

30.6.2016

מבני נתונים
פתרון תרגיל מס' 11

1. שאלה זו הופיע במבחן מועד א 2015

בעקבות גלי החום שהיו בשנים האחרונות באירופה, הוחלט במשרד התירות הצרפתי, להקים מערכת מחשוב ששומרת נתונים לגבי לינות של תירים במקומות הארוז השונים בצרפת כפי שמתואר בהמשך.

לכל מקום ארוז שומרים: שם מקום הארוז (משמש לזיהוי מקום הארוז), כתובתו, האם בחדרי הארוז שבמקום יש מזגנים או לא, וכן נתונים על לינות של תירים במקום הארוז.

הנח/הניחי שעבור כל מקום ארוז, או שבכל חדרי הארוז שבמקום יש מזגנים או שבכל חדרי הארוז במקום אין מזגנים. (במילים אחרות, אין מקום ארוז שבו בחלק מהחדרים יש מזגנים ובחלק מהחדרים אין מזגנים).

לכל עיר/עירה שומרים: שם העיר/עירה (משמש לזיהוי העיר/עירה), שם האיזור בצרפת אליו שיכת העיר, וכן נתונים על הלינות של התירים במקומות הארוז בעיר.

הנח/הניחי שמספר האיזורים בצרפת הוא קבוע ועומד על 22.

להשכלה כללית: האיזורים בצרפת נקראים חבלים, למשל חבל אלזס, חבל לורן, חבל פרובנס, וכו'...

המשך בעמוד הבא...

הצע/הציעי מבנה נתונים עבור המערכת הנ"ל ששומר את הנתונים הנ"ל ותומך בפעולות הבאות:

- הוספת/הוצאת הנתון שתיר מסוים לן במקום ארוח מסוים בתאריך מסוים בזמן $O(\log x + \log z)$ בממוצע כאשר x מציין את מספר הערים/עירות בצרפת ו- z מתאר את מספר מקומות הארוח בצרפת.
- בהינתן חודש שנה ואיזור הדפסת סך כל הלינות שהיו בשנה זו בחודש זה באיזור זה, במקומות ארוח ללא מזגן, בזמן $O(1)$ בממוצע.
- בהינתן שם איזור, הדפסת כל הערים/עירות באיזור זה שלנו בהן תירים במקומות ארוח עם מזגן, ממוינות לפי המספר הכולל של לינות במקומות ארוח עם מזגן, בזמן $O(c)$ במקרה הגרוע, כאשר c מציין את מספר הערים/עירות ברשימה שתודפס. (שימו לב שעיר/עירה שלא היו בה לינות במקומות ארוח עם מזגן לא תופיע ברשימה).
- בהינתן עיר/עירה חודש ושנה הדפסת כל מקומות האירוח בעיר/עירה שאין בהם מזגנים ולנו בהם תירים בשנה זו בחודש זה, ממוינים לפי מספר הלינות שהיו במקומות אלה בשנה זו בחודש זה, בזמן $O(h)$ בממוצע, כאשר h מציין את מספר מקומות הארוח ברשימה שתודפס.
- בהינתן מספר k , הדפסת k הערים/עירות שמספר הלינות הכולל בהן במקומות אירוח בלי מזגן הוא הגדול ביותר בזמן $O(k)$ במקרה הגרוע.

בנוסף לתאור מבנה הנתונים שהצעת, תאר/י באופן מילולי איך מתבצעות שלושת הפעולות האחרונות.

פתרון

בהמשך בכל פעם שנכתוב עיר הכוונה לעיר/עירה (דהינו עיר או עירה).

המבנה המוצע מכיל את המבנים הבאים:

לכל איזור טבלת **hash** של חודש-שנה (מפתח: חודש-שנה) כל איבר בטבלה מכיל מספר שמצין את סך כל מספר הלינות שהיו באיזור זה בחודש זה בשנה זו במקומות ארוח ללא מזגן.

