MODULE *LeastCircularSubstring*

An implementation of the lexicographically-least circular substring algorithm from the 1980 paper by *Kellogg S.* Booth. See: https://*doi.org*/10.1016/0020 − 0190(80)90149 − 0

EXTENDS *Integers*, *ZSequences*

CONSTANTS *CharacterSet*

ASSUME $CharacterSet \subseteq Nat$

```
****************************************************************************
--algorithm LeastCircularSubstring
  variables
    b ∈ Corpus ;
    n = ZLen(b) ;
    f = [index ∈ 0 .. 2 * n ↦ nil] ;
    i = nil ;
    j = 1 ;
    k = 0 ;
  define
    Corpus ≜ ZSeq(CharacterSet)
    nil ≜ − 1
  end define ;
  begin
L3: while j < 2 * n do
L5:     i := f[j − k − 1] ;
L6:     while b[j%n] ≠ b[(k + i + 1)%n] ∧ i ≠ nil do
L7:       if b[j%n] < b[(k + i + 1)%n] then
L8:         k := j − i − 1 ;
          end if ;
L9:       i := f[i] ;
        end while ;
L10:    if b[j%n] ≠ b[(k + i + 1)%n] ∧ i = nil then
L11:      if b[j%n] < b[(k + i + 1)%n] then
L12:        k := j ;
          end if ;
L13:      f[j − k] := nil ;
        else
L14:     f[j − k] := i + 1 ;
        end if ;
LVR:  j := j + 1 ;
    end while ;
end algorithm ;
****************************************************************************
```

BEGIN TRANSLATION ($chksum(pcal) =$ "$c2e05615$" $\land\ chksum(tla) =$ "$81694c33$")
VARIABLES $b,\ n,\ f,\ i,\ j,\ k,\ pc$

define statement
$Corpus \triangleq ZSeq(CharacterSet)$
$nil \triangleq -1$

$vars \triangleq \langle b,\ n,\ f,\ i,\ j,\ k,\ pc\rangle$

$Init \triangleq$ Global variables
$\land\ b \in Corpus$
$\land\ n = ZLen(b)$
$\land\ f = [index \in 0\ ..\ 2 * n \mapsto nil]$
$\land\ i = nil$
$\land\ j = 1$
$\land\ k = 0$
$\land\ pc =$ "L3"

$L3 \triangleq$ $\land\ pc =$ "L3"
$\land$ IF $j < 2 * n$
THEN $\land\ pc' =$ "L5"
ELSE $\land\ pc' =$ "Done"
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j,\ k\rangle$

$L5 \triangleq$ $\land\ pc =$ "L5"
$\land\ i' = f[j - k - 1]$
$\land\ pc' =$ "L6"
$\land$ UNCHANGED $\langle b,\ n,\ f,\ j,\ k\rangle$

$L6 \triangleq$ $\land\ pc =$ "L6"
$\land$ IF $b[j\%n] \neq b[(k + i + 1)\%n] \land i \neq nil$
THEN $\land\ pc' =$ "L7"
ELSE $\land\ pc' =$ "L10"
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j,\ k\rangle$

$L7 \triangleq$ $\land\ pc =$ "L7"
$\land$ IF $b[j\%n] < b[(k + i + 1)\%n]$
THEN $\land\ pc' =$ "L8"
ELSE $\land\ pc' =$ "L9"
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j,\ k\rangle$

$L8 \triangleq$ $\land\ pc =$ "L8"
$\land\ k' = j - i - 1$
$\land\ pc' =$ "L9"
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j\rangle$

$L9 \triangleq$ $\land\ pc =$ "L9"

$\land\ i' = f[i]$
$\land\ pc' = \text{“L6”}$
$\land$ UNCHANGED $\langle b,\ n,\ f,\ j,\ k\rangle$

$L10 \triangleq$ $\land\ pc = \text{“L10”}$
$\land$ IF $b[j\%n] \neq b[(k+i+1)\%n] \land i = nil$
THEN $\land\ pc' = \text{“L11”}$
ELSE $\land\ pc' = \text{“L14”}$
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j,\ k\rangle$

$L11 \triangleq$ $\land\ pc = \text{“L11”}$
$\land$ IF $b[j\%n] < b[(k+i+1)\%n]$
THEN $\land\ pc' = \text{“L12”}$
ELSE $\land\ pc' = \text{“L13”}$
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j,\ k\rangle$

$L12 \triangleq$ $\land\ pc = \text{“L12”}$
$\land\ k' = j$
$\land\ pc' = \text{“L13”}$
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ j\rangle$

$L13 \triangleq$ $\land\ pc = \text{“L13”}$
$\land\ f' = [f$ EXCEPT $![j-k] = nil]$
$\land\ pc' = \text{“LVR”}$
$\land$ UNCHANGED $\langle b,\ n,\ i,\ j,\ k\rangle$

$L14 \triangleq$ $\land\ pc = \text{“L14”}$
$\land\ f' = [f$ EXCEPT $![j-k] = i+1]$
$\land\ pc' = \text{“LVR”}$
$\land$ UNCHANGED $\langle b,\ n,\ i,\ j,\ k\rangle$

$LVR \triangleq$ $\land\ pc = \text{“LVR”}$
$\land\ j' = j+1$
$\land\ pc' = \text{“L3”}$
$\land$ UNCHANGED $\langle b,\ n,\ f,\ i,\ k\rangle$

Allow infinite stuttering to prevent deadlock on termination.
$Terminating \triangleq pc = \text{“Done”} \land$ UNCHANGED $vars$

$Next \triangleq L3 \lor L5 \lor L6 \lor L7 \lor L8 \lor L9 \lor L10 \lor L11 \lor L12 \lor L13 \lor L14$
$\lor\ LVR$
$\lor\ Terminating$

$Spec \triangleq Init \land \Box[Next]_{vars}$

$Termination \triangleq \Diamond(pc = \text{“Done”})$

END TRANSLATION

$TypeInvariant \triangleq$
$\land\ b \in Corpus$
$\land\ n = ZLen(b)$
$\land\ f \in [0\,..\,2 * n \to 0\,..\,2 * n \cup \{nil\}]$
$\land\ i \in 0\,..\,2 * n \cup \{nil\}$
$\land\ j \in 0\,..\,2 * n \cup \{1\}$
$\land\ k \in ZIndices(b) \cup \{0\}$

Is this shift the lexicographically-minimal rotation?

$IsLeastMinimalRotation(s,\ r) \triangleq$
LET $rotation \triangleq Rotation(s,\ r)$IN
$\land \forall\, other \in Rotations(s) :$
$\land\ rotation \preceq other.seq$
$\land\ rotation = other.seq \Rightarrow (r \leq other.shift)$

$Correctness \triangleq$
$pc =$ "Done" $\Rightarrow IsLeastMinimalRotation(b,\ k)$