Algodat 2 Theoriefragen

Suchen in Texten

Laufzeiten:

Naive Suche immer Theta von N*M wenn N = Textlänge und M = Musterlänge

Boyer-Moore = Theta von N+M (wenn beide Signaturverfahren angewendet werden)

Knuth-Morris-Pratt dasselbe

1) **Welchen der oben angefuhrten Algorithmen w urden Sie bevorzugen, wenn der Text nur sequentiell durchsucht werden kann? Begrunden Sie Ihre Antwort!**

Knuth-Morris-Pratt, da der Index i niemals zurückgesetzt wird und das besonders bei externen Speichermedien von Vorteil ist!

2) **Welchen aus der Vorlesung und Skriptum bekannten Algorithmus zur Textsuche wurden Sie bevorzugen, wenn die Zeichen des Musters P nur selten in T vorkommen? Begrunden Sie Ihre Antwort!**

Signaturverfahren von Karp und Rabin, da damit nicht von Anfang bis Ende im Text gesucht wird sondern ein Hashwert des Musters erstellt wird und nachgeschaut wird ob dieser irgendwo in Teilen des Textes vorkommt.

3) **Signaturverfahren von Karp und Rabin:**
   Welche Bedeutung hat es, wenn
   der Hash-Wert des Musters h(P1; P2; : : : ; PM) mit dem Hash-Wert eines Textblocks h(Ti+1; Ti+2; : : : ; Ti+M) ubereinstimmt?
   der Hash-Wert des Musters h(P1; P2; : : : ; PM) mit dem Hash-Wert eines Textblocks h(Ti+1; Ti+2; : : : ; Ti+M) nicht ubereinstimmt?

Übereinstimmung: Muster kommt im Text vor

Keine Übereinstimmung: Muster kommt im Text nicht vor

Randomisierte Algorithmen

1) **Gegeben sei die Suchen()-Funktion f ur eine Randomisierte Skipliste.**

   Beschreiben Sie grob, wie diese modiziert werden muss, damit sie auf Per-

   fekte Skiplisten anwendbar ist.

Muss nicht geändert werden – FANGFRAGE!!!

2) **Gegeben sei die Entfernen()-Funktion f ur eine Randomisierte Skipliste.**

   Beschreiben Sie grob, wie diese modiziert werden muss, damit sie auf Per-

   fekte Skiplisten anwendbar ist.

nach dem Entfernen muss gecheckt werden ob die Eigenschaften für randomisierte Skiplisten noch gegeben sind!

3) **Beschreiben Sie in Stichworten den gesamten Loschvorgang eines beliebigen Elements aus einer perfekten Skipliste.**

Suche nach dem Element, Aushängen der Container und umhängen mit Hilfe der Update-Struktur, Höhen aktualisieren, Eigenschaften der perfekten Skipliste überprüfen und erneuern, Element löschen.

4) **Wieviele Zeiger werden in Theta-Notation in Abhangigkeit der Anzahl der Elemente n bei einer perfekten Skipliste im Vergleich zu einer einfach verkettetenListe zusatzlich benötigt?**

Theta(n) Zeiger

5) **Wieviele Container werden im Worst-Case in Theta-Notation in Abhangigkeit der Anzahl der Elemente n bei einer perfekten Skipliste bei einer**
   erfolgreichen und
   erfolglosen Suche
   betrachtet?

In beiden Fällen werden Container mit Anzahl Theta(log n) benötigt.

6) **Randomisierte Skiplisten:**
   Es wird ein neues Element in eine randomisierte Skipliste eingefugt. Wie groß ist die Wahrscheinlichkeit, dass es die Hohe 3 annimmt? (Die M unzwurf-Methode aus der Vorlesung bzw. aus dem Skriptum wird verwendet.)

$$prob(c.höhe = i) = \frac{1}{2^{i+1}}$$

7) **Randomisierte Algorithmen:**
   Erl autern Sie in wenigen Worten den Unterschied zwischen Las-Vegas-Verfahren und Monte-Carlo-Verfahren. Geben Sie weiters zu jedem der beiden Klassen einen Algorithmus als Beispiel an (nur den Namen).

Las-Vegas: Verfahren bringt immer eine richtige Lösung (z.B. Random-Quicksort)

Monte-Carlo: haben eine gewisse Fehlerwahrscheinlichkeit (z.B. Miller-Rabin-Primzahlentest)

8) **Welchen Vor- und Nachteil haben perfekte Skiplisten im Vergleich zu einfach verketteten Listen ?**

Vorteil: Container mit verschiedenen Höhen, dadurch können Einträge auch Zeiger auf weiter hinten liegende Einträge haben. Dadurch ergibt sich logarithmische Laufzeit.

Nachteil: um nach dem Einfügen und Löschen weiterhin logarithmische Laufzeit zu erzielen, also die Höhen der Container zu berichtigen, benötigt man zusätzlich Reorganisation mit Aufwand O(n).

