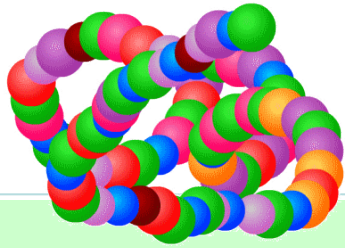


התאמת מילים

התאמת מילה???

המטרה: היא למצוא את כל המופעים של המילה בטקסט נתון



התאמת מילים

- **פיתוח אפליקציות בביולוגיה חישובית**

- רצף DNA הוא מילה ארוכה מעל הא"ב בעל 4 אותיות {A,C,T,G}:

```
GTTTGAGTGGTCAGTCTTTTCGTTTCGACGGAGCCCCCAATTAATA  
AACTCATAAGCAGACCTCAGTTCGCTTAGAGCAGCCGAAA.....
```

- חיפוש תבניות מסוימת ברצפים של DNA

- **מציאת המופעים של מילת תבנית בטקסט, שנוצר באמצעות האלפבית גדול**

- בתכניות עיבוד תמלילים

- חיפוש באינטרנט

- Desktop search (Google, MSN)

- **התאמת, מילת סיביות, המכילה**

- נתונים גרפיים

- קוד המכונה

- **תכנות בעזרת מיליוני דוגמאות: חיפוש קוד לשימוש חוזר, בעזרת מערכת חדשנית**

- **חיפוש שורות שמתאימות לביטויים מסוימים ב-Linux**

- הפקודה grep מחפשת בקובץ או במספר קבצים, שורות שמתאימות לביטוי הנתון.

התאמת מילים

- **הבעיה:** נתון טקסט T באורך n ותבנית P באורך m ($n \ll m$). מטרת האלגוריתם היא לבדוק האם התבנית P מופיעה בתוך הטקסט T והיכן.
- שימושים לדוגמה:

```
ATGGGAATCCCAATGGGGAAGTCGATGCTGGTGCCTTCACCTTCCTGGCCCTCGCCCTCGTGCATTCCTGCTGGCCG  
GCCGACCGGTCACCATCACCATCACCACGCGCCGGGGCCGTAACGACATCTTCGAGGCCCAGAAGATCGAGTGGCAGCAAC  
CTTTCGATCTGACTGAATGTTTCATTCAAAGATGGGGCTGGCAACTCCTTCGACCTCTCGTCCCTGTCAAGGTACAGTGAC  
AACTGGGAAGCCATCACTGGGACGGGGGACCCGGAGCACTACCTCATCAATGTCTGCAAGTCTCTGGCCCCGAGGCTGG  
CACTGAGCCGTGCCCTCCAGAAGCAGCCGCGTGTCTGCTGGGTGGCTCCAAGCCCGTGAACCTCGGCAGGGTAAGGGACG  
GACCTCAGTGGAGAGATGGCATAATTGTCCTGAAATACGTTGATGGCGACTTATGTCCAGATGGGATTCGGAAAAAGTCA  
ACCACCATCCGATTCACCTGCAGCGAGAGCCAAGTGAACCTCAGGCCCATGTTTCATCAGCGCCGTGGAGGACTGTGAGTA  
CACCTTTCGCTGGCCACAGCCACAGCCTGTCCCATGAAGAGCAACGAGCATGATGACTGCCAGGTCACCAACCCAAGCA  
CAGGACACCTGTTTGATCTGAGCTCCTTAAGTGGCAGGGCGGGATTCACAGCTGCTTACAGCGAGAAGGGGTGGTTTAC  
ATGAGCATCTGTGGGGAGAATGAAAAGTGCCTCCTGGCGTGGGGCCTGCTTTGGACAGACCAGGATTAGCGTGGGCA  
GGCCAACAAGAGGCTGAGATACGTGGACCAGGTCCTGCAGCTGGTGTACAAGGATGGGTCCCTTGTCCCTCCAAATCCG  
GCCTGAGCTATAAGAGTGTGATCAGTTTCGTGTGCAGGCCCTGAGGCCGGGCCAACCAATAGGCCCATGCTCATCTCCCTG  
GACAAGCAGACATGCACTCTCTTCTCTCCTGGCACACGCCGTGGCCTGCGAGCAAGCGACCGAATGTTCCGTGAGGAA  
TGGAAAGCTCTATTGTTGACTTGTCTCCCTTATTCATCGCACTGGTGGTTATGAGGCTTATGATGAGAGTGAGGATGATG  
CCTCCGATACCAACCCTGATTTTACATCAATATTTGTAGCCACTAAATCCCATGCACGGAGTGCCTGTCTGCCGGA  
GCCGCTGTGTGCAAAGTTCCTATTGATGGTCCCCCATAGATATCGGCCCGGGTAGCAGGACCACCAATACTCAATCCAAT  
AGCAAATGAGATTTACTTGAATTTTGAAGCAGTACTCCTTGCTTAGCGGACAAGCATTTCAACTACACCTCGCTCATCG  
CGTTTCACTGTAAGAGAGGTTGTGAGCATGGGAACGCCTAAGCTGTTAAGGACCAGCGAGTGCAGACTTTGTGTTTGAATGG  
GAGACTCCTGTGCTGTCTGCTGATGAAGTGAGGATGGATGGCTGACCTGACAGATGAGCAGCTCCTCTACAGCTTCAA  
CTTGTCCAGCCTTTCACAGGACCTTTAAGGTGACTCGGACTCGCGCACTACAGCGTTGGGGTGTGCACCTTTGCGAG  
TCGGGCCAGAACAAGGAGGCTGTAAGACCGGAGGAGTCTGTCTGCTCTCAGGCACCAAGGGGGCATCCTTTGGACGGCTG  
CAATCAATGAAACTGGATTACAGGCACCAGGATGAAGCGGTGCTTTTAAAGTTACGTGAATGGTGATCGTTGCCCTCCAGA  
AACCAGTACGGCGTCCCTGTGTCTTCCCTTCATATTTCAATGGGAAGAGCTACGAGGAGTGCATCATAGAGAGCAGGG  
CGAAGCTGTGGTGTAGCACAACTGCGGACTACGACAGAGACCAGAGTGGGGCTTCTGCAGACACTCAAACAGCTACCGG  
ACATCCAGCATCATATTTAAGTGTGATGAAGATGAGGACATTTGGGAGGCCACAAGTCTTCAGTGAAGTGGCTGGGTGCTG
```

