

問2 リスト構造で管理されているセルとガーベジコレクタ (ソフトウェア) (H28 春・FE 午後問2)

【解答】

【設問1】 aーア, bーウ, cーオ, dーウ

【設問2】 イ

【解説】

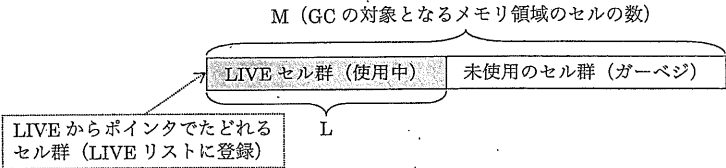
リスト構造で管理されているセルとガーベジコレクタに関する問題である。
アプリケーションソフトを実行するために確保されたメモリ領域が、アプリケーションの終了後も確保されたままになっていると、他のアプリケーションソフトが使用できるメモリ領域が少なくなってしまう。そこで、そのようなメモリ領域を自動的に解放して再利用できるようにする機能をガーベジコレクション (garbage collection ; ゴミ集め) という。アプリケーションが利用していないにもかかわらず、解放されず再利用が不可能なメモリ領域が発生することをメモリリークというが、OS や常用しているアプリケーションでメモリリークが発生すると、メモリ不足となり、コンピュータを正常に使用できなくなることがある。しかし、ガーベジコレクションの機能があると、そのような事態を未然に防ぐことができる。このガーベジコレクションを実行するためのモジュールがガーベジコレクタ (以下、GC という) である。

ガーベジコレクション、メモリリークは試験対策用に多用される用語であり、やや上級試験向けである。しかし、問題を解く上で、特別な知識は必要としない。この問題では、プログラムが要求するデータ領域はセルという単位で管理され、リスト構造をもつが、リスト構造の具体的な内容にも言及していない。したがって、問題文から、解答に必要な内容を読み取れば解ける。

ポイントは、設問1の内容を理解できるかどうかであり、設問2はマークビットが1のセルと0のセルの意味と区別ができればよい。

【設問1】

マークアンドスイープ方式によるGC処理の説明を基に、考察する。
・空欄a: 設問文の記述から、GCの対象となるメモリ領域のセルの数MとGC開始時のLIVEセルの数Lの関係を図Aに示す。



図A MとLの関係

図Aで示したように、GCの対象となるメモリ領域のセルの数をMとした場合、プログラムが使用しているLIVEセルの数がLであれば、M-Lが未使用のセル (ガーベジ) の数となる。つまり、「M=LIVEセルの数+未使用のセル (ガーベジ) の数」が成立する。

【マークアンドスイープ方式によるGC処理の説明】に、マーキングは「LIVEからのポインタをたどって到達できる全てのLIVEセルのマークビットを1にする」処理とあり、その処理量をセル数で考えるとLとなる。Lが小さければ処理量も小さいが、Lが大きくなれば処理量も大きくなる。したがって、マーキングの処理量はLに比例することが分かる。正解は、「L」の (ア) である。

・空欄b: 図4から、マーキング終了後は、マークビットが1であればLIVEセル、0であればガーベジのセルであることが分かる。

また、スイープの処理内容は、【マークアンドスイープ方式によるGC処理の説明】(2)の記述から次のようになっている。

- i セルが格納されているメモリ領域中の全てのセル (M) を重複がないように走査し、各セルのマークビットの値を調べる。
- ii マークビットが0ならば、そのセルはガーベジであるので、供給源に未使用セルとして返却する。
- iii マークビットが1であれば0にする。
 - iから、処理を行うセル数はMであり、マークビットの値によってii、又はiiiどちらかの処理を行う。iiとiiiで処理の違いはあるが、全体のセルの数 (M) を処理するため、Mが小さければスイープの処理量も小さく、Mが大きければスイープの処理量も大きくなるといえる。したがって、スイープの処理量はMに比例することが分かる。正解は、「M」の (ウ) である。

・空欄c: 図Aで示したように、GCの1回の作動で再び利用できるようになるガーベジの数は、「M-L」で表すことができる。正解は (オ) である。

・空欄d: 難易度の高い設問であるが、このようなときは具体例で考えるとよい。ここで注意が必要なのは、グラフの縦軸が何を指すかという点である。空欄dの前には「単位時間当たりに再び利用できるようになるセルの数 (GCの効率)」とある。また、図5の後には「GCが作動している間、プログラムの実行は一時中断される。中断の時間はマーキングの処理量とスイープの処理量に依存する」とあり、空欄a, bの解説にあるように「マーキングの処理量はLに比例」、「スイープの処理量はMに比例」するので、GC1回の処理量は「M+L」(「L+M」の順序を入替え) に比例すると言える。そして、この1回のGCによって再び利用可能になるセルの数は空欄cの解説のとおり「M-L」である。このため、「単位時間当たりに再び利用できるようになるセルの数 (GCの効率)」は「M-L」を「M+L」で割ったものとして示すことができる。処理量÷処理時間であり、時間はαを比例定数として、α (M+L) の形で記述することができる。

