

מימוש בסיסי נתונים עבודה 2

בעבודה זאת נעסוק בשליפות נתונים מקובץ על ידי Bitmap. ב-Bitmap עבור כל אחד מהערכים של כל עמודה מאונדקס וקטור בינארי אשר 1 מסמל שלרשומה יש את הערך ו-0 אם אין לה.

gender
m
m
f
f
m

דוגמה- עבור העמודה gender הבאה -

עבור הערך "m" הווקטור הבינארי אשר ייצג אותה יהיה- 11001 ועבור הערך "f" הווקטור הבינארי אשר ייצג אותה יהיה- 00110.

לאחר האינדוקס כאשר נרצה לבצע שליפה, למשל כל העובדים ממין זכר- נוכל לשלוף את כל הרשומות אשר בווקטור של "m" יהיה 1.

כאשר נרצה לבצע שליפה בעלת כמה תנאים, עם AND או OR בניהם נצטרך בהתאמה לעשות זאת עבור הווקטורים הרלוונטיים (למשל AND עבור שני הווקטורים – 110, 001 ייתן 000, כלומר אף רשומה לא ענתה על שני התנאים).

בתרגיל זה יהיה עליכם לממש 4 פונקציות כך שכאשר נריץ אותן אחת אחרי השנייה נוכל לקבל את המפתחות של הרשומות אשר ענו על תנאי השאילתה.

הגשה-

מצורף לכם פרויקט בו עליכם לממש את כל השיטות הנמצאות במחלקה Assignment2. מצורף לכם קובץ CSV בשם Employees_data לדוגמה כך שתוכלו לבדוק את עצמכם. שימו לב כי בדיקות העבודה יתבצעו באמצעות מערכת לבדיקת עבודות, כאשר העבודות שלכם יבדקו על קובץ CSV אחר כך שהעבודה שלכם צריכה לעבוד על כל קובץ בפורמט זה.

את העבודות יש להגיש מהמשתמשים של שני חברי הצוות למערכת ההגשה בכתובת <https://subsys.ise.bgu.ac.il/submission/login.aspx> תחת תיקיית Databases תחת Assignment2.

ההגשה הינה תיקייה מכווצת של הפרויקט אשר ניתן לכם לאחר שמישתם את פונקציות המחלקה Assignment2. **שימו לב לכווץ את אותה התיקיה אשר הורדתם מהמודל!** שימו לב שה-main שלכם יידרס. הנכם רשאים להעלות את התרגיל כמה פעמים שתמצאו, אך שימו לב שרק ההרצה האחרונה היא זאת שתישמר ותקבע את הציון. כאשר ההגשה שלכם רצה ללא שגיאות אנא אל תריצו אותה שוב עקב בעיות קודמות שקיימות במערכת הבדיקות.

להלן הפונקציות אשר עליכן לממש:

- `public void Index(string DBFilePath, string vectorFilePath)`

פונקציה זו מקבלת את הנתיב לקובץ הנתונים (DBFilePath) ואת הנתיב לקובץ אשר אליו עליכם לרשום את כל הווקטורים הבינאריים עבור כל ערך של כל עמודה (חוץ מעמודת id). אתם רשאים לכתוב את הווקטורים בכל פורמט שאתם מעוניינים.

דוגמה: עבור העמודה gender נדרש לכתוב שני ווקטורים (אחד עבור כל ערך אפשרי – f,m). דוגמה לאופן בו וקטור יכול להירשם לקובץ –

עבור הערך "m" של העמודה gender הקובץ יכיל את השורה-

gender,m,11001

- `public List<string> SelectVectors(XmlDocument xmlDoc, string vectorFilePath)`

הפונקציה הבאה מקבלת קובץ XML אשר יכיל את מבנה השאילתה אותה נרצה לבצע ואת הנתיב לקובץ אשר בו שמרתם את הווקטורים הבינאריים מהפונקציה הקודמת.

הפונקציה מחזירה את רשימת הווקטורים הבינאריים עליהם נצטרך לבצע את האופרטור הלוגי עבור ביצוע השאילתה. קובץ ה-XML אשר מתקבל בפונקציה יהיה בפורמט הבא:

```
<DB_EX2_QUERY>
  <Logical_Operation>AND</Logical_Operation>
  <Query_Elements>
    <Element column_Name="gender">
      <Value>m</Value>
      <Value>f</Value>
    </Element>
    <Element column_Name="Marital Status">
      <Value>married</Value>
    </Element>
  </Query_Elements>
</DB_EX2_QUERY>
```

כאשר כל תגית מסוג <Element> מייצגת עמודה (תנאי) וכל תת תגית מסוג <Value> מייצגת את הערך עבור העמודה.