התאמת מילים

All of the sudden fits of soft, but high pitched laughter came from the far bank. Farrel had expected to see a goblin leap down from the bushes to eat him. Alas though he saw nothing, yet the laughter continued to filter out from the bushes on the far bank. He peered intently along the far side of the stream, but still saw nothing but trees.

“You will never see me when I wish to blend with the trees,” sang the voice. Terrified, Farrel dashed back to the water’s edge where he had dropped his carry sac, opened it and withdrew his spell book.

“Come out! Or I shall cast a spell!”

“Oh...a magician you be. Cast spells will thee?” the voice sang melodiously. “Magic I know too. Much more I think than you.”

“Who are you...what are you?” demanded Farrel.

At that moment without warning something burst out of the bushes right where Farrel was looking. It flew quickly across the stream and right at Farrel’s nose. Out of self preservation he shut his eyes and held his spell book over his face to try and shield the collision, but it never came. Farrel opened his eyes slowly and lowered the book. His eyes almost touched the end of his nose and his heart skipped several beats at the sight of what was now hovering about two inches from his face.

“A human you must human be. You are very large compared to me,” sang the little creature.

Farrel was totally dumbfounded. He could not believe his eyes. In front of his nose was the

התאמת מילים

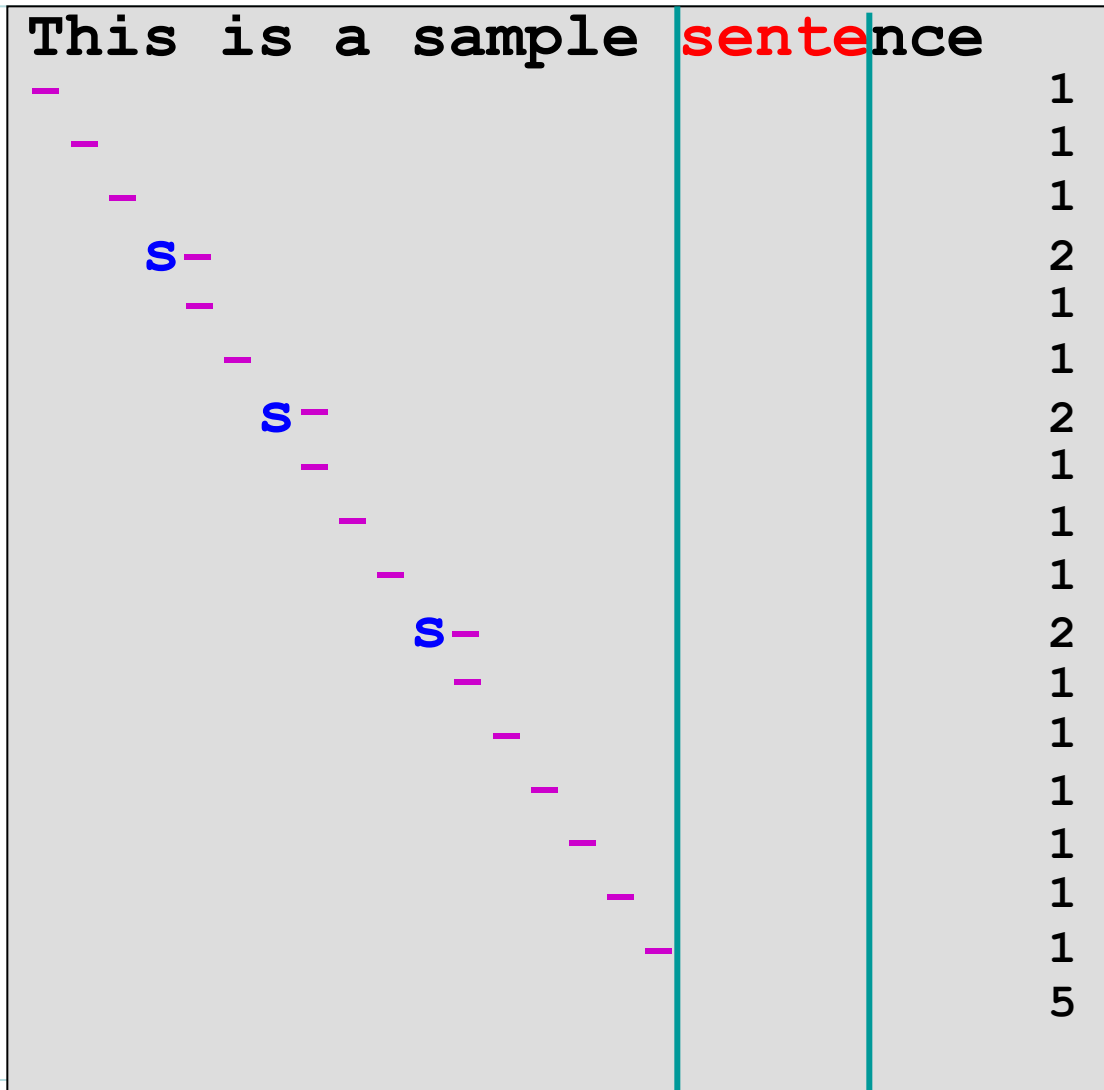
- טקסט T $T[0..n-1]$, כאשר n אורך הטקסט
 $T = \text{“abacaabaccabacabaabb...”}$
- מילת תבנית P $P[0..m-1]$, כאשר m אורך התבנית
 $P = \text{“abacab”}$
- המופע הראשון של מילת התבנית P בטקסט T
 $T[10..15] = P[0..5]$
- בדרך כלל $m \gg n$

התאמת מילים



- האלגוריתם הנאיבי להתאמה
- $22+6=28$ השוואות

התאמת מילים



- האלגוריתם הנאיבי להתאמה
השוואות $20+5=25$
- כמו בדוגמה הקודמת, מספר
ההשוואות כמעט שווה
למספר אותיות הטקסט
שנסרקו.
- המקרה הטוב ביותר

התאמת מילים

- האלגוריתם הנאיבי ניתן לדמות להחלקת "סרגל" המכיל את התבנית לאורך הטקסט, וחיפוש המופעים, שעבורם כל האותיות בתבנית, שוות לאותיות במקומות המקבילים בטקסט.
- באלגוריתם הנאיבי, במקרה הגרוע ביותר, צריך לעבור על אותיות הטקסט יותר מפעם אחת! לא תמיד אפשרי לעבור על הטקסט מספר פעמים.
- פונקצית זמן הריצה של אלגוריתם היא:
$$F(n + m) = (n - m + 1) \cdot m = n \cdot m - m^2 + m$$
- סדר הגודל: $O(n \cdot m)$

