# Tutorial 10: March 25

1. **Karp-Rabin**
For Karp-Rabin pattern matching, consider the following hash function for strings over the alphabet $\{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}\}$:

$$h(P) = (\#\text{ of occurrences of }\texttt{A}) + 2 \times (\#\text{ of occurrences of }\texttt{C}) + 3 \times (\#\text{ of occurrences of }\texttt{G}) + 4 \times (\#\text{ of occurrences of }\texttt{T})$$

Given the pattern $P = \texttt{TAGCAT}$ and sequence $T = \texttt{TGCCGATGTAGCTAGCAT}$, use the table below to show all the character comparisons performed during Karp-Rabin pattern matching. Start a new pattern shift (in which character comparison occurs) in a new row. You may not need all the available space.

| T | G | C | C | G | A | T | G | T | A | G | C | T | A | G | C | A | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

Table 1: Table for Karp-Rabin problem.

2. **Boyer-Moore, Revisited**
Let $P[0...5] = \texttt{payday}$ and let $T[0...12] = \texttt{daypayplayayaya}$.

   a) Compute the last-occurence array $S[0...5]$ for the pattern $P$.

   b) Show the search for $P$ in $T$ using the Boyer-Moore algorithm using only the bad character heuristic. Also, put square brackets around characters that are known to be matched, even if the algorithm matches them again.

3. **Moore's first name is 'J', not abbreviated**
Consider using the Boyer-Moore algorithm with only the Bad Character heuristic to search for a pattern $P$ of length $m$ in a text $T$ of length $n$, with $n > m$, where $P$ does **not** appear in $T$.

   a) Give an example of a pattern $P$ with length $n$ and text $T$ with length $n$ that achieves the worst-case runtime for searching. Do not consider preprocessing time.

   b) Same question, but for the best-case runtime.