I collaborated with: W3023339

Problem
Let Σ be a finite alphabet (evocative synonym for 'set') and let $w = \tau_1\tau_2 \dots \tau_m$ and $p = \sigma_1\sigma_2 \dots \sigma_n$ be two finite sequences of elements of Σ.[1] We say that the word w *contains* the pattern p if $\sigma_1, \sigma_2, \dots, \sigma_n$ occur in w in the same order (but not necessarily consecutively) that they occur in p. That is, w contains p if for some $\tau_{i_1}, \dots, \tau_{i_n}$ with $i_1 < i_2 < \dots < i_n$ we have $\tau_{i_1} = \sigma_1, \dots, \tau_{i_n} = \sigma_n$. Design an $O(n + m)$ (greedy) algorithm to determine, given a pair w and p, whether w contains the pattern p, and prove that your algorithm correctly solves the problem in this amount of time.
Solution

Algorithm 1 Determines whether w contains the pattern p

Require: sequences w and p in Σ.

1: $index \leftarrow 0$
2: **for** every τ_i in w **do**
3: **if** $\tau_i = \sigma_{index}$ **then**
4: $index \leftarrow index + 1$
5: **end if**
6: **end for**
7: **return** true if $index = n + 1$ else false

Proof. Base Case: Let p be of length 1. As long as $\sigma \in w$, then w contains p. Our algorithm will walk through w, comparing every letter to σ and, if a match is found, $index$ will be 1, so true will be returned.
Inductive Hypothesis: Our algorithm works for a pattern p of size k, where k is an integer. That is, the algorithm returns true if a word w contains p.
Inductive Step: We must show our algorithm works for a pattern of size $k + 1$. From our inductive hypothesis, we know it will check if the first k elements of p are in w, so we have to see what happens with this extra element. If w is of size less than $k + 1$, then our w does not contain p and $index$ from our algorithm will not get to be $k + 2$. If w is of size greater than k, then it already checked the first k elements of p if there were k matches, and $index$ is $k + 1$. It will then check for an element τ that matches this $k + 1$ element. If there is a match, w contains p and the algorithm returns true, otherwise w does not contain p and the algorithm returns false. □

[1] Tradition: Greek letters are used when denoting alphabets and their "letters".