

רטוב 2 – חלק יבש

הצגת מבנה הנתונים

מבנה הנתונים שלנו מורכב בכלליות מטבלת ערבול של קורסים (בשיטת ה-chaining) ועץ דרגות של הרצאות להן זמן צפייה גדול מאפס.

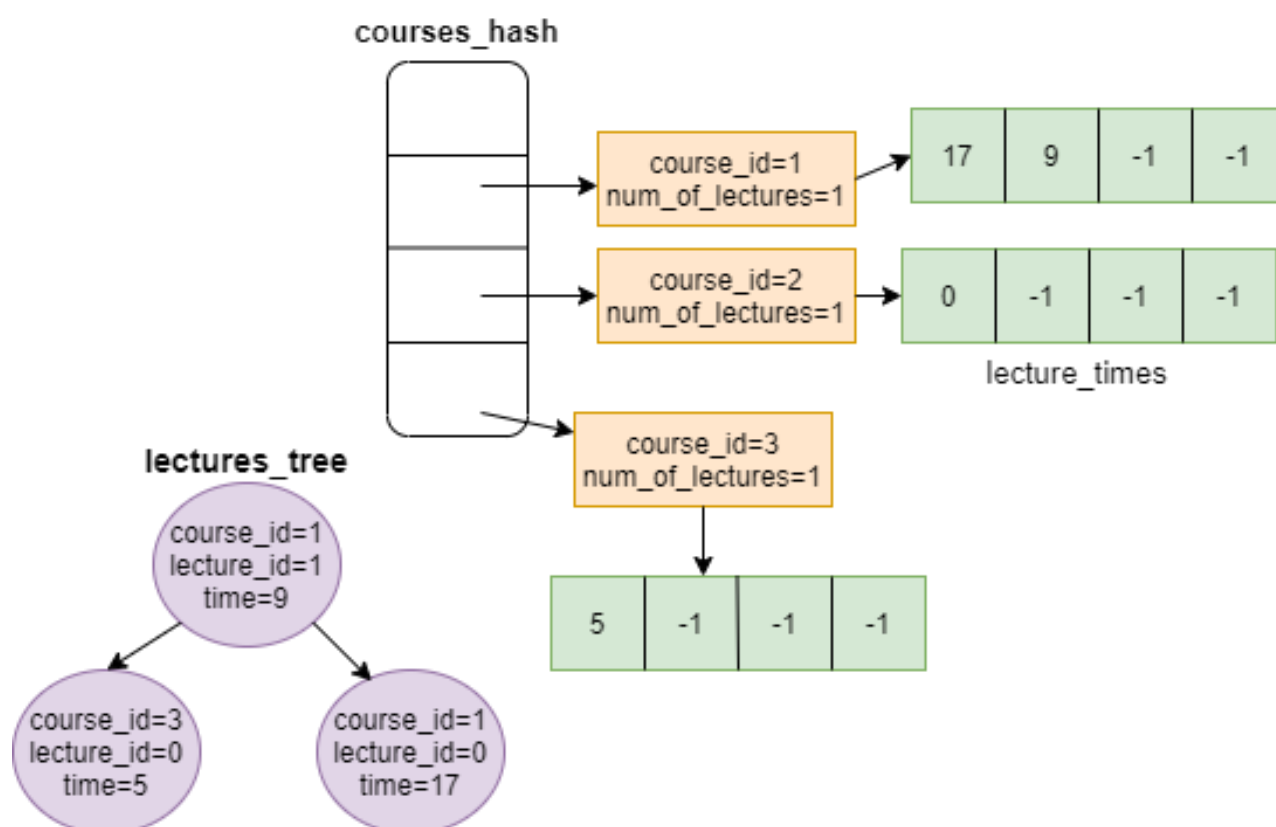
נתחיל מפירוט האובייקטים בהם השתמשנו לייצוג המידע במבנה הנתונים:

- Course – אובייקט המייצג קורס. כל אובייקט שכזה מחזיק במספר ID ייחודי, מספר ההרצאות עד כה, ומערך דינאמי של `lecture_times, int`, המייצג את זמני הצפייה של הרצאות הקורס.
- Lecture – אובייקט המייצג הרצאה. כל אובייקט שכזה מחזיק מספר ID של הקורס לו שייכת ההרצאה, מספר ID של ההרצאה (אשר ייחודי לה מבין הרצאות אותו הקורס), ואת זמן הצפייה בהרצאה זו.