לכל איזור שני עצי 2-3 ראשי ומשני. העץ 2-3 הראשי (מפתח שם עיר) מכיל את כל הערים שלנו בהן תירים במקומות ארוח עם מזגן. העץ 2-3 המשני (מפתח מספר לינות) מכיל עבור כל עיר שבעץ הראשי מספר שמצין את סך כל מספר הלינות היו בעיר זו במקומות ארוח עם מזגן.

רשימה מקושרת של איזורים, כל איזור ברשימה מכיל מצביעים לעצי 2-3 הראשי והמשני שתוארו לעיל, מצביע לעלה השמאלי ביותר בעץ 2-3 המשני שתואר לעיל ומצביע לטבלת ה- **hash** של חודש-שנה עבור האיזור שתוארה לעיל.

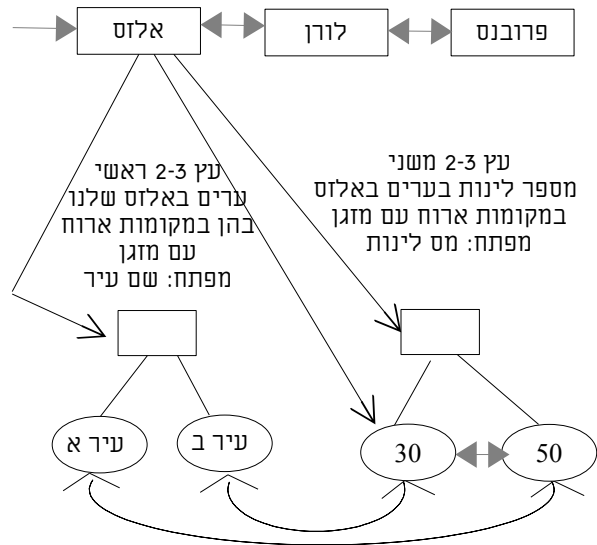
טבלת **hash** של מקומות ארוח (מפתח: שם מקום הארוח) כל איבר בטבלה מכיל את כתובת מקום הארוח, העיר והאיזור אליו הוא שיך והאם במקום הארוח יש מזגנים או לא.

טבלת **hash** של עיר-חודש-שנה (מפתח: שם עיר-חודש-שנה). כל איבר בטבלה מכיל מצביעים לשני עצי 2-3 ראשי ומשני, ומצביע לאיבר השמאלי ביותר בעץ 2-3 המשני. העץ 2-3 הראשי (מפתח: שם מקום ארוח) מכיל את מקומות הארוח ללא מזגן בעיר שלנו בהם תירים בעיר זו בשנה זו בחודש זה. העץ 2-3 המשני (מפתח: מספר לינות) מכיל עבור כל מקום ארוח שבעץ הראשי מספר שמצין את מספר הלינות שהיו במקום ארוח זה בחודש זה ובשנה זו. לכל איבר בעץ המשני ישנו מצביע דו כווני לאיבר המתאים לו בעץ הראשי. העלים בעץ המשני מחוברים על ידי רשימה מקושרת דו כוונית.

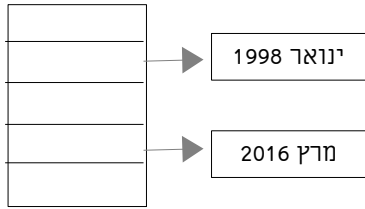
שני עצי 2-3 ראשי ומשני. העץ 2-3 הראשי (מפתח: שם עיר) מכיל את כל הערים. העץ 2-3 המשני (מפתח: מספר לינות במקומות ארוח ללא מזגן) מכיל עבור כל עיר שבעץ הראשי מספר שמצין את מספר הלינות שהיו בעיר זו במקומות ארוח ללא מזגן. לכל איבר בעץ המשני ישנו מצביע דו כווני לאיבר המתאים לו בעץ הראשי. העלים בעץ המשני מחוברים על ידי רשימה מקושרת דו כוונית. בנוסף ישנו מצביע לעלה הימני ביותר בעץ המשני.

