiel mittels Mismatch-Counter realisiert werden, der bei jeden Mismatch hochzählt. Sobald dieser eine Schwelle überschreitet, passt der untersuchte Text nicht mehr.

Simulated Annealing: Beschreiben Sie die Grundidee des Metropolis-Kriteriums. (Sie müssen die genaue Formel für die Akzeptanzwahrscheinlichkeit nicht angeben.)
Simulated Annealing erweitert die lokale Suche, indem auch Verschlechterungen akzeptiert werden. Dadurch wird ein Entkommen aus lokalen Optima ermöglicht. Die Akzeptanzbedingung für schlechtere Lösungen nennt man Metropolis-Kriterium.

Signaturverfahren von Karp und Rabin: Welche Bedeutung hat es, wenn der Hash-Wert des Musters h(P1,P2,...,PM) mit dem Textblocks h(Ti+1, Ti+2, ..., Ti+M) übereinstimmt?
Dann könnte ein Match aufgetreten sein. Weitere genau Überprüfung des Textblocks notwendig.

Welche Bedeutung hat es, wenn der Hash-Wert des Musters h(P1,P2,...,PM) mit dem Textblocks h(Ti +1, Ti+2,..., Ti+M) nicht bereinstimmt?
Dann liegt sicher kein Match vor.

Welchen der oben angeführten Algorithmen (KMP & BM) würden Sie bevorzugen, wenn der Text nur sequentiell durchsucht werden kann? Begründen Sie Ihre Antwort!
KMP durchläuft den Text ausschließlich sequentiell , da der Index niemals zurückgesetzt wird. Daher ist dieser auch hervorragend geeignet für externe Medien , die nur sequentiell in einer Richtung durchlaufen werden können.

Geben Sie die Laufzeit für den Knuth-Morris-Pratt Algorithmus und für die InitNext Prozedur möglichst genau in Abhängigkeit geeigneter Kenngrößen in Θ-Notation an.
KMP: $\Theta(N+M)$ InitNext: $\Theta(N)$ N: Textlänge M: Musterlänge

Traveling Salesman Problem (TSP)

Das symmetrische Traveling Salesman Problem: Angenommen die Distanzwerte zwischen den Knoten im TSP sind paarweise verschieden, ist die mittels der Spanning-Tree-Heuristik (gilt auch für Christophides-Heuristik) erzeugte Lösung immer eindeutig? Begründen Sie Ihre Antwort!
Nein, Lösung ist nicht immer eindeutig. Abhängig von Startknoten gibt es unterschieldiche Lösungen.

Welche Bedingung muss für das euklidische TSP im Vergleich zum allgemeinen TSP erfüllt sein?
Die Dreiecksungleichung muss erfüllt sein.

Was ist ein Polynomial Time Approximation Scheme (PTAS)?
Eine Familie der Approximationsalgorithmen, bei denen die Gütegarantie e > 1 mit der Laufzeit O (N^(1/e-1)) approximiert werden kann,

Branch-and-Bound

Branch-and-Bound: Wann bzw. unter welchen Voraussetzungen kann beim Branch-and-Bound Algorithmus ein Teilproblem aus der Problemliste entfernt werden?
Ist für eine (oder mehrere) Teilmengen die für sie berechnete untere Schranke nicht kleiner als die beste überhaupt bisher gefundene obere Schranke, braucht man die Lösungen in dieser Teilmenge nicht mehr betrachten.

Algorithmen

Cache optimale Algorithmen: Was bedeutet "Zeitliche Lokalität"?
Wenn öfters innerhalb einer kurzen Zeitspanne auf einen Speicherbereich zugegriffen wird, spricht man von zeitlicher Lokalität. Caches nutzen diese aus, indem sie versuchen, angesprochene Speicherbereiche möglichst lange im Chache zu lassen.

Cache optimale Algorithmen: Was bedeutet "Örtliche Lokalität"?
Existiert, wenn öfter Speicherbereiche angesprochen werden, die beieinander liegen. Caches nutzen diese aus, indem sie bei jedem Miss jeweils ganze Speicherblöcke in den Cache laden.