כעת נוכל לפרט בהרחבה אודות מבני הנתונים העיקריים (שרטוט מצורף מטה):

1. טבלת ערבול של קורסים בשיטת ה-chaining – `courses_hash`: טבלת ערבול אשר האובייקטים בה הם קורסים (למעשה מערך דינאמי המכיל רשימות מקושרות של קורסים מכיוון שפעלנו בשיטת ה-chaining). כל קורס בטבלה מכיל בתוכו מערך דינאמי של זמני הצפייה בהרצאות של אותו הקורס.
2. עץ דרגות AVL של הרצאות בעלות זמן צפייה גדול מאפס – `lectures_tree`: ההשוואה בעץ נעשית כפי הנדרש בתרגיל, ראשית ע"פ זמן וכן הלאה. בכל צומת אובייקט Lecture.

שרטוט לדוגמא של מבנה הנתונים ברגע נתון:



- הערה 1: תא במערך הדינאמי `lecture_times`, בו מוזן הערך -1, משמעו שאין הרצאה עם מזהה id שכזה בקורס הנ"ל. לעומת זאת, תא בו מוזן הערך 0, מסמן שכן קיימת הרצאה עם המזהה id ולהרצאה זו זמן צפייה 0 עד כה.
- הערה 2: בעץ `lectures_tree` מוכנסים אך ורק הרצאות להן זמן צפייה חיובי.

פעולות על מבני הנתונים:

עץ דרגות: בעץ קיימות הפעולות הסטנדרטיות (הכנסה, הוצאה תוך כדי עדכון הדרגות) שנכתבו לפי האלגוריתם שניתן בהרצאות, ולכן סיבוכיות פעולות אלו הן $O(\log n)$. בנוסף, קיימת הפונקציה:

Select(i): מחזירה את האיבר שלו אינדקס i כאשר איברי העץ מסודרים בסדר יורד, ומחזירה NULL אם אין מספיק איברים כאלה.

סיבוכיות: הפעולה נכתבה לפי אלגוריתם בהרצאה עבור Select, וכפי שראינו סיבוכיות אלגוריתם זה היא $O(\log n)$.

רשימה מקושרת דו-כיוונית: ברשימה קיימות פעולות ההכנסה הסטנדרטיות (בתחילת הרשימה, בסוף הרשימה ואחרי איבר נתון) שנכתבו לפי האלגוריתם שניתן בהרצאות (שינוי המצביעים של האיבר הקודם והאיבר הבא אחרי האיבר שהוכנס), וכן פעולת הסרה (שגם בה מתבצע שינוי מצביעים בהתאם). כפי שראינו בהרצאה, מדובר במספר סופי של פעולות לכל הכנסה והוצאה ולכן סיבוכיות פעולות אלו הן $O(1)$.

בנוסף, קיימת פעולת הסרה לפי מידע. פעולה זו עוברת על כל איברי הרשימה, ומסירה את האיבר בו נמצא המידע. מכיוון שפעולה זו עלולה לעבור על פני כל הרשימה, הסיבוכיות שלה הוא $O(n)$.

מערך דינמי: במערך הדינמי קיימות פעולה עריכת מידע במקום ה-i כמו במערך רגיל, ולכן סיבוכיות פעולה זו היא $O(1)$.

בנוסף, המערך משנה את גודלו בהתאם למספר האיברים הקיימים במערך, כפי שנלמד בהרצאה (כאשר המערך מלא, מוקצה מערך חדש בגודל כפול). לכן, כפי שראינו בהרצאה, פעולת הכנסת איבר והגדלת המערך דורשות סיבוכיות זמן משוערכת של $O(1)$. פעולת הכנסת איבר מוגדרת כשינוי ערך של איבר שלא שונה מיצירת המערך (שווה לערך ההתחלתי של המערך).

אתחול המערך הדינמי: נאתחל מערך בגודל 4. לכן, סיבוכיות האתחול היא $O(1)$.

טבלת ערבול: טבלת הערבול ממומשת באמצעות פונקציית הערבול $h(x) = (x) \bmod n$, כאשר n הוא גודל המערך הנוכחי.

בטבלת הערבול קיימות פונקציית הכנסה, הוצאה ומציאה. כפי שנלמד בהרצאה, עבור פונקציות ההכנסה וההוצאה סיבוכיות הזמן המשוערך היא $O(1)$ בממוצע על קלט תחת הנחת הפיזור האחיד הפשוט, וכן עבור מציאה מדובר בסיבוכיות זמן של $O(1)$ בממוצע על קלט תחת הנחת הפיזור האחיד הפשוט.

