

アルゴリズムとデータ構造 2023 第6回 演習課題 「文字列探索とジェネリクス」

1. 下記は、最も単純な**文字列探索**アルゴリズムである。このメソッドは、文字列 `text` の中に文字列 `key` が含まれるか先頭から調べ、最初に見つかった位置 (0 以上) を返す。見つからなかった場合は-1 を返す。空欄を適切に埋めて実行し、動作を確認せよ。(補足: `charAt` は文字列の中の指定位置の 1 文字を返す)

```
public static int findString(String text, String key) {  
  
    int n = text.length();  
    int m = key.length();  
  
    // text の中の位置 i から key が入っているか調べる (残りが m 文字以上なら続ける)  
    for (int i = 0; i <=          ; i++) {  
  
        // その位置から順に、key の中の文字と同じであるか調べる (k は key の中の位置)  
        int k;  
        for (k = 0; k <          ; k++) {  
  
            if (text.charAt(          ) != key.charAt(          )) /* ★ */  
                break;  
        }  
        if (k == m) return i; // key の全文字 (m 文字) が一致したら発見  
    }  
  
    return  
}
```

2. 下記のクラスについて問いに答えよ (このクラスには適当なメソッドを追加してもよい)。

```
public class Item {          // 商品  
    public int code;        // 品番  
    public String name;    // 品名  
}
```

この `Item` の配列に対して、挿入ソートを用いて商品を品番順に整列するメソッドと、2分探索を用いて引数の (仮の) 商品データと同じ品番の商品を探し出すメソッドを作成し、動作を検証せよ。探索で見つからなかった場合は負数 (-1) を返すようにせよ。以下にメソッドの名前と引数を示す。

```
public static void sort(Item[] data)  
public static int binarySearch(Item[] data, Item key) // key は探索値を入れた仮の商品データ
```

3. 配列の整列や探索は自作するよりも `java.util.Arrays` の機能を使う方が便利である。Java では、クラスに `java.util.Comparable` (比較可能) インタフェースを実装すると、整列や探索のための順序が定義される。これを「自然な順序」という。まず、2.の `Item` を品番によって順序が付くように修正せよ。

```
public class Item implements Comparable<Item> { // Item が比較可能であることを示す  
    // 2.で定義した部分はそのまま  
  
    @Override  
    public boolean equals(Object obj) { /* プロ II の教科書 14.2.6 を参考に実装 */ }  
  
    @Override  
    public int compareTo(Item item2) { /* code の小さい順になるように実装 */ }  
  
    @Override  
    public int hashCode() { return this.code; } // equals をオーバーライドするとき適切に定義  
}
```

次に、この `Item` の配列に対して `Arrays.sort` (プロ II 教科書 6.6.1 参照) と `Arrays.binarySearch` (各自調べること) という静的メソッドを適用させて、2.と同じ結果を得るプログラムを作成せよ。

4. 配列を「自然な順序」以外の方法で整列・探索したい場合、まず、要素同士の新しい比較方法を定義するために `java.util.Comparator` インタフェースを実装したクラスを作る。以下に例を示す。

```
class ItemComparator implements Comparator<Item> { // Item のための比較方法のクラスの実装例
    @Override
    public int compare(Item a, Item b) { // 順番を比較し, a<b なら負, a==b なら 0, a>b なら正
        return b.code - a.code; // この例では, 順序として品番の「大きい順」が定義される
    }
}
```

そして、`Arrays.sort(items, new ItemComparator())` などとして、定義した `Comparator` クラスのインスタンスを追加の引数として整列および探索のメソッドに渡す。この機能を利用して、`Item` の配列を品番ではなく、品名で並べ替え、品名で探索するプログラムを作成せよ。

5. 普通の配列は作成時のサイズ（要素数）を変更することができないが、それが不便なことも多い。そこで、Java では、`java.util` パッケージに `ArrayList<E>`（`E` には要素のクラス名が入る）という、サイズが動的に変更できる配列（のようなクラス）が用意されている（プロ II 教科書 16.1.1 参照）。

ここで、`<E>` はジェネリクス（総称型）と呼ばれる記法で、`E` の部分にはインスタンス生成時に任意のクラス名を指定できる。ただし、`int`、`double` などの基本型は、`Integer`、`Double` などの対応するラッパークラス（プロ II 教科書 16.1.3 参照）で代用する。以下に、`ArrayList` の使用例をいくつか示す。

必要な import 文

```
import java.util.ArrayList;
```

動的配列の生成（クラス）

```
ArrayList<String> strArray = new ArrayList<String>();
strArray.add("abc"); // 末尾に要素を追加するのは.addを使う(配列の長さは自動的に拡張)
```