Cache optimale Algorithmen: Wann heißt ein Algorithmus "Cache-Aware" (Cache-bewusst)?
Ein Algorithmus heißt "Cache-bewusst" wenn er Parameter enthält, mit denen die Cache-Komplexität für die gegebene Cachegröße und Zeilenlänge optimiert werden kann. Andernfalls heißt er "Cache-oblivious" (Cach-ignorierend).

Evolutionäre Algorithmen: Beschreiben Sie die Grundidee der Fitness-proportionalen Selektion.
Die Selektion kopiert aus der aktuellen Population Lösungen, die in den weiteren Schritten neue Nachkommen produzieren werden. Dabei werden grundsätzlich bessere Lösungen mit höherer Wahrscheinlichkeit gewählt. (Schlechtere können aber auch gewählt werden, um lokalen Optima zu entkommen)
Bei der Fitness-proportionalen Selektion sollen gute Lösungen bevorzugt werden.

Was sind die Aufgaben der Rekombination und Mutation?
Rekombination: aus 2 selektierten Elternlösungen eine neue Lösung generieren (neue Lösung soll möglichst ausschließlich aus Merkmalen der Eltern aufgebaut sein: starke Vererbung)
Mutation: Move zur Nachbarlösung, um neue oder verlorengegangene Merkmale in die Population einzubringen

Skiplisten

Randomisierte Skiplisten: Es wird ein neues Element in eine randomisierte Skipliste eingefügt. Wie groß ist die Wahrscheinlichkeit, dass es die Höhe 1 annimmt? (Die Münzwurf-Methode aus der Vorlesung bzw. aus dem Skriptum wird verwendet.)
Zuerst setzt man die Höhe auf 0. Dann wirft man eine Münze und vergrößert die Höhe um 1, falls "Zahl" geworfen wird. Dies tut man solange, bis zum ersten Mal "Kopf" erscheint. Daraus ergibt sich die folgende Wahrscheinlichkeit, dass die Höhe des neuen Containers gleich i (hier: 1) ist:
prob(c.höhe = 1) = 2^(-1-1)

Gegeben sei die Entfernen()-Funktion für eine Randomisierte Skipliste. Beschreiben Sie grob, wie diese modifiziert werden muss, damit sie auf Perfekte Skiplisten anwendbar ist.
Bei einer Perfekten Skipliste kann es bei der Entfernen()-Funktion zu einem sehr großen Aufwand kommen, da unter Umständen eine komplette Reorganisation der Liste vonnöten ist. (Schlagwort Conatinerhöhen!)

Gegeben sei die Einfügen()-Funktion für eine Randomisierte Skipliste. Beschreiben Sie grob, wie diese modifiziert werden muss, damit sie auf Perfekte Skiplisten anwendbar ist.
Bei perfekten Skip-Listen bedeutet das Einfügen oder Entfernen von Listenelementen eine vollständige Rekonfiguration der Liste (Container Höhen)

Gegeben sei die Suchen()-Funktion für eine Randomisierte Skipliste. Beschreiben Sie grob, wie diese modifiziert werden muss, damit sie auf Perfekte Skiplisten anwendbar ist.
Eine Perfekte Skipliste und eine Randomisierte Skipliste verwenden den gleichen Suchalgorithmus.

Welche Vor- und/oder Nachteile besitzt die randomisierte Skipliste gegenüber der perfekten Skipliste?
Geringerer Aufwand beim Einfügen oder Entfernen bei randomisierten Skiplisten.

Miller-Rabin Primzahlentest

Randomisierter Primzahltest von Miller-Rabin: Welche Aussage hat das Ergebnis, wenn a für "n ist prim" zeugt?
Hierbei liegt ziemlich sicher eine Primzahl vor, da es sich jedoch um ein Monto-Carlo Verfahren handelt, ist das Ergebnis mit einer gewissen Fehlerwahrscheinlichkeit behaftet.

Welche Aussage hat das Ergebnis, wenn a für "n ist nicht prim" zeugt?
Hierbei kann man sicher sein, dass keine Primzahl vorliegt.

