Actividad Integradora 1

María Fernanda Ramírez Barragán	A01750879	ITC
Diego Armando Ulibarri Hernández	A01636875	ITC
Raul Youthan Irigoyen Osorio	A01750476	ITC

Fecha: 13/10/21

Instituto Tecnológico y de Estudios Superiores de Monterrey

Analysis and Design of Advanced Algorithms (Gpo 602)

For this integrative activity we used the **KMP** algorithm, an algorithm that helps us in identifying patterns in a certain sequence, this process consists of comparing the given pattern within a string sequence, finding mismatches and detecting matches; the activity's main objective is to simulate the search for malware within a transmission: we must compare two sequences of characters: our pattern (*mcoden.txt*), which pretends to act as malign code, and the larger sequence (*transmission.txt*) where the malicious code could be present.

The algorithm iterates through the main string until the first character of the pattern is found within the larger sequence, then it iterates between both strings to see if the pattern exists within said sequence. Finally, the algorithm looks up for any potential repeated patterns within the match it just found, if there are no more potential matches, the algorithm skips to the last index it traveled to, which is why the algorithm has a time complexity of O (m + n) in it's worst-case scenario, because of the complexities of both parts of this algorithm, where the first one has a complexity of O(m) and the second has a complexity of O(n), whereas it's best-case time complexity is O(m), with m being the length of the pattern, since the table is still being built.

Pattern matching is commonly used in computer science and information processing in cases of the following nature:

- **DNA sequence matching:** DNA Pattern matching is applied to gather, store, analyze and merge biological data. These types of algorithms solve the problem of finding subsequences within a long DNA sequence. "The pattern matching algorithms can be applied for detecting the unusual patterns present in the gene database. It can show how the disease can be transformed from parents to their children and can identify the presence of the disease on hereditary basis and its impact." (IJSETR, 2014)
- **Language syntax checker:** The syntax checker looks for syntax errors in each statement, according to the dataset type, if an error is found the syntax checker highlights the syntax that is incorrect and displays an error message so the user can correct the error.
- **Search engines:** Search engines are designed to carry out web searches. The search results are presented in a long chain of links. Pattern matching is useful, so that the only links that are shown to the user are the ones they were looking for

References

- IJSETR. Comparison of Three Pattern Matching Algorithms using DNA Sequences. *International Journal of Scientific Engineering and Technology Research. 03. 6916-6920. Recuperado de:* http://ijsetr.com/uploads/625413IJSETR2868-162.pdf