9) **Wie unterscheidet sich das Randomisierte Quicksort vom klassischen Quicksort? Verbessert ich dadurch das Laufzeitverhalten? Wenn ja, inwiefern? Verbessert sich dadurch der Speicherverbrauch? Wenn ja, inwiefern?**

Pivotelement wird zufällig aus dem Bereich der zu sortierenden Folge gewählt. Laufzeit hat einen Erwartungswert von O(n log n).

10) **Randomisierter Primzahltest von Miller-Rabin:**
Welche Aussage hat das Ergebnis, wenn a für „n ist prim" zeugt?
Welche Aussage hat das Ergebnis, wenn a für „n ist nicht prim" zeugt?

N ist prim: Wir haben keinen Zeugen für die Zusammengesetztheit von n gefunden, wir müssen eine weitere Basis wählen um zu garantieren, dass n wirklich prim ist.

N ist nicht prim: Wir haben einen Zeugen gefunden und n ist sicherlich nicht prim!

Geometrische Algorithmen

1) **Was ist eine Ereignisstruktur? Was wird darin gespeichert? Womit wird sie initialisiert?**

ES wird anfangs mit allen möglichen Ereignissen, die auftreten können aufgefüllt (Segmentanfang, Schnitt von Segmenten, Segmentende), geordnet nach der x-Achse.

2) **Was ist eine Scanline-Status-Struktur? Was wird darin gespeichert und womit wird sie initialisiert?**

Sie ist anfangs leer und enthält während des Algorithmus nur die aktiven Segmente, sprich keine toten Segmente. Geordnet nach der y-Achse

3) **Geben Sie die Laufzeit fur die eindimensionale Bereichssuche, nachdem die Vor-verarbeitung schon durchgefuhrt ist, m oglichst genau in Abh angigkeit geeigneter Kenngroen in O-Notation an. Beschreiben Sie, wofur die verwendeten Variablen stehen.**

Bereichssuche hat Laufzeit von O(R + log n), wobei R die Anzahl der Punkte ist, die tatsächlich im Bereich liegen und n die Anzahl der Punkte die im gesamten Raum liegen.

4) **Geben Sie den Speicherplatzbedarf im Worst-Case fur den Algorithmus fur das Segment-Schnitt Problem in allgemeiner Lage in Abhangigkeit der Anzahl der Segmente n in Theta-Notation an, wenn fur die Anzahl der Schnittpunkte k gilt:**
k = O(n).

Theta(n^2)

Algorithmen für große Datenmengen

1) **Cache-optimale Algorithmen:**
Beschreiben Sie in einfachen Worten, was unter dem Begriff „örtliche Lokalität" verstanden wird. Wie kann diese vom Cache ausgenutzt werden?

… existiert, wenn öfter Speicherbereiche angesprochen werden, die nahe beieinander liegen. Caches nutzen diese aus indem sie bei jedem MISS ganze Speicherblöcke in den Cache holen.

2) **Cache-optimale Algorithmen:**
Beschreiben Sie in einfachen Worten, was unter dem Begriff „zeitliche Lokalität" verstanden wird. Wie kann diese vom Cache ausgenutzt werden?

… existiert, wenn der gleiche Speicherbereich öfter innerhalb eines kurzen Zeitraums angesprochen wird. Wird ausgenutzt indem versucht wird, die betroffenen Speicherbereiche möglichst lange im Cache zu belassen.

Tries

1) **Welche Vor- und/oder Nachteile besitzt der Indexed Trie gegen uber dem Linked Trie?**

Bei größeren Tries, die recht dicht sind kann durch das Iterieren in den Listen ein deutlich höherer Aufwand von O(m*|A|) -> A ist Alphabet entstehen.

2) **Welchen wesentlichen Vorteil bietet der Packed-Trie gegenüber dem Indexed-Trie?**

Indexed-Trie enthält Felder, die entweder NULL sind oder False, diese werden nicht gebraucht, auf diese Optimierung geht der Packed-Trie ein.

3) **Wieviele Datens atze kann ein Radix Trie mit H ohe m maximal aufnehmen?**

2^m

4) **Tries:**
   Erlautern Sie in wenigen Worten das Prinzip des Linked Tries

Feld für jeden Knoten aus dem Indexed-Trie wird durch eine lineare Liste ersetzt, wodurch nur next- und end-Variablen enthalten sind, für die gilt dass next != NULL oder end != False. Alle leeren Einträge werden damit nicht mehr gespeichert.

Optimierungsalgorithmen

1) **Evolution are Algorithmen:**
   Was versteht man unter dem Selektionsdruck? Welche Probleme k onnen bei einem zu geringen bzw. zu hohem Selektionsdruck auftreten?

Selektionsdruck ist der Faktor, um den die beste Lösung häufiger selektiert wird als eine durchschnittliche. Wenn zu niedrig, werden gute Lösungen wenig bevorzugt und es wird zu einer Zufallssache, wenn zu hoch dann werden sie zu stark bevorzugt und der Algorithmus wird zu einem lokalen Optimum.