את גודל המערך הדינמי נשנה כפי שנלמד בהרצאה: כאשר המערך מלא, מוקצה מערך חדש בגודל כפול ונעתיק את טבלת הערבול למערך זה, וכאשר המערך רבע מלא, נקצה מערך חדש בחצי הגודל ונעתיק את טבלת הערבול למערך זה.

אתחול טבלת הערבול: נאתחל את טבלת הערבול להיות בגודל 4. לכן, סיבוכיות האתחול היא $O(1)$.

הפעולות הנדרשות:

:Init

ניצור אובייקט חדש מסוג CoursesManager. ישנם 2 מבני נתונים באובייקט זה:

- עץ דרגות של הרצאות להן זמן גדול מאפס.
- טבלת ערבול של קורסים.

נדרש לאתחל כל מבנה נתונים. כל אתחול של מבנה נתונים, עץ וטבלת ערבול ריקה, לוקח $O(1)$, (שכן ללא איברים כעת) ולכן סך הכל סיבוכיות היא $O(1)$.

:AddCourse

ראשית נבדוק תקינות קלט כדרוש בתרגיל. לאחר מכן נבדוק האם הקורס קיים בטבלת ערבול הקורסים, ובמידה שכן נשמור מצביע אליו (סיבוכיות של $O(1)$ בממוצע על הקלט). במקרה הצורך נחזיר שגיאה מתאימה.

נכניס איבר חדש מטיפוס Course, עם המזהה course_id שלו לטבלת הערבול, המכיל 0 הרצאות (סיבוכיות משוערכת של $O(1)$ בממוצע על הקלט).

סיבוכיות כוללת של הפונקציה: פעולת מציאת הקורס דורשת סיבוכיות של $O(1)$ בממוצע על הקלט. הכנסת הקורס דורש סיבוכיות משוערכת של $O(1)$ בממוצע על הקלט. ולכן, סך הכל הסיבוכיות היא $O(1)$ בממוצע על הקלט, משוערך.

:RemoveCourse

ראשית נבדוק תקינות קלט כדרוש בתרגיל. לאחר מכן נבדוק אם קיים הקורס בטבלת ערבול הקורסים, ובמידה שכן נשמור מצביע אליו (סיבוכיות של $O(1)$ בממוצע על הקלט). במקרה הצורך נחזיר שגיאה מתאימה.

נרוץ על מערך זמני ההרצאות של קורס זה, ולכל הרצאה לה זמן צפייה גדול מ-0 (גישה לאיבר במערך דינמי ב- $O(1)$), נסיר אותה מעץ ההרצאות (כל הסרה לוקחת $O(\log M)$, כאשר M הוא מספר ההרצאות בכל הקורסים להם זמן צפייה גדול מאפס). לאחר מכן נסיר את הקורס מטבלת הערבול (סיבוכיות משוערכת של $O(1)$ בממוצע על הקלט).

סיבוכיות כוללת של הפונקציה: פעולת מציאת הקורס דורשת סיבוכיות של $O(1)$ בממוצע על הקלט. לאחר מכן, גישה למערך הדינמי של זמני ההרצאות לוקח $O(1)$ לכל הרצאה, והסרת הרצאה זו מהעץ דורשת $O(\log M)$, כאשר M הוא מספר האיברים בעץ, כלומר מספר ההרצאות הכולל להם זמן צפייה גדול מאפס (שחסום מלמעלה על ידי מספר ההרצאות הכולל). לכל היותר, יידרשו m הסרות כאלו, כאשר m הוא מספר ההרצאות של הקורס הנוכחי. לאחר מכן, הסרת הקורס מטבלת הערבול דורשת סיבוכיות משוערכת של $O(1)$ בממוצע על הקלט. לכן, סך הכל הסיבוכיות היא $O(m \cdot \log(M))$ בממוצע על הקלט. משוערך, כאשר m הוא מספר ההרצאות של הקורס הנוכחי ו- M הוא מספר ההרצאות הכולל במערכת בזמן הקריאה לפעולה.

AddClass

ראשית נבדוק תקינות קלט כדרוש בתרגיל. לאחר מכן נבדוק אם קיים הקורס בטבלת ערבול הקורסים, ובמידה שכן נשמור מצביע אליו (סיבוכיות של $O(1)$ בממוצע על הקלט). במקרה הצורך נחזיר שגיאה מתאימה.

ניגש למערך ההרצאות של הקורס, ונכניס במקום האחרון הפנוי במערך 0. את מקום זה ניתן לדעת מכיוון ששמרנו את מספר ההרצאות הנוכחי של כל קורס (הוספת איבר חדש למערך דינמי מבוצע בסיבוכיות משוערכת של $O(1)$).