להלן דוגמה לציור של מבנה הנתונים המוצע. (השדות של האיברים במבנה הנתונים, לא מפורטים בציור).

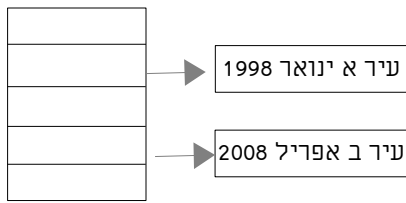
רשימה מקושרת של איזורים
מפתח: שם איזור



טבלת hash של איזור אלזס
מפתח: חודש-שנה

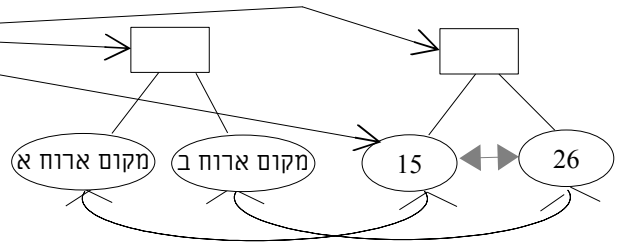


טבלת hash עיר-חודש-שנה
מפתח: עיר-חודש-שנה

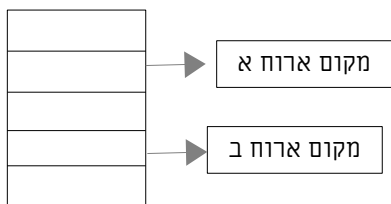


עץ 2-3 ראשי מקומות ארוח בעיר א ללא מזגן שלנו בהם תירים בחודש ינואר 1998 מפתח: שם מקום ארוח

עץ 2-3 משני מספר לינות במקומות הארוח שבעץ הראשי בחודש ינואר 1998 מפתח: מס לינות

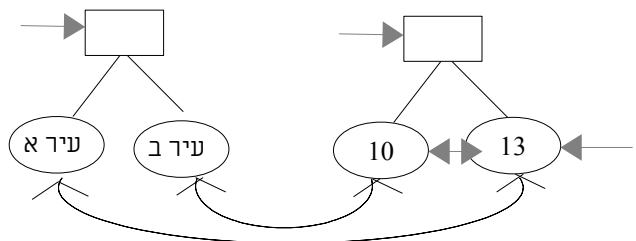


טבלת hash מקומות ארוח
מפתח: שם מקום הארוח



עץ 2-3 ראשי ערים מפתח: שם עיר

עץ 2-3 משני מספר לינות במקומות ארוח ללא מזגן בערים שבעץ הראשי מפתח: מס לינות



בהינתן שם איזור, הדפסת כל הערים/עירות באיזור זה שלנו בהן תירים במקומות ארוח עם מזגן, ממוינות לפי המספר הכולל של לינות במקומות ארוח עם מזגן, מתבצעת באופן הבא: מחפשים את האיזור ברשימה המקושרת של האיזורים ($O(1)$ במקרה הגרוע), בעזרת המצביע מגיעים לאיבר השמאלי בעץ 2-3 המשני, עוברים על האיברים בעץ 2-3 המשני משמאל לימין ומדפיסים את האיברים המתאימים להם בעץ 2-3 הראשי ($O(c)$ במקרה הגרוע). בסך הכל הפעולה לוקחת $O(c)$ במקרה הגרוע.

בהינתן עיר/עירה חודש ושנה הדפסת כל מקומות האירוח בעיר/עירה שאין בהם מזגנים ולנו בהם תירים בשנה זו בחודש זה, ממוינים לפי מספר הלינות שהיו במקומות אלה בשנה זו בחודש זה, מתבצעת באופן הבא: מחפשים את העיר-חודש-שנה בטבלת העיר-חודש-שנה, ($O(1)$ בממוצע), בעזרת המצביע מגיעים לאיבר השמאלי בעץ 2-3 המשני, עוברים על האיברים בעץ 2-3 המשני משמאל לימין ומדפיסים את האיברים המתאימים להם בעץ 2-3 הראשי ($O(h)$ במקרה הגרוע). בסך הכל הפעולה לוקחת $O(h)$ בממוצע.