Algorithmus von Miller-Rabin: In der Zeuge()-Funktion muss der Term a^(n-1) ≡ 1 mod n möglichst effizient ausgewertet werden. Beschreiben Sie, wie dieses Problem gelöst wird.
Das Problem wird durch so genanntes "Repeated Sqauring" gelöst, bei der die Vertauschbarkeit von Multiplikation und Modulo-Berechnung ausgenutzt wird.

Geben Sie die (Bit-)Komplexität der Zeuge()-Funktion möglichst genau in Abhängigkeit geeigneter Kenngrößen in O-Notation an. Beschreiben Sie, wofür die verwendeten Variablen stehen.
O(s*k^3) (s = Aufrufe der Zeuge()-Funktion, k = Zahl besteht aus k Bits (Länge binär))

Prioritätswarteschlangen

Prioritätswarteschlangen: Geben Sie den Aufwand für folgende Operationen in Abhängigkeit der Anzahl der Elemente n in Θ-Notation im Worst-Case an, wenn für die Realisierung der Prioritätswarteschlange ein binärer Heap verwendet wird:
Get Min: Θ(n), Del Min: Θ(n), Insert: Θ(n)

Schnitt von allgemeinen Liniensegmenten

Schnitt von allgemeinen Liniensegmenten: Was ist eine Scan-Line-Status Struktur? Was wird darin wie gespeichert? Womit wird sie initialisiert?
In der Scan-Line-Status Struktur werden die jeweils aktiven Kanten gespeichert. Sie wird leer initialisiert.

Scanline-Algorithmus: Geben Sie den Speicherplatzbedarf für den Algorithmus für das Segment-Schnitt Problem in allgemeiner Lage möglichst genau in Abhängigkeit geeigneter Kenngrößen in O-Notation an.
Der Speicherplatzbedarf ist im schlimmsten Fall quadratisch, weil die Kreuzungen in der Ereignisstruktur abgelegt werden und es quadratisch viele Kreuzungen geben kann.
O(n^2)

Schnitt von allgemeinen Liniensegmenten: Was ist eine Ereignisstruktur? Was wird darin wie gespeichert? Womit wird sie initialisiert?
Die Ereignisstruktur wird mit allen Anfangs- bzw. Endpunkten der Linien aufsteigend nach x-Koordinate initialisiert. Weiters werden beim Auftreten auch noch die Schnittpunkte eingefügt.

Scanline-Algorithmus: Geben Sie die Laufzeit für den Algorithmus für das Segment-Schnitt Problem in allgemeiner Lage möglichst genau in Abhängigkeit geeigneter Kenngrößen in O- Notation an. Beschreiben Sie, wofür die verwendeten Variablen stehen.
Laufzeit beläuft sich auf O((n+k)*log n) (n = Anzahl d. Segmente, k = Anzahl d. Schnittpunkte)
Speicherplatzbedarf ist O(n^2)

Scan-line: Welche vereinfachende Annahmen sind notwendig, damit der Algorithmus zum Schnitt von allgemeinen Liniensegmenten reibungslos abläuft? Nennen Sie 3 davon!
1. Der Schnitt zweier Segmente ist stets leer oder genau ein Punkt
2. Es schneiden sich nie mehr als zwei Segmente in einem Punkt
3. Alle Endpunkte und Schnittpunkte haben paarweise verschiedene x-Koordinaten

Approximative Algorithmen

Approximative Algorithmen: Was bedeutet es, wenn ein Algorithmus für ein Maximierungsproblem Gütegrantie ε besitzt? Was bedeutet es, wenn ein Algorithmus für ein Minimierungsproblem Gütegarantie ε besitzt?
Die Güte sagt etwas darüber aus, ob optimale Lösungen gut oder schlecht angenähert werden. Gütegarantie e=1 wäre, dass der Algorithmus die optimale Lösung liefert. In der Praxis gibt es so etwas bei NP-schwierigen Problemen jedoch nicht, daher werden Lösungen angenähert.

Approximative Algorithmen: Welche Gütegarantie besitzt die Spanning-Tree-Heuristik für das Euklidische Traveling Salesman Problem? Ca(P) / Copt(P) <= E
Welche Gütegarantie besitzt die Christophides-Heuristik für das Euklidische Traveling Salesman Problem? Cch(P) / Copt(P) <= 3/2 (Forum sagt 14% Abweichung von optimalen Lösung)

Randomisierte Algorithmen

Randomisierte Algorithmen: Verbessert sich beim Randomisierten Quicksort, im Vergleich zum klassischen Quicksort, das Laufzeitverhalten? Wenn ja, inwiefern?
Nein, die Laufzeit bleibt erwartungsgemäß bei O(n log n). Das Pivotelement (nach dem sortiert wird) wird zufällig gewählt.