לאחר ההכנסה נגדיל את מספר ההרצאות של הקורס הנוכחי ב-1 (מבוצע ב- $O(1)$).

נעדכן את הערך של $classID$ להיות מספר ההרצאה (האינדקס) שהכנסנו כעת.

סיבוכיות כוללת של הפונקציה: פעולת מציאת הקורס דורשת סיבוכיות של $O(1)$ בממוצע על הקלט. לאחר מכן, הכנסת איבר חדש למערך דינמי דורשת סיבוכיות משוערכת של $O(1)$, ושאר העדכונים דורשים $O(1)$. לכן, סך הכל הסיבוכיות היא סיבוכיות משוערכת של $O(1)$ בממוצע על הקלט.

WatchClass

ראשית נבדוק תקינות קלט כדרוש בתרגיל. לאחר מכן נבדוק אם קיים הקורס בטבלת ערבול הקורסים, ובמידה שכן נשמור מצביע אליו (סיבוכיות של $O(1)$ בממוצע על הקלט) והאם מספר הרצאה זה קיים בקורס (סיבוכיות של $O(1)$, מכיוון שאנחנו שומרים לכל קורס בטבלת הערבול את מספר ההרצאות הקיימות לו). במקרה הצורך נחזיר שגיאה מתאימה.

ניגש למערך ההרצאות של הקורס במקום ההרצאה המבוקשת ונבדוק את זמן הצפייה בו (גישה לאיבר במערך דינמי ב- $O(1)$). אם הוא גדול מאפס, נסיר את ההרצאה המתאימה מעץ ההרצאות ($O(\log(M))$, כאשר M הוא מספר ההרצאות בכל הקורסים להם זמן צפייה גדול מאפס). נעדכן את זמן הצפייה (באיבר המתאים במערך, גישה לאיבר במערך דינמי ב- $O(1)$), ניצור איבר חדש מטיפוס Lecture עם זמן הצפייה המעודכן ונכניס לעץ שלנו ($O(\log(M))$, כאשר M הוא מספר ההרצאות בכל הקורסים להם זמן צפייה גדול מאפס).

סיבוכיות כוללת של הפונקציה: פעולת מציאת הקורס דורשת סיבוכיות של $O(1)$ בממוצע על הקלט, ובדיקת תקינות ההרצאה דורשת סיבוכיות של $O(1)$. לאחר מכן, גישה למערך לוקחת $O(1)$, והסרה מהעץ לוקחת $\log(M)$, כאשר M הוא מספר האיברים בעץ, כלומר מספר ההרצאות הכולל להם זמן צפייה גדול מאפס (שחסום מלמעלה על ידי מספר ההרצאות הכולל). עדכון זמן הצפייה לוקח $O(1)$, והכנסה לעץ לוקחת $\log(M)$. לכן, סך הכל הסיבוכיות היא $O(\log(M))$ בממוצע על הקלט, כאשר M הוא מספר ההרצאות הכולל במערכת בזמן הקריאה לפעולה.

:TimeViewed

ראשית נבדוק תקינות קלט כדרוש בתרגיל. לאחר מכן נבדוק אם קיים הקורס בטבלת ערבול הקורסים, ובמידה שכן נשמור מצביע אליו (סיבוכיות של $O(1)$ בממוצע על הקלט) והאם מספר הרצאה זה קיים בקורס (סיבוכיות של $O(1)$, מכיוון שאנחנו שומרים לכל קורס בטבלת הערבול את מספר ההרצאות הקיימות לו). במקרה הצורך נחזיר שגיאה מתאימה.

ניגש לאיבר המתאים במערך ההרצאות של הקורס (גישה לאיבר במערך דינמי ב- $O(1)$), נעדכן את הערך של *timeViewed* להיות מספר זמן צפייה של איבר זה.

סיבוכיות כוללת של הפונקציה: פעולת מציאת הקורס דורשת סיבוכיות של $O(1)$ בממוצע על הקלט, ובדיקת תקינות ההרצאה דורשת סיבוכיות של $O(1)$. לאחר מכן, גישה למערך לוקחת $O(1)$, ועדכון ערך ההחזרה לוקח $O(1)$. לכן, סך הכל הסיבוכיות היא $O(1)$ בממוצע על הקלט.

:GetIthWatchedClass

ראשית נבדוק תקינות קלט כדרוש בתרגיל. במקרה הצורך נחזיר שגיאה מתאימה.

נקרא לפונקציה *Select* של העץ ההרצאות (סיבוכיות של $O(\log M)$, כאשר M הוא מספר ההרצאות בכל הקורסים להם זמן צפייה גדול מאפס). במידה ואין i הרצאות עם צפיות במבנה הנתונים שלנו (*Select* מחזיר NULL), נחזיר שגיאה מתאימה.