בהינתן מספר k , הדפסת k הערים/עירות שמספר הלינות הכולל בהן במקומות אירוח בלי מזגן הוא הגדול ביותר, מתבצעת באופן הבא: בעזרת המצביע מגיעים לעלה הימני ביותר בעץ 2-3 המשני עוברים על k האיברים הימניים ביותר בעץ המשני ועבור כל איבר מדפיסים את האיבר המתאים לו בעץ 2-3 הראשי. בסך הכל הפעולה לוקחת $O(k)$ במקרה הגרוע.

2. שאלה זו הופיע במבחן מועד ב 2015

לקראת תחילת שנת הלימודים תשע"ו הוחלט במשרד החינוך להקים מערכת מחשוב ששומרת נתונים לגבי לימודי המתמטיקה בבתי הספר השונים בישראל כפי שמתואר בהמשך.

לכל שנה שומרים את נתוני התלמידים שסימו כיתה י"ב בשנה זו, כאשר לכל תלמיד שומרים: שם פרטי ושם משפחה (משמשים לזיהוי התלמיד), שם בית הספר בו הוא למד, האם סיים בגרות במתמטיקה, ובמידה וכן מה הציון שלו במתמטיקה ובכמה יחידות הוא נבחן. ניתן להניח שהציון הוא מספר שלם בין 0 ל-100 ומספר היחידות הוא מספר שלם בין 3 ל-5.

לכל בית ספר שם בית הספר משמש לזיהוי בית הספר.

אין להניח שמספר השנים שבהם שומרים נתונים על בית הספר הוא קבוע. ניתן להניח שמספר הציונים האפשריים הוא קבוע ושווה ל-101, וכן שמספר היחידות במתמטיקה האפשריות להבחן עליהן בבגרות הוא קבוע ושווה ל-3.

הערה: האופן בו נשמרים הנתונים הנ"ל אינו מפורט ויהיה עליך לצין אותו כחלק מפתרון השאלה.

המשך בעמוד הבא...

- הצע/הציעי מבנה נתונים עבור המערכת הנ"ל ששומר את הנתונים הנ"ל ותומך בפעולות הבאות:
- הוספת/הוצאת הציון שתלמיד מסוים שלומד בבית ספר מסוים קיבל בשנה מסוימת בבגרות במתמטיקה במספר יחידות מסוים בזמן $O(\log x + \log z)$ בממוצע כאשר x מצייין את מספר בתי הספר ו- z מתאר את מספר השנים שבהם שומרים נתונים במערכת על ציוני המתמטיקה בבתי הספר.
 - בהינתן שם בית ספר הדפסת נתונים על השנים שעבורן שומרים נתונים במערכת על בית הספר כאשר עבור כל שנה מודפס: אחוז התלמידים שסימו 3 יחידות מתמטיקה בשנה זו בבית ספר זה, אחוז התלמידים שסימו 4 יחידות מתמטיקה בשנה זו בבית ספר זה, ואחוז התלמידים שסימו 5 יחידות מתמטיקה בשנה זו בבית ספר זה. על הרשימה להיות ממוינת לפי אחוז התלמידים שסימו 5 יחידות מתמטיקה בסדר עולה. את הנתונים הנ"ל יש להדפיס בזמן $O(y)$ בממוצע כאשר y מצייין את מספר השנים ברשימה שתודפס.
 - בהינתן שנה, הדפסת בתי הספר ששומרים עבורם נתונים במערכת בשנה זו, כאשר עבור כל בית ספר מודפס: ממוצע הציונים של התלמידים שלמדו 3 יחידות מתמטיקה בשנה זו, ממוצע הציונים של התלמידים שלמדו 4 יחידות מתמטיקה בשנה זו וממוצע הציונים של התלמידים שלמדו 5 יחידות מתמטיקה בשנה זו. על הרשימה להיות ממוינת לפי ממוצע הציונים ב- 5 יחידות מתמטיקה בסדר עולה. את הנתונים הנ"ל יש להדפיס בזמן $O(s)$ בממוצע, כאשר s מצייין את מספר בתי הספר ברשימה שתודפס.
 - בהינתן שנה ומספר k הדפסת k התלמידים שקיבלו ציון גבוה ביותר ב- 5 יחידות מתמטיקה בשנה זו בזמן $O(n+k \log k)$ בממוצע כאשר n מצייין את מספר התלמידים שסימו 5 יחידות מתמטיקה בשנה זו.