2) **Welche G utegarantie besitzt die Christophides-Heuristik f ur das Euklidische Traveling-Salesman-Problem? Beschreiben Sie die Bedeutung dieses Faktors auch in Worten.**

$$\frac{c_{CH}(P)}{c_{opt}(P)} \leq \frac{3}{2} \; \forall \, P \in \pi$$

Christophides ist ein Minimierungsalgorithmus, daher er ist ein 3/2-approximativer Algorithmus. Diese Zahl sagt aus wie gut die Heuristik ein Problem löst.

3) **Branch-and-Bound fur das asymmetrische TSP:**
   Aus welchen Kanten besteht die Kantenmenge eines Nulldigraphen?
   Wofur wird der Nulldigraph im Branch-and-Bound verwendet?

Der Nulldigraph soll (bei einer gültigen Lösung) eine Tour enthalten, ist somit eine Heuristik um eine Tour zu finden.

4) **Tabu-Suche:**
   Was ist eine Tabu-Liste? Die Wahl der Lange ist häufig kritisch. Was können Probleme bei zu kurzer bzw. zu langer Tabu-Liste sein?

Wenn zu kurz, dann werden zu viele nicht optimale Lösungen im Algorithmus weiterhin betrachtet

Wenn zu lang, dann wird die Suche nach einer Lösung zu stark eingeschränkt

5) **Symmetrisches TSP:**
   Analysieren Sie die 3-OPT Nachbarschaft fur das symmetrisches TSP auf einem vollstandigen Graphen mit N Knoten: Wieviele unterschiedliche Moglichkeiten kann es maximal geben, um nach dem Entfernen von entsprechend vielen Kanten die entstandenen Teilpfade wieder zu einer gultigen Tour zu verbinden?

Es gibt dabei 4 3-opt-Schritte und 3 2-opt-Schritte, somit insgesamt 7 Möglichkeiten.

6) **Evolutionare Algorithmen:**
Beschreiben Sie die Grundidee der Tournament Selektion. Wie wird darin der Selektionsdruck gesteuert?

a) Wähle aus der Population k Lösungen gleichverteilt zufällig aus, Mehrfachauswahl ist erlaubt
b) Die beste der k Lösungen ist die selektierte

Der Selektionsdruck kann über die Gruppengröße k gesteuert werden

7) **Approximative Algorithmen:**
Was bedeutet es, wenn ein Algorithmus A fur ein Maximierungsproblem Gutegarantie Epsilon besitzt?
Welche Gutegarantie besitzt die Spanning-Tree-Heuristik fur das Euklidische Traveling Salesman Problem?

A ist ein Epsilon-approximativer Algorithmus, dabei ist Epsilon immer > 0 und es gilt:

$$\frac{c_A(P)}{c_{opt}(P)} \geq \varepsilon \, \forall \, P \in \pi$$

Für ST-Heuristik gilt: $\frac{c_A(P)}{c_{opt}(P)} \leq 2 \, \forall \, P \in \pi$

8) **Approximative Algorithmen:**
Was bedeutet es, wenn ein Algorithmus A fur ein Minimierungsproblem Gutegarantie " besitzt?
Welche Gutegarantie besitzt die Christophides-Heuristik fur das Euklidische Traveling Salesman Problem?

A ist ein Epsilon-approximativer Algorithmus, Epsilon größer 0 und es gilt:

$$\frac{c_A(P)}{c_{opt}(P)} \leq \varepsilon \, \forall \, P \in \pi$$

Gütegarantie von Christophides-Heuristik, siehe oben!

9) **Branch-and-Bound:**
Beschreiben Sie mindestens zwei Situationen, in denen beim Branch-and-Bound Algorithmus ein Teilproblem aus der Problemliste entfernt werden kann.

Lokale Schranke ist größer als obere Schranke

Neue gefundene untere Schranke ist größer als obere Schranke

10) **Simulated Annealing:**
Beschreiben Sie die Grundidee des Metropolis-Kriteriums. (Sie mussen die genaue Formel fur die Akzeptanzwahrscheinlichkeit nicht angeben.)

Ausgangslösung mit Konstruktionsheuristik, die beliebig gewählt wird. Ausgehend von Lösung x wird eine Nachbarlösung x‘ durch eine kleine zufällige Änderung abgeleitet, ist diese besser als x, so wird sie als aktuelle Lösung akzeptiert. Ansonsten wird diese mit einer gewissen Wahrscheinlichkeit (Formel) akzeptiert. Damit kann man lokalen Optima gut entkommen. In der Formel kommt die Variable T vor, die man als „Temperatur“ bezeichnet.

11) **Lokale Suche:**
Fur die Auswahl einer Nachbarschaftsl osung k onnen 3 verschiedene Strategien verfolgt werden. Erlautern Sie kurz die Vorgangsweise der folgenden Verfahren:

Best improvement: Nachbarschaft N(x) wird vollständig durchsucht und die beste Lösung x‘ wird ausgewählt (ist am zeitaufwändigsten)

First improvement: Nachbarschaft wird systematisch durchsucht, bis die erste bessere Lösung gefunden wird bzw. alle Nachbarn betrachtet wurden.

Random neighbour: Nachbarlösung wird zufällig abgeleitet (ist die schnellste Variante, jedoch häufig, dass x‘ schlechter als x ist)