Sequence search with preprocessing

Author : Deepak Khatri

This way of searching for a seqence is faster than the frame shift method.

```
Algorithm:

Step 1: find all the position in seqence that match searchseqence[0]
i.e first letter of the string we are searching for in the original seqence.
And save all the position in a list. this stem is called the preprocessing step.

Step 2: use the position from step 1 and match the second letter of searchseqence in sequence only at positions + 1.
if not found remove the position from the list.

Step 3: move on with repeating for position 3 to length of searchsequence i.e repeat Step 2 until the string found.

Step 4: either we find the string else we return an appropriate error message.
```

In [1]:

```python
# data
seq="MOPMOUNTAINMADMONKEYMENMAD"
searchseq="MAD"
seq_len = len(seq)
searchseq_len = len(searchseq)
```

In [2]:

```python
# pre-processing
positions = []
for i in range(seq_len):
    if seq[i] == searchseq[0]:
        positions += [i]
positions
```

Out[2]:

```
[0, 3, 11, 14, 20, 23]
```

In [3]:

```python
# algorithm
for j in range(searchseq_len):
    for i in positions:
        if seq[i+j] != searchseq[j]:
            del positions[positions.index(i)]

if positions == []:
    print("Error ! seqence not found")
```

In [4]:

```python
# validating the result
# loop if multiple results are found
for i in positions:
    print(seq[i:i+searchseq_len], "at", i)
```

```
MAD at 11
MAD at 23
```