בנוסף לתאור מבנה הנתונים שהצעת, תאר/י באופן מילולי איך מתבצעות שלושת הפעולות האחרונות.

המבנה המוצע מכיל את המבנים הבאים:

טבלת **hash** של בתי ספר (מפתח: שם בית ספר). כל איבר בטבלה מכיל מצביעים לשני עצי 2-3 ראשי ומשני, ומצביע לאיבר השמאלי ביותר בעץ 2-3 המשני. העץ 2-3 הראשי (מפתח: שנה) מכיל את השנים עבורן שומרים נתונים על בית ספר זה, עבור כל שנה שומרים: סך כל הסטודנטים שסימו כיתה י"ב בשנה זו בבית ספר זה, סך כל הסטודנטים שסימו 3 יחידות מתמטיקה בשנה זו בבית ספר זה, ובאופן דומה לגבי 4 ו-5 יחידות מתמטיקה. העץ 2-3 המשני (מפתח: אחוז מסימים 5 יחידות) מכיל עבור שנה שבעץ הראשי מספר שמציין את אחוז המסימים 5 יחידות מתמטיקה בבית ספר זה בשנה זו. לכל איבר בעץ המשני ישנו מצביע דו כווני לאיבר המתאים לו בעץ הראשי. העלים בעץ המשני מחוברים על ידי רשימה מקושרת דו כוונית.

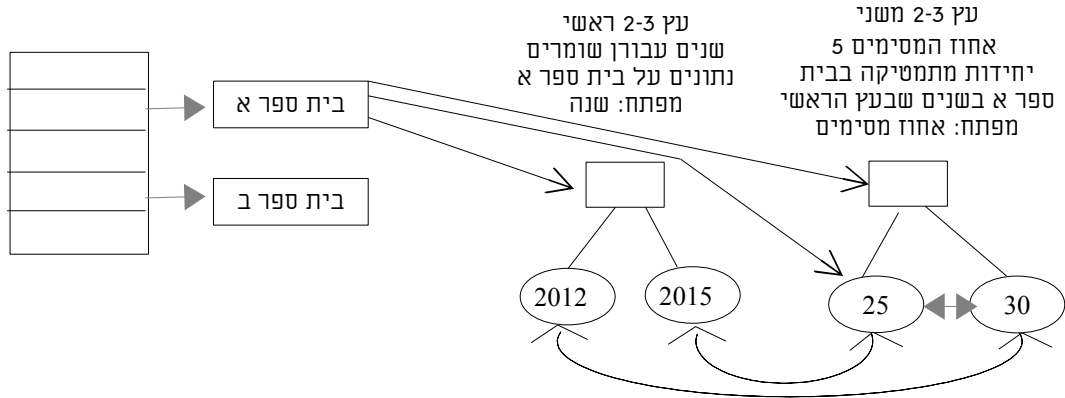
טבלת **hash** של שנים (מפתח: שנה). כל איבר בטבלה מכיל מצביעים לטבלת **hash** ולרשימה מקושרת של תלמידים באותה שנה (שמתוארים בהמשך), ולשני עצי 2-3 ראשי ומשני, ומצביע לאיבר השמאלי ביותר בעץ 2-3 המשני. העץ 2-3 הראשי (מפתח: שם בית ספר) מכיל את בתי הספר עבורן שומרים נתונים בשנה זו, עבור כל בית ספר שומרים: סך כל הסטודנטים שנבחנו ב-3 יחידות מתמטיקה בבית ספר זה בשנה זו, סך כל הציונים של כל התלמידים שנבחנו ב-3 יחידות מתמטיקה בבית ספר זה בשנה זו, ובאופן דומה שומרים נתונים לגבי 4 ו-5 יחידות מתמטיקה. העץ 2-3 המשני (מפתח: ממוצע ציונים ב-5 יחידות) מכיל עבור בית ספר שבעץ הראשי מספר שמציין את ממוצע הציונים ב-5 יחידות מתמטיקה של התלמידים שנבחנו ב-5 יחידות מתמטיקה בשנה זו בבית ספר זה. לכל איבר בעץ המשני ישנו מצביע דו כווני לאיבר המתאים לו בעץ הראשי. העלים בעץ המשני מחוברים על ידי רשימה מקושרת דו כוונית.