動的配列の使用例（基本型）

```
ArrayList<Integer> intArray = new ArrayList<Integer>();
Integer a = new Integer(32); // クラス版の整数(Integerクラス)のインスタンスを生成する
Integer b = 32; // 自動変換(autoboxing)機能を使えば基本型から直接生成できる
intArray.add(a);
intArray.add(64); // 引数に直接基本型を渡した場合も自動変換(autoboxing)される
```

配列要素へのアクセス

```
int n = intArray.size(); // 要素数を得るには.size()を使う(通常の配列の.lengthと同じ)
for (int i = 0; i < n; i++) {
    int c = intArray.get(i); // i番目の要素を取得するには.get(i)とする(この例ではInteger)
    System.out.println(c); // その際、自動変換(auto-unboxing)で基本型に直接代入している
}
strArray.set(0, "xyz"); // i番目の要素をxに入れ替えるには.set(i, x)とする
for (String str : strArray) // 通常の配列のように、拡張for文も使える
    System.out.println(str);
```

配列要素の探索

```
int j = intArray.indexOf(64); // 先頭から線形探索でデータを探す
int k = intArray.lastIndexOf(32); // 末尾から線形探索でデータを探す
```

これらの例を参考にして、2.で作成したプログラムを普通の配列ではなく `ArrayList<Item>` を利用するように書き換え、動作を検証せよ。`ArrayList` の詳細についてはマニュアル等を参照すること。

```
public static void sort(ArrayList<Item> data)
public static int binarySearch(ArrayList<Item> data, Item key)
```

6. `ArrayList` に対して、`Arrays.sort` や `Arrays.binarySearch` は使えないが、代わりに `java.util.Collections` の静的メソッドである `Collections.sort` および `Collections.binarySearch` をまったく同じように使える。3.および4.と同様の処理を行うプログラムを `ArrayList<Item>` を用いて作成し、動作を確認せよ。

7. 【発展】以下のプログラムは、効率的な文字列探索アルゴリズムであるボイヤー・ムーア法（簡易版）の例である。このアルゴリズムの特徴は、探索文字列の後ろから文字比較を行うことによって、不一致だった場合に探索位置を大きくスキップできることである。このアルゴリズムと 1.のアルゴリズムの文字比較回数を比べるために、それぞれのプログラムの/* ★ */の if 文の**前に**実行回数を数える処理を加え、それを最後に表示させるように改造せよ。さらに、ある程度長い英文テキストを使って回数を比較してみよ。

```
public class Main {

    // 日本語にも対応するため、マスクをかけて文字コードの下位 8 ビットだけ抽出して利用する
    final static int BITMASK = 0xff;

    // あらかじめ、テキストと探索文字列を比較中に、照合された文字が異なっていた場合、
    // テキスト側の文字から判断して、探索位置を何文字スキップできるか表にしておく
    public static int[] makeSkipTable(String pattern) {
        int len = pattern.length();
        int[] skipTable = new int[BITMASK + 1]; // 256 文字分の表

        // 探索文字列に含まれない文字の場合、探索文字列の長さの分だけ探索位置をスキップ
        for (int i = 0; i <= BITMASK; i++)
            skipTable[i] = len;

        // 探索文字列に含まれる文字の場合、その文字の位置が合うように探索位置をスキップ
        for (int i = 0; i < len - 1; i++) {
            int code = pattern.charAt(i) & BITMASK;
            skipTable[code] = len - i - 1;
        }

        return skipTable;
    }

    // ボイヤー・ムーア法（簡易版）による文字列探索
    public static int findStringBM(String text, String key) {

        // あらかじめ探索文字列からスキップ表を作成して準備しておく
        int[] skipTable = makeSkipTable(key);
        int n = text.length();
        int m = key.length();

        int i = 0;
        while (i <= n - m) {
            // 探索文字列の末尾の文字（m 文字目）から前に 1 文字ずつ照合していく
            int k;
            for (k = m - 1; k >= 0; m--) {
                // テキストから照合位置の文字を取り出し、探索文字列の対応位置の文字と比較する
                int code = text.charAt(i + k);
                if (code != key.charAt(k)) { /* ★ */
                    // 文字が合わなかった場合、その文字からスキップ量を調べる
                    i += skipTable[code & BITMASK];
                    break;
                }
            }
            if (k < 0) return i;
        }
        return -1;
    }

    public static void main(String[] args) {
        String text = "Tamagawa Gakuen Kogakubu";
        System.out.println(findStringBM(text, "gawa"));
        System.out.println(findStringBM(text, "ken"));
        System.out.println(findStringBM(text, "gaku"));
        System.out.println(findStringBM("T a m a g a w a   G a k u e n", "g a w a"));
    }
}
```