פונקציות:

Insert:

Main function- insert:

הפונקציה תקבל key, value ותיצור IAVLnode.

תחילה הפונקצייה בודקת אם העץ ריק- אם כן מעדכנת את כל השדות (מינימום ומקסימום להיות הנקודה ומחזירה 0- לא היו סיבובים ועדכונים.

אחרת, נחפש את הצומת שאליה נכניס את הצומת החדש/ אם הצומת קיימת נחזיר אותה באמצעותplace to insert .

נחבר את הצומת החדש למקום המתאים ונבצע עדכוני מינימום ומקסימום.

נקרא לrebalance ולבסוף לupdate till root.

Place\_to\_insert:

טיול ימינה כל עוד המפתח גדול מהצומת החדש ושמאלה אם ההפך.

זאת נבצע כל עוד האיבר שאליו אנו רוצים ללכת אינו צומת פיקטיבית אם כן נחזיר את הצומת האחרונה שאינה פיקטיבית וגם אם הגענו למפתח של הצומת החדש (כבר קיים בעץ) נחזיר אותו.

Rebalance:

פונקצייה רקורסיבית שמאזנת את העץ לאחר הכנסה בהינתן צומת שממנו צריך להתחיל\ להימשך תהליך האיזון.

בהינתן קלט של צומת מסוים בעץ ומס' השינויים שנעשו בתהליך האיזון הנוכחי עד כה, הפונקציה בודקת את getrankright, getrankleft- הפרשי הדרגות (הגבהים) של הצומת הנתון ביחס לבניו ולפי זה מסווגת לאיזה מקרה איזון מחדש צריך לקרוא או האם הצומת מאוזן וישר לסיים את הריצה הרקורסיבית.

בחלק ממקרה הסיווג יש צורך הבחנה לתתי מקרים באופן הבא:

הפרש דרגות (0,1)/(1,0)- נבצע promotion ונקרא שוב לrebalace (בתנאי שזה לא הroot ואז אין צורך).

הפרש דרגות (0,2)+ הפרש דרגות של בן שמאלי (1,2) – סיבוב ימינה. נקרא לright rotation. ראינו כי פעולה זאת מביאה למצב תקין ומאוזן בעץ ולכן נסיים ונחזיר את מספר הפעולות שאספנו עד כה.

הפרש דרגות (0,2)+ הפרש דרגות של בן שמאלי (2,1) – סיבוב כפול – תחילה שמאלה של הבן השמאלי ואז ימינה של הצומת העליונה והבן השמאלי.

\*המקרים הסימטרים של (2,0) מטופלים באופן סימטרי בתת תנאי נוסף בפונקציה.

בסיום כל פעולה נקרא לupdateNode עבור הצמתים שיש לעדכן מהנמוך יותר בעץ אשר תלוי בצמתים שלא השתנו ולגבוה יותר- לכל היותר 3 עדכונים.

הפונקציה אוספת באופן רקורסיבי בעזרת משתנה מוחזר מטיפוס int אשר אוסף את מספר הפעולות -עבור כל rotation, promotion, demotion (promotion, demotion ייספרו באמצעות updateNode).

updateTillRoot:

הפונקציה מקבלת צומת שממנה צריך להתחיל לעדכן. בכל איטרציה הפונקציה בודקת האם יש לצומת אבת אם כן תעלה אליו ותעדכן אותו גםץ כלומר נבצע עדכונים מהצומת הנתון ועד לשורש העץ.

סיבוכיות כוללת של פעולת הinsert- בהתאם לכלל הפונקציות המשוייכות אליה:

נראה כי הסיבוכיות הינה logn:

Place\_to\_insert, במקרה הגרוע תרד משורש העץ אל עבר העלה הרחוק ביותר – כגודל גובה העץ-

Rebalance, כפי שהוכחנו בכיתה promotion עשוי לגרור את הבעיה מעלה ואילו סיבוב/ סיבוב כפול יובילו לreturn - כלומר ליציאה מהרקורסיה.

לאחר פעולת הpromote יש קריאה לrebalance עם parent- כלומר לכל היותר קריאות כגובה העץ.

בrotation\_right/left מבצעים ניתוקים וחיבורים קבועים- לכל היותר 6 ולכן סיבוכיות הזמן קבועה - .

בעדכון הצמתים נבצע טיפוס בעץ כגודל גובהו סיבוכיות O(logn).

סה"כ .

Join:

Join:

פונקציית מעטפת אשר מקבלת 2 עצים וקודקוד וממפה למצבים:

\*אם 2 העצים ריקים נקבע את x להיות השורש ונעדכן את שדות המינימום והמקסימום.

\*אם אחד העצים ריק ניצור קודקוד פיקטיבי ונשלח לפונקצייה הפנימית בהתאם לסדר האיברים (smaller, x, bigger).