לכל שנה טבלת **hash** של תלמידים (מפתח: שם פרטי ושם משפחה) שנבחנו ב-5 יחידות מתמטיקה בשנה זו. כל איבר בטבלה מכיל את הציון ב-5 יחידות מתמטיקה של התלמיד בשנה זו, וכן מכיל מצביע לאיבר המתאים לו ברשימה המקושרת של תלמידים בשנה זו (שמתוארת בהמשך).

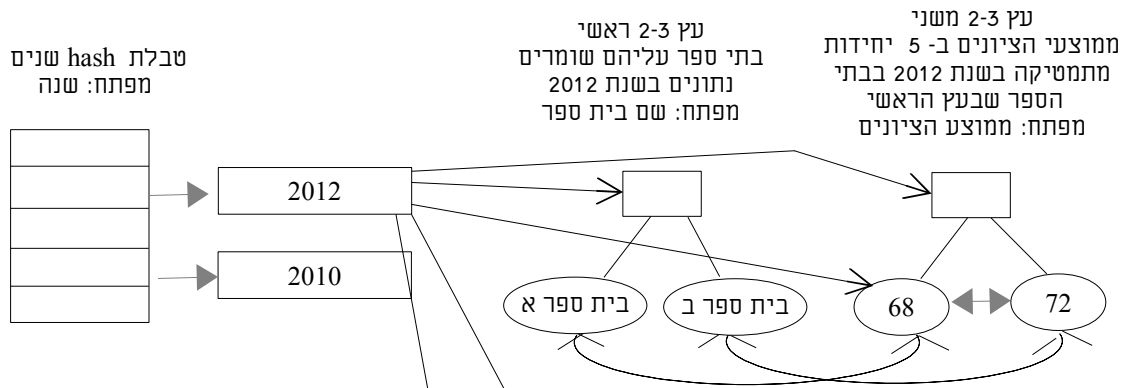
לכל שנה רשימה מקושרת של תלמידים (מפתח: שם פרטי ושם משפחה) שנבחנו ב-5 יחידות מתמטיקה בשנה זו. כל איבר ברשימה מכיל מצביע לאיבר המתאים לו בטבלת ה-**hash** של התלמידים בשנה זו שתוארה לעיל.

להלן דוגמה לציור של מבנה הנתונים המוצע. (השדות של האיברים במבנה הנתונים, לא מפורטים בציור).