טקסט T

תבנית P

תבנית P

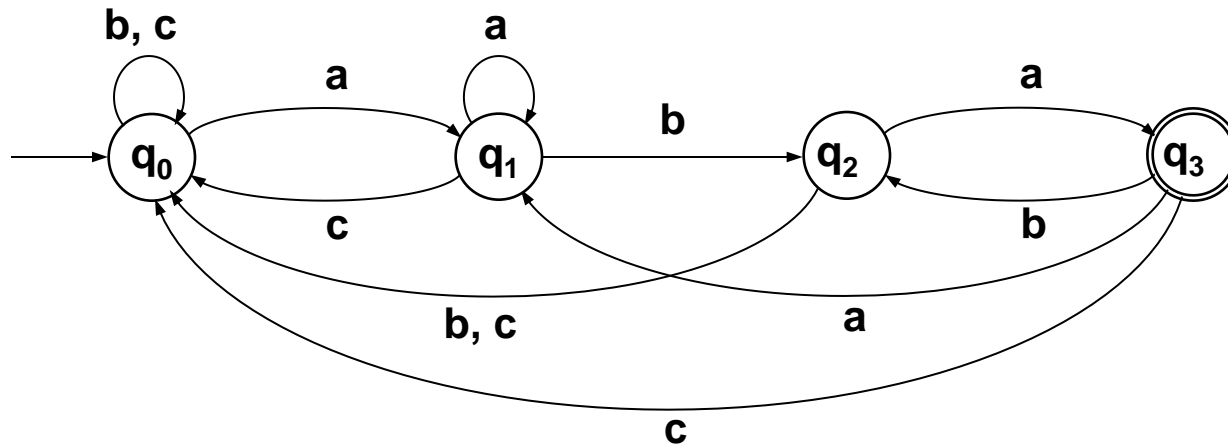
תבנית P

התאמת מילה באמצעות אוטומט סופי

- נציג אלגוריתם אחר הפותר את הבעיה.
- אלגוריתם זה משתמש באוטומט סופי דטרמיניסטי, המקבל את שפת כל המילים מעל הא"ב Σ , שמסתיימות ברצף האותיות של מילת התבנית P .
- לדוגמה: $P = aba$, הא"ב $\Sigma = \{a, b, c\}$.

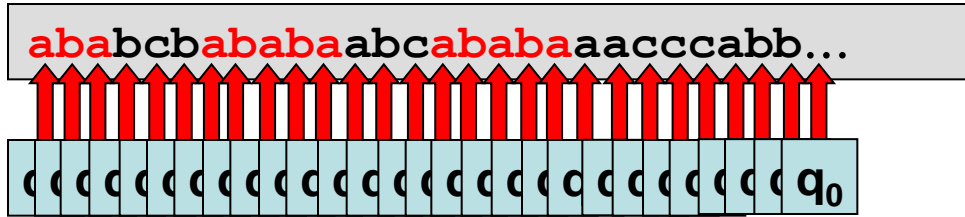
התאמת מילה באמצעות אוטומט סופי

אוטומט סופי דטרמיניסטי המקבל את שפת כל המילים מעל הא"ב $\Sigma = \{a, b, c\}$ שמסתיימות ברצף האותיות של המילה $P = aba$.



- המצב q_0 – מצב התחלתי הזוכר כי לא נקרא עדיין קטע מן המילה.
- המצב q_1 – זוכר כי האות האחרונה שנקלטה הייתה **a** והיא האות הראשונה של המילה.
- המצב q_2 – זוכר כי שתי האותיות האחרונות שנקלטו היו **ab** והן שתי האותיות הראשונות של המילה.
- המצב q_3 – זוכר כי המילה **aba** כולה נקלטה והוא שוב במצב המקבל.

התאמת מילה באמצעות אוטומט סופי



- מופע המילה **aba**
- מופע המילה **aba**
- מופע המילה **aba**
- מופע המילה **aba**
- מופע המילה **aba**

