22.1.09

KMP

יהי T טקסט באורך n ותבנית P באורך m. KMP מחשב לכל $1 \le i \le n$ את $\sigma[i]$ - אורך הרישא המקסימלית של P שהיא סיפא של T_i (כאשר $T_i = T[1...i]$).
לפני כן, נחשב לכל $1 \le j \le m$ את $\pi[j]$ = אורך הרישא המקס' באורך $< j$ של P שהיא סיפא של P_j.

דוגמא:

$P = ababc$; $T = cabcababca$

$\quad 00120 \qquad 0120123451$

הגדרה: עבור מחרוזת T (נסמן) ב-T^R את המחרוזת ההפוכה. למשל אם $T = aabc$ אז $T^R = cbaa$. T תקרא פלינדרום אם $T = T^R$.

תרגיל: נתונה מחרוזת T, מצאו את הרישא הארוכה ביותר של T שהיא פלינדרום ביעילות.

דוגמא: $T = abacabacbab$

פתרון: נרשום את $T = \underline{x}y$ ונניח שמצאנו כלומר $x = x^R$. לכן $T^R = y^R\underline{x}$.
אם כן, נרצה למצוא רישא מקסימלית של T שהיא גם סיפא של T^R.

נריץ KMP עם טקסט T^R ותבנית T. האורך שאנו רוצים הוא בדיוק $\sigma[n]$.

זמן הריצה: $O(|T|+|T^R|) = O(2|T|) = O(|T|)$

תרגיל: תהי T באורך $n \ge 10$ ורוצים למצוא חלוקה $T = xyx$ כך ש $|y| \ge 10$ ו- $|x|$ מקסימלי. ביעילות.

פתרון: בכל חלוקה כזו $|x| \le \frac{n-10}{2}$. נריץ אם כן, KMP כשהתבנית היא $T_{\lfloor\frac{n-10}{2}\rfloor}$ והטקסט הוא $\lfloor\frac{n-10}{2}\rfloor$ התווים האחרונים של T.

$\sigma\left[\substack{\text{המקום}\\\text{האחרון}}\right]$ $\left(\lfloor\frac{n-10}{2}\rfloor\right)$ יתן את אורך ההתאמה המקסימלית.

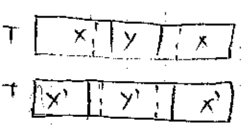
זמן הריצה לינארי.

תרגיל: יהי T באורך n, רוצים למצוא ביעילות חלוקה $T = xyx$ כך ש y היא תת מחרוזת של x. (אם קיימת כזו).

פתרון: נריץ KMP עם טקסט T ממקום $T[\lfloor\frac{n}{2}\rfloor+1...n]$ עם תבנית $T[1....n]$ כדי למצוא חלוקה $T = xyx$ עם x מקסימלי.

נריץ KMP עם טקסט x ותת מחרוזת y כדי לבדוק אם y היא תת מחרוזת של x.

אם כן, נחזיר "כן" אחרת, נחזיר "לא".



נכונות: אם מצאנו חלוקה - ברור.
נניח שהאלגוריתם "לא" ובשלילה נניח שכן קיימת חלוקה מתאימה.
ממקסימליות x, מתקיים כי x' היא רישא של x.
y לעומת זאת היא תת מחרוזת של y'. y' היא תת מחרוזת של x' (לפי ההנחה בשלילה) שהיא תת מחרוזת של x.
y היא תת מחרוזת של y', ולכן קיבלנו ש y היא תת מחרוזת של x וזו סתירה.

סיבוכיות: זמן ריצה לינארי.