אחרת, נעדכן במשתנה *courseID* את מזהה הקורס של ההרצאה המבוקשת, ובמשתנה *ClassID* את ההרצאה המבוקשת.

סיבוכיות כוללת של הפונקציה: פונקציית *Select* דורשת $O(\log(M))$, כאשר M הוא מספר האיברים בעץ, כלומר מספר ההרצאות הכולל להם זמן צפייה גדול מאפס (שחסום מלמעלה על ידי מספר ההרצאות הכולל). עדכון ערכי ההחזרה לוקח $O(1)$. לכן, סך הכל הסיבוכיות היא $O(\log(M))$, כאשר M הוא מספר ההרצאות הכולל במערכת בזמן הקריאה לפעולה במקרה הגרוע.

:Quit

נמחק את *DS*, וכך יקרא ההורס של כל מבנה הנתונים. נציב ערך NULL ב-*DS*.

לכל קורס יש $\Theta(m)$ איברים, כאשר m הוא מספר ההרצאות של הקורס, כפי שראינו בהרצאה עבור מערך דינמי. לכן, בשחרור כל המערכים הדינמיים של הקורסים, מתבצעות $\Theta(M)$ פעולות, כאשר M הוא מספר ההרצאות הכולל.

בטבלת הערבול יש בכל רגע $\Theta(n)$ איברים, כאשר n הוא מספר הקורסים, כפי שראינו בתרגילים עבור טבלאות ערבול עם מערך דינמי. לכן, בשחרור הטבלה הדינמית, מתבצע $\Theta(n)$ פעולות, כאשר n הוא מספר ההרצאות הכולל.

בעץ ההרצאות יש לכל היותר M איברים, כאשר M הוא מספר ההרצאות הכולל. לכן, בהריסת העץ מתבצעות $O(M)$ פעולות מחיקה לכל היותר.

לכן, סך הכל הסיבוכיות היא:

$$O(M) + O(n) + O(M) = O(M) + O(n)$$

כלומר, הסיבוכיות היא $O(n + M)$, כאשר n הוא מספר הקורסים ו- M הוא מספר ההרצאות.

הוכחת סיבוכיות המקום של מבנה הנתונים:

- בעץ הדרגות של ההרצאות - lectures_tree - בכל רגע נתון ישנם לכל היותר m צמתים שכן בעץ תופיע כל הרצאה אשר לה זמן צפייה גדול מאפס, ותופיע פעם אחת בלבד. כלומר בעץ $O(m)$ הרצאות כאשר m מספר ההרצאות הכולל במערכת.
- נתבונן כעת בטבלת הערבול שלנו. בהרצאה ראינו כי בשימוש ב-chaining מתקבל הקשר: $\alpha = \frac{n}{m}$ כאשר α הינו פקטור העומס, קבוע כלשהו, n מספר האיברים בטבלה, ו- m גודל הטבלה. אזי $m = \frac{1}{\alpha} n$. כלומר מספר התאים בטבלה הוא $O(n)$, כאשר n מספר הקורסים במערכת.
נזכור כי כל תא בטבלה שלנו הוא למעשה רשימה של קורסים, ולכן עלינו כעת לקחת בחשבון את n הקורסים הללו בסכימה.
כפי שהסברנו על מבנה הנתונים שלנו, לכל קורס מערך דינאמי של הרצאות. מספר התאים במערך דינאמי חסום ע"י פי 2 ממספר האיברים בו (במקרה הקיצוני, מיד לאחר הגדלתו פי 2), ולכן נוכל לחסום את מספר התאים בכלל המערכים הדינאמיים של ההרצאות ע"י $2m =$ מספר ההרצאות $\cdot 2$, כאשר m מספר ההרצאות הכולל במערכת.

נסכום כעת את המקום המוקצה למערכת שלנו מכלל מבני הנתונים אשר בידנו:

$$\begin{aligned}
 & \underbrace{m}_{\text{lectures}_{\text{tree}}} + \underbrace{\frac{1}{\alpha} n}_{\substack{\text{cells in} \\ \text{courses}_{\text{hash}}}} + \underbrace{n}_{\substack{\text{courses in} \\ \text{courses}_{\text{hash}}}} + \underbrace{2m}_{\substack{\text{cells in} \\ \text{dynamic arrays} \\ \text{lecture times}}} \\
 &= \left(1 + \frac{1}{\alpha} \right) n + 3m = O(n + m)
 \end{aligned}$$

α is constant