

badchar.py

```

1 def preprocess_bad_character(pattern):
2     """Preprocesses the bad character heuristic table using a dictionary."""
3     bad_char = {}
4     for i, ch in enumerate(pattern):
5         bad_char[ch] = i
6     return bad_char
7
8 def preprocess_good_suffix(pattern):
9     """Preprocess strong good suffix rule (standard approach)."""
10    m = len(pattern)
11    shift = [0] * (m + 1)
12    bpos = [0] * (m + 1)
13
14    # Step 1: Preprocess strong suffix
15    i = m
16    j = m + 1
17    bpos[i] = j
18    while i > 0:
19        while j <= m and pattern[i-1] != pattern[j-1]:
20            if shift[j] == 0:
21                shift[j] = j - i
22            j = bpos[j]
23            i -= 1
24            j -= 1
25        bpos[i] = j
26
27    # Step 2: Preprocess case 2 (when there is no occurrence of suffix in pattern)
28    j = bpos[0]
29    for i in range(m + 1):
30        if shift[i] == 0:
31            shift[i] = j
32        if i == j:
33            j = bpos[j]
34    return shift
35
36 def boyer_moore(text, pattern):
37     """Full Boyer-Moore string search algorithm with correct strong good suffix rule."""
38     n = len(text)
39     m = len(pattern)
40     if m == 0:
41         return []
42
43     bad_char = preprocess_bad_character(pattern)
44     good_suffix = preprocess_good_suffix(pattern)
45
46     result = []
47     s = 0 # shift of the pattern
48     while s <= n - m:

```

```
49     j = m - 1
50     while j >= 0 and pattern[j] == text[s + j]:
51         j -= 1
52
53     if j < 0:
54         result.append(s)
55         s += good_suffix[0]
56     else:
57         bc_shift = j - bad_char.get(text[s + j], -1)
58         gs_shift = good_suffix[j + 1]
59         s += max(bc_shift, gs_shift)
60
61     return result
62
63 # Example usage
64 text = "ABAAABCDABCDA"
65 pattern = "ABCDA"
66 matches = boyer_moore(text, pattern)
67 print("Pattern found at indices:", matches)
```