1. A don’t care character # is a special character that matches any single character. For example, the pattern #oke#i matches sokeri, pokeri and tokeni.

(a) Modify the Shift-And algorithm to handle don’t care characters.

(b) It may appear that the Morris–Pratt algorithm can handle don’t care characters almost without change: Just make sure that the character comparisons are performed correctly when don’t care characters are involved. However, such an algorithm would be incorrect. Give an example demonstrating this.

2. Let \( P_k = \{ P_1, \ldots, P_{2k} \} \) be a set of patterns such that

- for \( i \in [1..k] \), \( P_i = a^i \) and
- for \( i \in [k+1..2k] \), \( P_i = P'_i a^k \) such that \( |P'_i| = k \) and each \( P'_i \) is different.

(a) Show that the total size of the sets \( \text{patterns}(\cdot) \) in the Aho–Corasick automaton for \( P_k \) is asymptotically larger than \( ||P_k|| \).

(b) Describe how to represent the sets \( \text{patterns}(\cdot) \) so that

- the total space complexity is never more than \( O(||P||) \) for any \( P \)
- each set \( \text{patterns}(\cdot) \) can be listed in linear time in its size.

3. Show that edit distance is a metric, i.e., that it satisfies the metric axioms:

- \( ed(A, B) \geq 0 \)
- \( ed(A, B) = 0 \) if and only if \( A = B \)
- \( ed(A, B) = ed(B, A) \) (symmetry)
- \( ed(A, C) \leq ed(A, B) + ed(B, C) \) (triangle inequality)

4. Let \( \Sigma = \{ a, b, c \} \). Define the function \( \gamma : \Sigma \times \Sigma \rightarrow \mathbb{R}_{\geq 0} \) as follows

\[
\gamma(a, a) = \gamma(b, b) = \gamma(c, c) = 0
\]
\[
\gamma(a, b) = \gamma(b, c) = \gamma(c, a) = 0.5
\]
\[
\gamma(b, a) = \gamma(c, b) = \gamma(a, c) = 1.5
\]

Let \( ed_\gamma \) be a weighted edit distance, where the cost of substituting a character \( x \) with a character \( y \) is \( \gamma(x, y) \). The cost of insertions and deletions is 1.

(a) It might seem that we can compute \( ed_\gamma(A, B) \) using the recurrence for the standard edit distance (slide 117 on the lecture notes) except \( \delta \) is replaced by \( \gamma \). Show that this is not the case by providing an example for which the recurrence produces an incorrect distance.

(b) Is \( ed_\gamma \) a metric?

5. Let \( P = \text{evete} \) and \( T = \text{neeteneeveteen} \). Use Ukkonen’s cut-off algorithm to find the occurrences of \( P \) in \( T \) for \( k = 1 \).