***דוח תרגיל 3 ביולוגיה חישובית***

מגישים:

ליאב טרבלסי 315870345 מדעי המחשב

ישי שלזינגר 208438119 ביולוגיה חישובית

על מנת להריץ את הקוד, צורף קובץ exe בגיט נפרד בקישור הזה

לאחר הרצת התוכנית, היא תבקש את הpath של הקובץ אותו מעוניינים לסווג, ואז התוכנית תתחיל לרוץ.

התוכנית בשלב הראשון עושה shuffle על הקלט ואז מתחילה לפתור.

כמבוקש – הפלט הוא רשת הsom ורשימת הערים בתוך כל תא ברשת :

A picture containing outdoor object, honeycomb

Description automatically generated

Text

Description automatically generated

(רשימת הערים ארוכה מידי ולכן צורפה פה בחלקה אך מודפסת במלואה בטרמינל בתוכנית עצמה)

כל תא מייצג כמו שנאמר כמה יישובים, והצבע בכל תא מייצג את הממוצע של הרמה הסוציואקונומית של כל היישובים בכל תא, לפי גווני האדום ( ככל שבהיר יותר הממוצע נמוך יותר, השתמשתי ב5 גוונים שחולקו בדרך הבאה – ממוצע 1-2 קיבלנו את הבהיר ביותר, 3-4 את הבא אחריו, וכן הלאה עד 9-10 אשר קיבלו את הגוון הכהה ביותר) ואפור בהיר לתא אשר לא מופו אליו יישובים כלל.

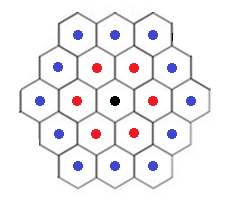
תשובות לשאלות התרגיל:

1. חישוב המרחק בין יישוב לתא המייצג אותו נעשה בדרך הבאה – כל תא אותחל כוקטור עם מספרים רנדומלים בטווח בין 0 ל20,000 כאשר הוקטור באורך של הנתונים של כל יישוב.

לאחר מכן, ע"י חישוב מרחק אוקלידי חישבנו את המרחק של כל יישוב מכל תא והתא עם המרחק הקטן ביותר נבחר כתא המייצג

1. קירוב התאים – לאחר שמצאנו את התא הקרוב ביותר, שלחנו אותו לפונקצית עדכון אשר מעדכנת בדרך הבאה – כאמור כל יישוב וכל תא מיוצגים כוקטור, ולכן כדי לעדכן עברנו ערך ערך בוקטור (הערך במקום ה0 של היישוב מול הערך במקום ה0 של התא וכן הלאה) חישבנו את המרחק בניהם (ע"י חיסור הערך) ולאחר מכן עדכנו את התא ב80% מהמרחק בניהם, לדוגמא אם הערך בתא היה 970 והערך ביישוב היה 870 – המרחק הוא 100 ולכן הערך בתא מעודכן במינוס 80 ל 890 בכדי לקרב אותו.

עדכון השכנים נעשה בדרך הבאה – הסתכלנו על 2 שלבים של שכנים



כאשר השחור זה התא המשוייך, האדום אלו השכנים בדרגה 1 והכחול זו הדרגה השניה

והעדכון נעשה באותה צורה כמו התא המרכזי רק במקום 80% דרגה 1 מעודכים ב40% ודרגה 2 מעודכנים ב10%.

1. בנוגע לסדר היישובים האם הוא משנה את איכות הפתרון, אז ערבבנו את הקלט לפני תחילת התהליך בצורה רנדומלית והדפסנו את התוצאות –

Example1 - [3424.724449977265, 4012.0447015674245, 3895.6627416415627, 3822.720529214361, 3691.579742232339, 3765.356334870936, 3703.0056654781924, 3891.210753242286, 3542.220990884155, 3565.132361736061]

Example2 - [4050.5611672961777, 3915.215431706132, 3632.5940334106813, 3678.6218385356888, 3825.729985628213, 4366.903988149832, 3770.175303076245, 3953.2126056089014, 4296.121570669498, 4069.9481486896107]

Example3 - [4044.2829122732533, 4185.229666493361, 4334.304537130717, 3585.6917397502943, 3944.792539350189, 3856.0210469060157, 3689.775180343835, 3932.2457552788237, 3684.508778789272, 4399.896709021773]

A picture containing outdoor object, honeycomb

Description automatically generated

אלו רק 3 דוגמאות מריצות רבות, אך ניתן ללמוד מהן כי למרות שהתהליך מקבל כל פעם סדר שונה של יישובים, איכות הפיתרון נשמרת ( גם בכללי בפתרונות וגם בטובה ביותר)

אך, כאשר שולחים את היישובים לפי סדר עולה של המצב הסוציואקונומית – מקבלים פתרונות פחות טובים –

Example1 - [4205.361334012519, 4920.660962363141, 4568.8781863441845, 4436.760070480248, 4357.879307013296, 4623.920309746263, 4396.77933847497, 4072.099225696401, 4291.931302022757, 4019.036809623843]

Example2 - [4458.773199837134, 4879.4868166139995, 4341.945538824098, 4865.715358868174, 4389.109506403357, 5418.904416370397, 4632.924273116313, 4195.002980197409, 4649.984461312482, 5049.779670875242]

Example3 - [4976.2754331966435, 4663.970196693939, 4420.327962199516, 4642.386758525646, 5120.374449782657, 4547.264612985652, 4597.0677459293065, 5414.861410198593, 4627.499694101472, 4008.305538930316]

A picture containing outdoor object, honeycomb

Description automatically generated

ההכי טוב שקיבלנו בסדר יורד הוא 4008 אך גם רוב הפתרונות באיזור ה5000

לעומת הכי טוב שקיבלנו היה 3424 ורוב הפתרונות באיזור ה4000. וגם ניתן לראות כי הפיזור פחות מסודר.

דבר זה קורה כיוון שכאשר המידע מגיע ממויין העדכונים הרבה יותר מקומיים ולכן פחות משמעותיים והפתרון פחות משתפר

1. כמו שצויין בתרגיל, חזרנו על כל התהליך 10 פעמים, בסיום כל ריצה חישבנו את המרחק הממוצע של כל היישובים מהתאים אליהם הם משוייכים, ושמרנו אותו, בסוף 10 הריצות, בדקנו מבין הממוצעים איזה ממוצע היה מהמינימלי וזה הפיתרון שאותו לקחנו כמוצג בגרף הבא –

Chart, bar chart

Description automatically generated

כפי שניתן לראות – הגרף מציג את המרחק הממוצע בסוף כל ריצה, והריצה עם המרחק המינימלי מסומנת לנו בירוק והיא הפיתרון שאותו אנו מציגים.

וכפי שניתן לראות במערך המצורף באמת הערך הנמוך ביותר היה באיטרציה התשיעית ולכן הוא נבחר-

[3753.6419078929966, 3738.440590932115, 4061.182079570375, 4050.4680640667043, 3841.4623597841405, 3876.6004523501233, 4005.6015938591913, 4120.214491511398, 3577.641150456943, 4119.837849859118]