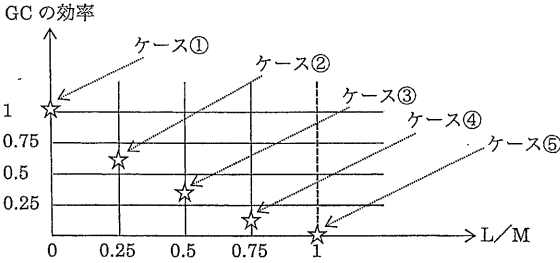
しかし、本問では、処理量、処理時間ともに単位が不明なので比例定数は無視して考える。これを踏まえて、ケース①～⑤のデータを考えてみると、表Aのようになる。ここではMを100として考える。

更に、表Aをグラフで表すと図Bのようになる。

表A M=100とした場合のデータ例と値

| ケース | M (全体のセル) | L (LIVEセル) | L/M (横軸) | GCの効率※ (縦軸) $\frac{M-L}{M+L}$ |
|-----|-----------|------------|----------|----------------------------------|
| ① | 100 | 0 | 0 | 1 |
| ② | 100 | 25 | 0.25 | 0.6 |
| ③ | 100 | 50 | 0.5 | 0.33 |
| ④ | 100 | 75 | 0.75 | 0.14 |
| ⑤ | 100 | 100 | 1 | 0 |

※GCの効率は、小数第3位を四捨五入している。



図B ケース①～⑤のグラフの様子

図Bで分かるように、L/Mの値は0～1の範囲となる。このとき、ケース①のようにLの値が0であれば、L/Mの値は0であり、全てのセルがガーベジとなり、「単位時間当たりに再び利用できるようになるセルの数 (GCの効率)」は理論上、多い。しかし、ケース⑤のようにLの値が100 (Mと等しい) であれば、L/Mの値は1となり、ガーベジのセルは存在せず、「単位時間当たりに再び利用できるようになるセルの数 (GCの効率)」は理論上、0である。ケース②～④の様子を確認すると、L/Mの値が0からだんだん大きくなるにつれ、GCの効率は小さくなることが分かる。このような傾向にあるグラフは、(ウ) だけである。

なお、他の選択肢については次のとおりである。
ア、エ: L/Mが1のとき、GCの効率が0となっていない。

イ: L/Mが1のとき、GCの効率は0となっているが、L/Mがある範囲の間、GCの効率が一定となっている。GCの効率はガーベジのセル数によって異なるはずであり、一定値にはならないと予測できる。

なお、参考として、図Bのグラフを反比例 (分数) 関数の式で示す。次のように変形することができる。

$$\begin{aligned} \frac{M-L}{M+L} &= \frac{1/M(M-L)}{1/M(M+L)} && \text{分数の分子、分母を同じ値で割っても値は同じになる。} \\ &= \frac{1-L/M}{1+L/M} && \text{例えば、}\frac{2}{4}=\frac{1}{2} \\ &= \frac{2-(1+L/M)}{1+L/M} && \text{分子を分母の}1+L/M\text{で割れる形にする。} \\ &= \frac{2}{1+L/M} - 1 \end{aligned}$$

L/M=xとし、一般的な式で示すと次のようになる。

$$y+1=\frac{2}{1+x}$$

この式は反比例関数y=1/xの原点をずらしたグラフとなり、例えば、(x, y)=(0, 1), (0.5, 0.33), (1, 0)を通り、「下に凸」なことが分かる。

【設問2】

GCが作動し、マーキング終了後の状態であるから、LIVEセルのマークビットは1となっており、ガーベジのセルのマークビットは0となっている。

マークビットが1であるセルはLIVEセルなので、LIVEリストに登録されているセルであり、「プログラムが使用中である」。ここで、正解は (イ) であることが分かったが、このことに注意して他の選択肢を吟味し、誤りの部分に対する説明を加える。

ア: 【セルの構造とその管理方法】の(3)①に「プログラムの実行開始時には、全てのセルは未使用セルとして、供給源が管理している」、②に「プログラムから取得要求があると、管理している未使用セルを提供する」と記述されている。また、【プログラムでのセルの使用方法】(1)に、「プログラムは新たにセルが必要になると、供給源に要求を出し、未使用セルを受け取り、LIVEから始まるリスト (以下、LIVEリストという) に登録して使用する。LIVEリストに登録され、プログラムが使用している状態のセルをLIVEセルという」と記述されている。つまり、プログラムが使用中のセルは、供給源の管理下ではなく、プログラムの管理下 (LIVEリスト) である。マーキング終了後のマークビットが1のセルはプログラムが使用中のセルなので、「供給源の管理下」ではなく、プログラムの管理下である。

ウ: 「LIVEリストから切り離された」セルは、マークビットが0のガーベジである。

エ: 「マーキングに続いて行われるスイープで、供給源に返却されることがある」のは、マークビットが0のガーベジである。