Randomisierte Algorithmen: Erläutern Sie in wenigen Worten den Unterschied zwischen Las-Vegas-Verfahren und Monte-Carlo-Verfahren! Geben Sie weiteres zu jedem der beiden Klassen einen Algorithmus als Beispiel an.
Las Vegas: Immer richtige Lösung. (zB Quicksort)
Monte Carlo Verfahren: mit gewisser Fehlerwahrscheinlichkeit behaftet. (zB. Miller Rabin Primzahltest)

Tries

Erläutern Sie in wenigen Worten das Prinzip des Linked Tries. Welche Vor- und/oder Nachteile besitzt der Linked Trie gegenüber dem Indexed Trie?
Beim Linked Trie wird das Feld für jeden Knoten durch eine lineare Liste ersetzt. Leere Einträge werden nicht gespeichert. Bei der Suche kann der Aufwand deutlich höher sein als beim Indexed Trie.

Tries: Wieviele Datensätze kann ein Radix Trie mit Höhe m maximal aufnehmen?
2^m

Geben Sie die Laufzeit für das Suchen nach einem Datensatz in einem Radix Trie möglichst genau in Abhängigkeit geeigneter Kenngrößen in O-Notation an.
O(m)

Können in einem Packed Trie zwei unterschiedliche Knoten denselben Anfangsindex erhalten?
Nein, da sonst nicht zwischen ihren Elementen unterschieden werden kann. So würden Daten verloren gehen.

Vorgehensweise zu Packed Tries (Greedy Heuristik):
Absteigend nach Belegung sortieren und der Größe nach absteigend einfügen. Bei gleich großen kann Auswahl zufällig gewählt werden. Dies ändert nichts mehr an der Dichte des Packed Tries. Ebenso sind auch negative Indizes möglich.

Bin Packing

Beschreiben die First-Fit-Heuristik für das Bin-Packing Problem.
Gegenstände werden in beliebiger Reihenfolge behandelt. Jeder Gegenstand wird in die erstmögliche Kiste gelegt, in die er passt.

Tabu-Suche

Tabu-Suche: Die Wahl der Länge einer Tabu Liste ist häufig kritisch. Was können die Probleme bei zu kurzer bzw. zu langer Tabu-Liste sein?
Die Tabu-Suche ist eine Erweiterung der lokalen Suche und verwendet ein Gedächtnis, um bereits bekannte Lösungen nicht gleich nochmal zu untersuchen. Lange Tabu-Liste: Viele Lösungen, Wahrscheinlichkeit gering zur Neuberechnung. Zeitaufwand und Speicherbedarf dafür groß.

Lokale Suche

Nachbarschaft: umfasst die Menge der Lösungen, die von einer aktuellen Lösung durch eine bestimmte Art von Änderung (*Zug*) erreicht werden können.
Alle Lösungen nennt man Nachbarschaftsstruktur, die man auch als Graph darstellen kann.

Beispiel: Rucksackproblem: Nachbarschaft *N(x)* könnte die Menge aller Bitvektoren sein, die sich in genau einem Bit unterscheiden.

Allgemein kann eine sinnvolle Nachbarschaft für ein Problem durch Austausch definiert werden: Entferne aus der aktuellen Lösung bis zu *r* Gegenstände.

Strategien zur Auswahl einer neuen Lösung:
Best improvement: Gesamte Nachbarschaft *N(x)* wird vollständig durchsucht und die beste Lösung *x'* wird ausgewählt. Dieser Ansatz ist am zeitaufwendigsten.

First improvement: Nachbarschaft *N(x)* wird durchsucht, bis eine neue Lösung gefunden ist, die besser ist als die alte.

Random neighbor: Eine neue Lösung wird zufällig bestimmt. Die Lösung kann sich dadurch auch verschlechtern.