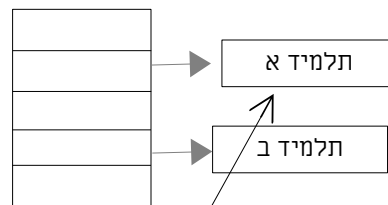
טבלת hash של בתי ספר
 מפתח: שם בית ספר



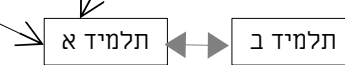
טבלת hash שנים
 מפתח: שנה



טבלת hash תלמידים בשנת 2012
 מפתח: שם תלמיד



רשימה מקושרת של תלמידים בשנת 2012
 מפתח: שם תלמיד



בהינתן שם בית ספר הדפסת נתונים על השנים שעבורן שומרים נתונים במערכת על בית הספר כאשר עבור כל שנה מודפס: אחוז התלמידים שסימו 3 יחידות מתמטיקה בשנה זו בבית ספר זה, אחוז התלמידים שסימו 4 יחידות מתמטיקה בשנה זו בבית ספר זה, ואחוז התלמידים שסימו 5 יחידות מתמטיקה בשנה זו בבית ספר זה, מתבצעת באופן הבא: מחפשים את בית הספר בטבלת בתי הספר ($O(1)$ בממוצע), בעזרת המצביע מגיעים לאיבר השמאלי בעץ 2-3 המשני, עוברים על האיברים בעץ 2-3 המשני משמאל לימין ומדפיסים את האיברים המתאימים להם בעץ 2-3 הראשי כאשר עבור כל איבר בעץ הראשי מדפיסים את אחוז המסימים 3 יחידות מתמטיקה בבית ספר זה בשנה זו שמתקבל מתוך חישוב החלוקה של מספר המסימים 3 יחידות מתמטיקה במספר התלמידים שסימו בשנה זו. באופן דומה לגבי 4 ו- 5 יחידות מתמטיקה. פעולה זו לוקחת $O(y)$ במקרה הגרוע. בסך הכל הפעולה לוקחת $O(y)$ בממוצע.

בהינתן שנה, הדפסת בתי הספר ששומרים עבורם נתונים במערכת בשנה זו, כאשר עבור כל בית ספר מודפס: ממוצע הציונים של התלמידים שלמדו 3 יחידות מתמטיקה בשנה זו, ממוצע הציונים של התלמידים שלמדו 4 יחידות מתמטיקה בשנה זו וממוצע הציונים של התלמידים שלמדו 5 יחידות מתמטיקה בשנה זו, מתבצעת באופן הבא: מחפשים את השנה בטבלת השנים ($O(1)$ בממוצע), בעזרת המצביע מגיעים לאיבר השמאלי בעץ 2-3 המשני, עוברים על האיברים בעץ 2-3 המשני משמאל לימין ומדפיסים את האיברים המתאימים להם בעץ 2-3 הראשי כאשר עבור כל איבר בעץ הראשי מדפיסים את ממוצע הציונים של התלמידים שלמדו 3 יחידות מתמטיקה בשנה זו בבית ספר זה שמתקבל מתוך חישוב החלוקה של סך כל הציונים של הנבחנים ב- 3 יחידות מתמטיקה בשנה זו בבית ספר זה בבסך כל התלמידים שנבחנו ב- 3 יחידות מתמטיקה בשנה זו בבית ספר זה. באופן דומה לגבי 4 ו- 5 יחידות מתמטיקה. פעולה זו לוקחת $O(s)$ במקרה הגרוע. בסך הכל הפעולה לוקחת $O(s)$ בממוצע.

בהינתן שנה ומספר k הדפסת k התלמידים שקיבלו ציון גבוה ביותר ב- 5 יחידות מתמטיקה בשנה זו מתבצעת באופן הבא: מחפשים את השנה בטבלת השנים ($O(1)$ בממוצע), בעזרת המצביע מגיעים לרשימה המקושרת של התלמידים שנבחנו ב- 5 יחידות מתמטיקה בשנה זו. עוברים על הרשימה המקושרת של התלמידים ועבור כל תלמיד מוצאים את הציון של התלמיד ב- 5 יחידות מתמטיקה בעזרת המצביע לאיבר המתאים לו בטבלת ה- hash ומוסיפים את הציון שלו למערך עזר. את המערך הזה הופכים לערימה ומדפיסים את k האיברים הגדולים ביותר בערימה. הפעולה הזו לוקחת $O(n+k \log k)$ במקרה הגרוע. בסך הכל הפעולה לוקחת $O(n+k \log k)$ בממוצע.