

LABORATORIUM NR 11

---

WYSZUKIWANIE WZORCA W TEKŚCIE

---

**Zadanie ALZ.11.1** Rozważmy tekst  $T = bbbbab$  oraz wzorzec  $P = bbba$ . Załóżmy, że tablice indeksujemy od 1, czyli np. dla wzorca  $P$  mamy:

pozycja	1	2	3	4
znak	$b$	$b$	$b$	$a$

- a) Prefiks  $P_3$  wzorca  $P$  to  $bbb$ . Jaki jest maksymalny prefiks właściwy  $P_3$ , który jest jego sufixem?
- b) Funkcja prefiksowa  $\pi: \{1, 2, 3, 4\} \rightarrow \{0, 1, 2, 3\}$  określona jest następująco:

$$\pi(i) \doteq \max\{k : k < i \text{ oraz } P_k \text{ jest sufixem } P_i\}.$$

Innymi słowy,  $\pi(i)$  jest maksymalną długością prefiksu wzorca  $P$  będącego właściwym sufixem  $P_i$ . Wyznacz funkcję prefiksów dla wzorca  $P = bbba$ .

- c) Jakie znaki tekstu i wzorca będą porównane ze sobą w czasie działania algorytmu naiwnego i algorytmu Knutha-Morrisa-Pratta, jeżeli danymi wejściowymi będą wzorzec  $P$  i tekst  $T$ ?

**Zadanie ALZ.11.2** (2,5 pkt) Zmodyfikuj algorytm Knutha-Morrisa-Pratta tak, aby wyszukiwał w tekście  $T$  danego wzorca  $X$  w przy założeniu, że każdy znak zapytania '?' w tekście  $T$  może odpowiadać dowolnemu znakowi, a następnie zaimplementuj tę modyfikację.

**Zadanie ALZ.11.3** ((1+2+2)+2\* pkt) Zaimplementuj i porównaj algorytmy (ich czas działania) wyszukiwania wzorca w tekście:

- a) oczywisty algorytm (1 pkt);
- b) algorytm Rabina-Karpa (2 pkt);
- c) algorytm Knutha-Morrisa-Pratta (2 pkt);
- d) oraz algorytm wykorzystujący automat skończony (+2 pkt)

na przykładowym wzorcu i tekście podanych w pliku `wzorzec-tekst.txt` pod adresem <http://inf.ug.edu.pl/~zylinski/dydaktyka/AiSD/wzorzec-tekst.txt>.

**Uwaga.** Należy ignorować znaki zmiany wiersza we wzorcu i w tekście. Program powinien wypisywać pozycje, na których znaleziono wystąpienie wzorca.