2008 年 2 月 1 日

データ構造とアルゴリズム　期末試験問題

1. 比較に基づくソーティングの時間計算量が $\Omega(n \log n)$ であることを説明せよ．ただし n はソートするデータ数とする．なお Stirling の公式 $n! > (\frac{n}{e})^n$ は適宜用いよ．(15 点)

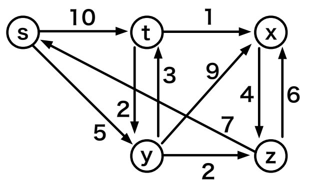
   クイックソートの実行に要する時間を $T(n)$ で表す．ただし n はソート対象の要素数とする．以下の問に答えよ．

   (a) 再帰呼び出しにおける要素集合の分割が常に 1 対 1（等分割）となる場合を考える．$T(n)$ の漸化式は（分割の際の端数を無視すれば）$T(n) = 2T(n/2) + O(n)$ と表せる．オーダー表記による $T(n)$ の上界を導出せよ．(5 点)

   (b) 再帰呼び出しにおける要素集合の分割が常に 1 対 $K-1$（ただし K は 2 以上の定数）となる場合を考える．$T(n)$ の漸化式は（分割の際の端数を無視すれば）$T(n) = T(n/K) + T(n(K-1)/K) + O(n)$ と表せる．オーダー表記による $T(n)$ の上界を導出せよ．(5 点)

   (c) クイックソートの実行時間が最悪になる場合はどのようなときかを述べ (証明は不要)，その場合でのオーダー表記による $T(n)$ の上界を導出せよ．(5 点)

2. 以下の図で示されるグラフにおいて，ダイクストラのアルゴリズムを用いて s から各節点への最短経路を求めよ．なお，計算の過程が分かるように解答せよ．(20 点)



3. 長さ n のテキスト，長さ m のパターンに対する文字列照合のラビン-カープ法について考える．計算時間を $O(n+m)$ にするには，ハッシュ関数がどのような性質を満たせばよいか，(a) ハッシュ関数の計算時間，および，(b) ハッシュ値の衝突，それぞれについて説明せよ．さらに，ラビン-カープ法では，どのようにしてその性質を保証しようとしているか説明せよ．(20 点)

4. クヌース-モリス-プラッツ法に関して，(a) 表 $next$ について説明し，(b) 文字列 $abcabcd$ に対する表 $next$ の内容を示せ．(20 点)

5. 次のアルゴリズムの戦略について説明し，その戦略を利用しているアルゴリズムの例を 1 つ示せ．(20 点)

   (a) 貪欲法（グリーディ法）

   (b) 分割統治法

   (c) 動的計画法

   (d) 分岐限定法

データ構造とアルゴリズム 期末試験対策

2007 年度後期期末試験 (2008/02/01) 解答例

この解答例は多分に間違えてる可能性があるので注意しながら閲覧すること．問題文は Web ページにアップロードされているのでそちらを参照．間違いを見つければできる限り連絡すること．

$\boxed{1}$ (1)

データ数 n の置換の数は $n!$ である．ここで下界を示すため if 文の回数について考える．if 文は 1 回あたり置換の数を最大で 1/2 にできることから，ソートする（置換の数を 1 にする）には少なくとも $\Omega(\log_2 n!)$ の実行時間が必要である．これはスターリングの公式より $\Omega(\log_2 n!) > \Omega(n \log_2 n)$ である．よってソーティングの時間計算量の下界は $\Omega(n \log n)$ である．

$\boxed{1}$ (2)

(a) 与式を二階微分をした時，$2T''(n) = T''(n/2)$ となることから，$T''(n) = 1/n$ である．これを二度，積分を行うと $T(n) = n \log n$ となり，$T(n)$ の上界は $O(n \log n)$ であることがわかる．

(b) $n = K$ であるので与式を全て n で表現すると，$T(n) = T(1) + T(n-1) + O(n) = 2T(n) + T(n-2) + 2O(n) = \cdots = nT(1) + nO(n)$ であるので，$T(n)$ の上界は $O(n^2)$ であることがわかる．

(c) 要素集合の分割をするときに常にその要素集合の最小または最大を仮の中間値として取ってしまった時に，(b) と同様のことが発生し最悪となる．この時，$T(n)$ の上界は (b) で求めた通り，$O(n^2)$ である．

$\boxed{2}$

s からの距離の小さい順に取り出される優先順位キューを考える．もし，ある距離未満で到達できる範囲の全ての節点がわかっていれば，それらの節点を経由して全ての節点へ行く場合の s からの最小距離を算出した時，ある距離以下では到達できなかった節点の内，最も s からの距離が小さいものの距離は s からの最短距離である．なぜならば，それ以外の点（ある距離以上の点）を経由した場合，全ての辺の重みは正の数であるので，その距離よりも長くなるはずだからである．

まず，s の距離を 0 とし，s から出ている辺について全てを追加する．

$5:y; 10:t$

次に先頭 $(y = 5)$ を決定し，y から出ている辺について全てを追加する．ここで t は現在 10 であるが，y を経由すると 8 になり小さくなるので変更する．

$\{7:z; 8:t; 11:x\}$

次に先頭 $(z = 7)$ を決定し，z から出ている辺について全てを追加する．ここで x への経路もあるが，もともとある経路が小さいので追加しない．同様に s への経路もあるが，s は既に決定しているので追加しない．

$\{8:t; 11:x\}$

次に先頭 $(t = 8)$ を決定し，t から出ている辺について全てを追加する．

$\{9:x\}$

次に先頭 $(x = 9)$ を決定し，x から出ている辺について全てを追加する．ここで優先順位キューの要素数が 0 になったので終了する．

これにより各節点への最短距離は

$\{s:0; y:5; z:7; t:8; x:9\}$

となった．

$\boxed{3}$

(a) ハッシュは初期化時は $O(m)$，繰り返し時は $O(1)$ で計算しなければならない．

(b) ハッシュ値は衝突しないくらい十分に長い必要がある．

$\{a_0, a_1, \cdots, a_{n-1}\}$ のハッシュが与えられた時，$\{a_1, a_2, \cdots, a_n\}$ のハッシュが $O(1)$ で計算できるようにしている．

4

(a) そこで検索した時に戻るべき長さが格納される．

(b)

a	b	c	a	b	c	a
-1	0	0	0	1	2	3

5

(a) 貪欲法（グリーディ法） アルゴリズムの各段階で局所的な最良の答えを求めて，序所に範囲を大きくしていき大局的な解を得る．ダイクストラの最短経路アルゴリズム．

(b) 分割統治法 部分問題に分解して各解を求めていくことで，最終的に全体の解を得る．クイックソート．

(c) 動的計画法 分割問題 $P(i)$ を定義する．$P(0)$ の最適解を求めて，$P(i)$ の最適解を i より小さな部分問題の最適解を元に構成する．ナップザック問題のアルゴリズム．

(d) 分岐限定法 明らかに不要な場合を除き，組み合わせ全てを検査することで解を得る．MAX-SAT に対するアルゴリズム．