נעדכן את שדות המינימום והמקסימום בהתאם.

\*2 העצים לא ריקים- נבדוק באיזה סדר לשלוח לפונקצייה הפנימית בהתאם ל(smaller, x, bigger).

נעדכן את שדות המינימום והמקסימום בהתאם.

בכל אחד מהשלבים נחזיר את הפרש הגבהים פלוס אחד כסיבוכיות זמן הריצה של join.

נשלח בכל פעם לjoin\_in את שלשת הקודקודים וjoin\_in מחזירה את הקודקוד המיועד להיות שורש העץ המעודכן. לכן נעדכן את שורש העץ להיות join\_in(..).

Join\_in:

הפונקציה מחברת בין קודקודים בהתאם לאלגוריתם join שלמדנו בכיתה, תוך שימוש בrebalance ובעדכון שדות הsize והhight של הקודקודים בקריאה לפונקצייה updateTillRoot . בנוסף, הפונקצייה תחזיר את הקודקוד המיועד להיות שורש העץ המעודכן.

נחלק למקרים:

אם העצים ריקים נחזיר את X.

אם גבהי העצים שווה, נחבר את איקס לשני הקודקודים בהתאם למיקום (קטן, גדול) ונחזיר את X.

אחרת נבדוק מי מבין הגבהים גדול יותר ונטייל עד לקודקודים שגובהו שווה לגובה העץ הקטן. נחבר את איקס אליו ונעדכן כלפי מעלה.

לבסוף, נקרא לפונקצייה find\_root אשר תעלה במעלה העץ עד לקודקוד שההורה שלו הוא null- זה יהיה שורש העץ החדש.

סיבוכיות כוללת של פעולת join- בהתאם לכלל הפונקציות המשוייכות אליה:

נראה כי הסיבוכיות הינה (|hight(t1)-hight(t2)|+1):

בjoin – פונקציית המעטפת ישנן פעולות אריתמטיות ועדכוני שדה ב עבודה.

נחקור את סיבוכיות פונקציית join\_in:

\*ירידה בעץ בעל הגובה הגדול יותר עד לגובה העץ הקטן- |hight(t1)-hight(t2)|.

*\*חיבורי X לקודקודים – פעולות קבועות בO(1).*

*\*הקריאה לrebalance מהאב של X – ראינו כי הפונקצייה הינה בסיבוכיות logn כאשר – כגובה העץ. כעת המעבר הינו על גובה של* |hight(t1)-hight(t2)|.

עדכוני שדות הקודקודים מX ועד לשורש העץ- מעבר על |hight(t1)-hight(t2)| קודקודים.

סה"כ נקבל סיבוכיות .(|hight(t1)-hight(t2)|+1

חלק ניסויי

שאלה 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| מספר סידורי | מספר חילופים במערך ממוין הפוך | עלות החיפושים במיון AVL עבור מערך ממוין הפוך | מספר חילופים במערך מסודר אקראית | עלות החיפושים במיון AVL עבור מערך מסודר אקראית |
| 1 | 1999000 | 38884 | 978828 | 31674 |
| 2 | 7998000 | 85764 | 4004900 | 73717 |
| 3 | 31996000 | 187524 | 15960205 | 161301 |
| 4 | 127992000 | 407044 | 64038517 | 361877 |
| 5 | 511984000 | 878084 | 255146028 | 762637 |

א.

ב.

נוכיח כי מספר החילופים:

במערך ממוין הפוך, לאיבר הראשון יש n-1 חילופים, לאיבר השני n-2 וכן הלאה. לאיבר האחרון אין חילופים כלל ולכן זוהי סדרה חשבונית כפי שהצגנו.

עלות החיפושים:

נוכיח כי עלות החיפושים הינה

נחשב באופן מדוייק את עלות החיפוש של האיבר הi:

בשלב זה יש בעץ i-1 איברים ולכן גובה העץ הינו . האיבר שנכניס הינו המינימלי ולכן בהינתן מצביע למקסימום נעלה במעלה העץ עד השורש ונרד עד הבן השמאלי ביותר. ולכן עלות החיפוש תהיה לכל היותר פעמיים גובה העץ.

נחסום מלמעלה:

נסכום את עלות כל החיפושים כפי שתיארנו למעלה ונקבל

כעת, נחסום מלמטה:

נתבונן בתת סדרה קשה – הכנסת n/2 האיברים האחרונים. באופן דומה, נסכום ונקבל:

בסה"כ קיבלנו שעלות סך החיפושים היא .

ג.

