t

باسمه تعالی

دانشگاه تهران - دانشکده مهندسی برق و کامپیوتر

ساختمان‌داده‌ها و الگوریتم‌ها

تمرین چهارم - مرتب‌سازی و هش

محمدامین یوسفی، فرشته باقری

تاریخ تحویل: ۱۴۰۴/۰۹/۲۵

۱. رشته با کمترین مقدار — ۱۵ نمره

فرض کنید مجموعه‌ی $S = \{s_1, s_2, \ldots, s_n\}$ شامل n رشته از حروف کوچک انگلیسی داده شده است. طول طولانی‌ترین رشته در این مجموعه را با L نمایش می‌دهیم ($L = \max_i |s_i|$). هدف، یافتن یک جایگشت π روی اندیس‌های $\{1, \ldots, n\}$ است تا رشته‌ی حاصل از چسباندن آن‌ها، یعنی $T = s_{\pi(1)} s_{\pi(2)} \ldots s_{\pi(n)}$، از نظر لغت‌نامه‌ای (Lexicographically) کمینه باشد.

۱. **طراحی الگوریتم:** الگوریتمی کارا ارائه دهید که مسئله‌ی فوق را در زمان $O(n \cdot L \cdot \log n)$ حل کند. دقیقاً توضیح دهید که معیار شما برای تعیین ترتیب قرارگیری رشته‌ها چیست.

۲. **اثبات صحت:** ثابت کنید که الگوریتم پیشنهادی شما همواره پاسخ بهینه (کمینه) را تولید می‌کند. به بیان دقیق‌تر، نشان دهید که اگر خروجی الگوریتم شما رشته‌ی T_{opt} باشد، برای هر جایگشت دلخواه دیگر که منجر به رشته‌ی T' شود، رابطه‌ی $T_{opt} \leq_{lex} T'$ برقرار است.

۲. عنصر اکثریت — ۱۵ نمره

فرض کنید آرایه‌ی A شامل n عدد صحیح $a_1, a_2, \ldots, a_n$ و یک عدد صحیح x (به عنوان target) داده شده است. یک زیرآرایه‌ی پیوسته از A را با اندیس‌های شروع و پایان l و r مشخص می‌کنیم ($1 \leq l \leq r \leq n$).

می‌خواهیم تعداد زیرآرایه‌هایی را بیابیم که در آن‌ها عدد x، «عنصر اکثریت» باشد. عدد x در زیرآرایه‌ی $A[l \ldots r]$ عنصر اکثریت است اگر و تنها اگر رابطه‌ی زیر برقرار باشد:

$$count(x, A[l \ldots r]) > \frac{r - l + 1}{2}$$

که در آن $count$ تعداد دفعات تکرار x در بازه‌ی مذکور است.

۱. الگوریتمی با پیچیدگی زمانی $O(n)$ طراحی کنید که تعداد این زیرآرایه‌ها را محاسبه کند.

۲. صحت الگوریتم خود را اثبات کنید.

۳. بادکنک‌ها — ۲۰ نمره

مجموعه‌ای از n بادکنک دراز داریم که هر کدام به‌صورت یک بازه‌ی عددی $[x_i, y_i]$ نمایش داده می‌شود. هر بادکنک در بازه‌ی بین x_i تا y_i روی محور اعداد قرار دارد. با پرتاب یک تیر در نقطه‌ی p (روی محور اعداد)، تمام بادکنک‌هایی که شرط زیر را ارضا کنند، می‌ترکند:

$$x_i \leq p \leq y_i$$

روشی برای یافتن کمترین تعداد تیری که کل بادکنک‌ها را بترکاند بیابید.

۴. Linear Probing vs Double Hashing — ۱۵ نمره

یک جدول هش با اندازهٔ m = ۱۳ داریم و می‌خواهیم کلیدهای زیر را از چپ به راست در جدول قرار دهیم:

۹۶، ۸۳، ۷۰، ۵۷، ۴۴، ۳۱، ۱۸

(الف) برای هر یک از کلیدهای بالا، مراحل درج آن‌ها را با توجه به توابع هش زیر با هر دو روش Linear Probing و Double Hashing به‌صورت مرحله به مرحله نشان دهید و جدول نهایی را نیز رسم کنید. تابع هش در Linear Probing همان تابع اولیه Double Hashing است.

$$h_۱(k) = k \bmod ۱۳, \qquad h_۲(k) = ۱ + (k \bmod ۱۲)$$

(ب) برای هر روش، تعداد کل تصادم‌ها را حساب کنید.

(ج) فرض کنید اجازه داریم ترتیب درج مجموعه کلیدها را به دلخواه خود عوض کنیم. آیا برای این مجموعهٔ خاص می‌توانیم با عوض کردن ترتیب درج، تعداد تصادم‌ها را کم کنیم؟

۵. ذخیره‌سازی نقاط — ۲۰ نمره

می‌خواهیم مجموعه‌ای از نقاط را ذخیره کنیم که هر نقطه به‌صورت (x, y) است و :

$$۰ \le x < ۱۰۰۰, \qquad ۰ \le y < ۱۰۰۰$$

۱. یک تابع طراحی کنید که هر نقطهٔ (x, y) را به یک کلید عدد صحیح یکتا $K(x, y)$ تبدیل کند.

۲. اگر جدول هش m خانه داشته باشد، توابع هش اولیه و ثانویهٔ زیر را تعریف کنید:

$$h_۱(K) = ? \qquad\qquad h_۲(K) = ?$$

به‌طوری که:

- $h_۲(K)$ هیچ‌وقت صفر نباشد،
- و $h_۲(K)$ نسبت به m اول باشد.

۳. اگر تابع هش دوم نسبت به m اول نباشد چه مشکلاتی ممکن است پیش بیاید؟ و اگر $h_۲(K) = ۰$ شود، چه اتفاقی می‌افتد؟

۴. شبه‌کد برای توابع $\text{INSERT}(x, y)$ و $\text{SEARCH}(x, y)$ با استفاده از توابع $h_۱$ و $h_۲$ که در قسمت قبلی تعریف کرده‌اید بنویسید.

۶. احتمال تصادم — ۱۵ نمره

فرض کنید یک جدول هش با m خانه داریم و قرار است تعداد n کلید متمایز $k_۱, \ldots, k_n$ را در این جدول ذخیره کنیم. تابع هش h به‌گونه‌ای عمل می‌کند که برای هر کلید k_i، خانه‌ی مقصد $(h(k_i))$ را به‌صورت کاملاً تصادفی و مستقل از بین اندیس‌های $\{۰, \ldots, m-۱\}$ انتخاب می‌کند.

۱. با فرض اینکه $m = n^۲$ باشد، فرمول دقیق احتمال این‌که هیچ‌گونه برخوردی رخ ندهد را بنویسید. (برخورد زمانی رخ می‌دهد که دو کلید متفاوت به یک خانه ارجاع داده شوند).

۲. در حالت کلی (بدون فرض قسمت قبل)، مقدار n بر حسب m باید حداقل چقدر باشد تا احتمال وقوع «حداقل یک برخورد» بیشتر از ۵۰ درصد شود؟ (برای محاسبه‌ی ضرایب عددی دقیق، می‌توانید از ماشین‌حساب یا اینترنت استفاده کنید).