תגית מסוג <Logical_Operation> מייצגת את סוג האופרטור הלוגי אותו נרצה ליישם בין התנאים השונים (יכול להיות AND או OR).

שימו לב שיכול להיות עבור אותה עמודה כמה ערכים שונים אותם נרצה (בדוגמה- עבור עמודה gender נרצה את הערכים "m" ו-"f") כך שבניהם תמיד נשתמש באופרטור OR כי כמובן ששניהם יחדיו לעולם לא יתקיימו.

למען הסר ספק השאילתה המיוצגת ע"י ה-XML הנ"ל הינה-

```
Select id from employees where (gender="m" OR gender="f") AND material status="married"
```

כאשר הווקטורים הבינאריים אותם הפונקציה תהיה צריכה להחזיר (לשלוף מתוך קובץ הווקטורים) הינם הווקטורים **הסופיים** שעליהם נבצע את האופרטור הלוגי (בדוגמה- AND). במילים אחרות, הפונקציה תחזיר ווקטור עבור כל תנאי.

עבור הדוגמה הנ"ל הפונקציה תחזיר שני וקטורים-

1. עבור עמודה gender - **וקטור יחיד** עבור הערכים "f" ו-"m" (וקטור הנוצר ע"י פעולת OR בין שני הווקטורים של שני הערכים אשר נשלפו מהקובץ- עבור "m"-00110 ועבור "f"-11001 נחזיר את הווקטור הבא ע"י הפעלת אופרטור OR-11111)
 2. עבור העמודה material status נשלוף מהקובץ את הווקטור עבור הערך "married".
- => סך הכל הפונקציה תחזיר **שני** וקטורים עבור שני התנאים הקיימים בשאילתה

שימו לב שיכולות להיות שאילתות בעלות תנאי אחד בלבד (Element בודד) כך שלא יהיה צורך בתגית <Logical_Operation>.

הנחות על הקלט-

כפי שלמדנו XML יכול להיות דינמי, ולכן אתם לא יכולים לעשות הנחות על הקלט. יכול להיות שלא יהיו תגיות element (תגית Query_Elements לא תכיל כלום). במקרה כזה אין תנאים לשאילתה- תצטרכו להחזיר את כל המפתחות.

הקובץ תמיד יכיל תגית <Logical_Operation> **אחת** בלבד או כלל לא- אם אין תנאים מרובים (element) לא תהיה גם תגית <Logical_Operation>.

הקובץ תמיד יכיל את התגיות DB_EX2_QUERY ו-Query_Elements.

קובץ ה-CSV אשר עליו תיבדקו יכול להיות בעל עמודות אחרות וערכים אחרים (חוץ מעמודת id אשר תהיה הראשונה בקובץ).

אתם יכולים להניח כי השאילתות יבוצעו על עמודות וערכים אשר נמצאים בקובץ.

מכיוון שאתם לא צריכים ליצור וקטורים בינאריים עבור עמודת id אתם יכולים להניח כי לא תצטרכו לתמוך בשאילתה שבתנאים שלה מתייחסים ל-id. השימוש בעמודה זו יהיה רק עבור ה-select.

- `public string CreateOutputVector(XmlDocument xmlDoc, List<string> vectors)`

פונקציה זו מקבלת רשימה של וקטורים (שהחזרנו בפונקציה הקודמת) ואת קובץ ה-XML ויוצרת את וקטור הפלט עבור השאילתה. כלומר, עליכם להוציא מתוך קובץ ה-XML את האופרטור מתגית <Logical_Operation> (בדוגמה- AND), לבצע את הפעולה על רשימת הווקטורים vectors ולהחזיר את וקטור הפלט שנוצר.

- `public List<string> SelectRecords(string DBFilePath, string outputVector)`

הפונקציה הבאה מקבלת את הנתיב לקובץ הנתונים ואת וקטור הפלט הסופי עבור השאילתה ומחזירה רשימה של המפתחות (id) של הרשומות אשר ענו על השאילתה.

למשל עבור הווקטור 10101 הרשימה המוחזרת צריכה להכיל את מפתחות 1,3,5.

בהצלחה!