בנוסף, הוספנו את קו הריבועים הפחותים (בכחול) שמסמל את הקשר הליניארי בין ערכי ציר הX לערכי ציר הY, ואת מקדם המתאם – R^2 שמעיד על "איכות" קו הריבועים הפחותים, כלומר כמה הקשר "קרוב" לליניארי, ככל שמקדם המתאם קרוב יותר ל1 הקשר יותר הדוק.

נציג את התוצאות:

הפונקציה:

מספר חילופים במערך ממוין הפוך:

בסעיף ב מצאנו כי מספר החילופים הינו ריבועי בn. ואכן, ניתן לראות בגרף את הקשר הריבועי. בנוסף נבצע טרנספורמציית שורש לציר הY (מספר החיפושים) ונקבל קשר ליניארי.

Chart, line chart

Description automatically generated

the correlation R^2 is: 0.9491598820436215

Chart, line chart

Description automatically generatedובטרנספורמציית שורש של ציר הY:

the correlation R^2 is: 0.9999999999999996

עלות החיפושים במיון AVL עבור מערך ממוין הפוך:

בסעיף ב קיבלנו כי עלות החיפושים הוא nlog(n).

בנוסף, ראינו כי שיפוע קו הריבועים הפחותים (המייצג את הקשר הליניארי) הינו 27. נשים לב כי הפונקציה 27n הינה קירוב טוב ל2nlogn ואכן ניתן לראות זאת בגרף.

Chart, line chart

Description automatically generated

the correlation R^2 is: 0.9989568182473707

ד. בהינתן רשימה עם h חילופים ניתן לפרק אותם לסכום החילופים של כל איבר ואיבר.

כעת, נתבונן באיבר הi בעל h\_i חילופים:

ישנם h\_i איברים אשר גדולים ממנו ולכן עלינו לעלות מהאיבר המקסימלי עד שנגיע לצומת שתת העץ שלה מכיל h\_i איברים- תת עץ זה הוא עץ AVL ולכן הגובה של הינו log(h\_i). איבר זה ימוקם כאיבר השמאלי ביותר בתת עץ זה (המינימלי). לכן, עלות החיפש עבורו תהיה 2log(h\_i).

כלומר, בסה"כ עלות החיפושים תהיה:

כאשר אי השיוויון הראשון נובע מאי-שיוויון הממוצעים.

נשים לב שכאשר h<n נקבל למעשה שכן עלות החיפוש היא לפחות n שכן חיפוש מינימילי לכל איבר עולה לנו לפחות 1.

ה. נתבונן בהכנסה של מערך ממויין בסדר עולה אז כל חיפוש עולה לנו ולכן סה"כ נקבל

כלומר לא ניתן לתת חסם הדוק מכיוון שכפי שהראינו בסעיף הקודם ייתכן כי h=n^2 ואז נקבל כי עלות החיפוש היא .

שאלה 2:

ב. נוכיח כי עלות join ממוצע לשני התרחישים הינו amortized O(1).

נוכיח בשיטת הבנק:

תהי צומת במרחק i מהשורש.

נזכיר כי במהלך פעולת הjoin אנו מטפסים במעלה העץ עד השורש (i טיפוסים) .

בכל טיפוס, אנו מבצעים פעולת join של האיברים הקטנים (השמאליים) או הגדולים (הימניים).

נפעל באופן הבא: נצבור עבור כל טיפוס אסימון בבנק. אם עשינו join של האיברים בקטנים (טיפסנו שמאלה) אז האסימון שהפקדנו ישמש לjoin עתידי של האיברים הגדולים ולהיפך.

נוכיח כי עבור פעולת join יקרה יהיו מספיק אסימונים בבנק:

תהי פעולת join בעלות K. בה"כ פעולת join של איברים גדולים (ימנית).

על מנת להגיע להפרש גבהים k עלינו לעלות במעלה העץ k פעמים מבלי לעשות join ימני, כלומר לעלות k פעמים שמאלה. ולכן יש k אסימונים בבנק.

כמו כן, עבור כל פעולה שמאלה כזו שילמנו לכל היותר 2- הפרש גבהים של אב ובן.

לכן סה"כ שילמנו בכל פעולה לכל היותר 3 אסימונים, i פעמים וביצענו i פעמים join ולכן סה"כ:

באותו אופן, עבור האיבר המקסימלי בתת העץ השמאלי, ניתן לראות כי בתרחיש זה, נטפס שמאלה O(logn) פעמים ולכל טיפוס נצבור אסימון.

לבסוף יהיו בבנק logn אסימונים עבור פעולת הjoin היקרה של תת העץ הימני הריק של האיבר עם תת העץ הימני של השורש שגובהו logn. לכן גם במקרה זה נקבל זמן O(1).

ניתן להבחין כי הניתוח התיאורטי תואם לתוצאות שקיבלנו בסעיף א, שאכן הjoin הממוצע עלה בשני התרחישים בין 1-2. כלומר .

ג.