מבני נתונים 2 — תרגיל בית 3

(2 עמודים)

הגשה: 02/Jul/2006 (בזוגות)

שאלה 13

תהי A קבוצה בת a איברים ו־$B = \{0, 1, \ldots, b-1\}$ כך ש־b מחלק את a. פונקציה $h : A \rightarrow B$ תקרא שוות פיזור אם לכל $i \in B$ מספר האיברים x המקיימים $h(x) = i$ הוא a/b.

תהי H קבוצת כל הפונקציות מ־A ל־B שהן שוות פיזור.

א. הוכיחו: $\delta_H(A, A) = |H|a(\frac{a}{b} - 1)$.

ב. הוכיחו ש־H אוניברסלית.

ג. הוכיחו או הפריכו: אם U היא משפחה אוניברסלית של פונקציות מ־A ל־B אז כל $h \in U$ היא שוות פיזור.

שאלה 14

א. נרחיב את הבעיה של string matching באופן הבא: התבנית P יכולה להכיל מופעים של התו המיוחד "*". תו זה אינו נמצא בא"ב של הטקסט T והוא מתאים למחרוזת כלשהי (כולל מחרוזת ריקה).

דוגמא: אם T=acdababac ו־ P=a*b* אז P מתאימה ל־T במקומות 1, 4 ו־6.

הציעו אלגוריתם המוצא את כל המקומות בהם התבנית מתאימה לטקסט בסיבוכיות זמן $O(n+m)$, כאשר n הוא אורך הטקסט ו־m הוא אורך התבנית.

ב. בהינתן שתי תבניות P ו־P' תאר כיצד בונים אוטומט סופי המוצא בטקסט את כל המופעים של שתי התבניות. נסה למזער את מספר המצבים באוטומט שלך.

שאלה 15

פלינדרום הוא מחרוזת ששוה למחרוזת ההפוכה לה (לדוגמא, המחרוזת abbfbba). עבור מחרוזת קלט $S[1..n]$, נאמר שהיא מכילה פלינדרום ברדיוס k סביב האינדקס i אם התת־מחרוזת $S[i-k..i+k]$ היא פלינדרום. (כמובן, יש לדרוש $1 \leq k < i \leq n-k$). נאמר ש־S מכילה פלינדרום מקסימלי ברדיוס k סביב האינדקס i אם היא מכילה פלינדרום ברדיוס k סביב האינדקס i ואינה מכילה פלינדרום ברדיוס $k+1$ סביב האינדקס i.

הציעו אלגוריתם שמקבל כקלט מחרוזת $S[1..n]$, ומוצא בסיבוכיות זמן $O(n)$ את כל האינדקסים סביבם יש פלינדרום מקסימלי, ואת רדיוסי הפלינדרומים המקסימליים האלה.

<u>שאלה 16</u>

הראו מבנה נתונים דינאמי שעונה על שאילתות מהסוג: בהנתן נקודה q, כמה קטעים מכילים את q?
יש לעמוד בדרישות הסיבוכיות הבאות:

בניה: $O(n \log n)$.
שאילתא, הכנסה, הוצאה: $O(\log n)$.
זכרון: $O(n)$.

<u>שאלה 17</u>

נדגים את הטכניקה של fractional cascading על הבעיה הבאה.

> נתונות קבוצות של מספרים $S_m \subseteq S_{m-1} \subseteq \cdots \subseteq S_1$ ורוצים לענות שאילתות מהסוג: בהנתן תחום $[a,b]$, מהם איברי S_i, לכל i, הנמצאים בתחום $[a,b]$?

<u>פתרון נאיבי:</u> בשלב הבניה נמיין כל S_i במערך. כאשר תגיע שאילתא $[a,b]$ נבצע בכל אחד מהמערכים חיפוש בינרי של a ונוציא לפלט את איברי S_i החל מהאיבר שמצאנו ועד שנגיע לאיבר שגדול מ־b.

סיבוכיות שאילתא: $\mathrm{O}(\sum_{i=1}^{m} \log |S_i| + k)$ (k הוא אורך הפלט).

סיבוכיות זו עלולה להיות גרועה משמעותית מ־$\mathrm{O}(\log |S_1| + k)$. **למשל, עבור** $m = \log |S_1|$ **ו־**$|S_{i+1}| = \frac{1}{2}|S_i|$ **נקבל** $\mathrm{O}((\log |S_1|)^2 + k)$.

<u>פתרון בעזרת fractional cascading:</u> הרעיון הוא לנצל את העובדה שהקבוצות מוכלות זו בזו בכדי לחסוך את כל החיפושים הבינריים, פרט לראשון. בשלב הבניה נמיין כל S_i במערך. בנוסף, לכל איבר $s \in S_i$ נצמיד מצביע לעותק של עצמו ב־S_{i+1}. אם $s \notin S_{i+1}$ המצביע יהיה לאיבר הבא בגודלו אחרי s ב־S_{i+1}. במילים אחרות, $s \in S_i$ מצביע לאיבר $t \in S_{i+1}$ המינימלי כך ש־$s \le t$, או `null` אם אין כזה. כאשר מגיעה שאילתא $[a,b]$ נחפש את a במערך של S_1. נסמן ב־a_1 את האיבר שנמצא. יהי a_2 האיבר המוצבע על ידי a_1, ובאופן כללי, יהי a_{i+1} האיבר המוצבע על ידי a_i. לכל i, נתחיל באיבר a_i ונתקדם במערך המתאים עד שנגיע לאיבר הגדול מ־b (או לסוף המערך).

סיבוכיות שאילתא: $\mathrm{O}(\log |S_1| + k)$.

הציעו מבנה נתונים סטטי עבור בעית החיפוש הדו מימדית (שנפתרת על ידי range trees).
יש לעמוד בדרישות הסיבוכיות הבאות:

זמן בניה: $\mathrm{O}(n \log n)$.
זמן שאילתא: $\mathrm{O}(\log n + k)$.
זכרון: $\mathrm{O}(n \log n)$.

<u>שאלה 18</u>

הראו סכימה חצי־משתמרת (partially persistence) לבניית פונקצית ערבול מושלת חצי־דינאמית (semi-dynamic), אין הוצאות ויש הכנסות רק לגרסא האחרונה. כלומר הפעולות הן:

Insert(x) הוסף את x לגרסא האחרונה של הטבלה, זמן משוערך O(1).

H(x,t) החזר את ערכה של פונקצית הערבול שהייתה קיימת אחרי ההכנסה ה־t זמן משעורך O(log(n)) עבור טבלה בת n איברים.

סיבוכיות מקום O(n).