

תרגיל בית מס' 4

נושאים: MongoDB, XML, Neo4j

מתרגלת אחראית: דיאנה כהן

- כדאי מאוד להסתייע במצגות התרגולים.
- לשאלות יש לפנות לפורום הייעודי **במודל**.
- ההגשה **בזוגות בלבד**, **מוקלדת** בפורמט PDF, עד לתאריך 30/06/2022 כולל.
- הגשות באיחור בהתאם לנהלי הקורס, המופיעים במצגת האדמיניסטרציה.
- שאלות אדמיניסטרטביות יש להפנות לדניאל ליברמן (מתרגל אחראי של הקורס).

להגשה:

ת.ו.

פירוט החלקים והניקוד:

שאלה	נושא	ניקוד	הערות
1	MongoDB	35	
2	XML	35	
3	Neo4j	35	
	סה"כ	100	ציון מירבי בתרגיל בית כולו לא יעלה על 100

בהצלחה!

שאלה 1 – MongoDB (35 נק')

על השאלה יש לענות בהיבט של MongoDB, לא לכתוב שאילתות SQL במענה על השאלה.

במהלך התרגולים חזרנו מדי פעם לדוגמא של הספרייה. לצורך השאלה הזו, נתמקד בטבלת Books שהיא אחת הטבלאות המרכזיות בספרייה (אין הרבה משמעות לספרייה ללא ספרים).

תזכורת מהתרגולים:

Books(Book_Id, Book_Name, Year, Max_Time, Pages, Faculty)

Book_Id – מזהה ייחודי של העותק (עותקים שונים של אותו ספר מקבלים מזהה שונה)

Book_Name – שם הספר

Year – שנת הדפסה

Max_Time – מספר ימים מירבי המותר להשאלת העותק

Pages – מספר הדפים

Faculty – שם הפקולטה

תוכן הטבלה:

<u>Book_Id</u>	Book_Name	Year	Max_Time	Pages	Faculty
1111	Database Systems	1998	7	348	CS
1112	Database Systems	1998	14	348	CS
1113	Database Systems	2001	7	424	CS
2222	Database And Knowledge	1998	1	390	CS
2223	Database And Knowledge	1998	7	390	EE
3333	Electronic Circuits	1998	21	180	EE
4444	Genes 7	1985	7	580	MED
5555	Anatomy	1988	7	450	MED

באשר ל-MongoDB: הניחו כי קיים Collection בשם **books**, המכיל Documents כך שכל מסמך מתאר ספר בודד. הגישה לספרים תיעשה ע"י התחילית **db.books**.

(1) (5 נק') תרגמו את אחת משורות הטבלה למסמך אפשרי.

(2) (8 נק') בפקולטה CS יש עניין לעדכן את זמן ההשאלה המירבי לשבוע ימים. כתבו שאילתת עדכון מתאימה.

(3) (10 נק') כתבו שאילתא שמחשבת את כמות הספרים לכל פקולטה.

(4) (12 נק') השתמשו ב-MapReduce כדי להחזיר לכל שם-ספר (`book_name`) את מספר הדפים הממוצע (`AvgPages`), אבל יש להתחשב רק בעותקים החל משנת 1990.

שאלה 2 – XML (35 נק')

גם בשאלה זו נתמקד בטבלת הספרים שהוצגה בשאלה 1.
תרגום הטבלה למסמך XML נותן מבנה שטוח יחסית.
אחת האפשרויות למסמך כזה, המכיל את הרשומה הראשונה מהטבלה:

```
<books>
  <book book_id="1111">
    <book_name>Database Systems</book_name>
    <year>1998</year>
    <max_time>7</max_time>
    <pages>348</pages>
    <faculty>CS</faculty>
  </book>
</books>
```

לצורך המרה לעץ, היעזרו בשלבים הבאים:

- ספר יזוהה עם השם שלו (book_name). לכל ספר ייתכנו מספר מהדורות.
- מהדורה של ספר תזוהה עם שנת הדפסה (year). מספר הדפים (pages) נקבע לפי המהדורה. לכל מהדורה ייתכנו מספר פקולטות שמחזיקות בה.
- פקולטה (faculty) יכולה להחזיק במספר עותקים של אותה מהדורה (של אותו הספר).
- כל עותק מזוהה עם book_id. לכל עותק בפקולטה נקבע max_time.

(1) (10 נק') תרגמו את טבלת הספרים למבנה של עץ (בהתאם לשלבי ההמרה נ"ל), תנו שמות משמעותיים לצמתים חדשים.

- (2) (10 נק') כתבו מסמך XML התואם לעץ שהתקבל בסעיף (1), עבור 5 הרשומות הראשונות מהטבלה, כלומר רק עבור הספרים "Database Systems" ו-"Database And Knowledge". מלאו את המסמך עם הנתונים שמופיעים בטבלה המקורית. צרפו את תוכן מסמך ה-XML שהתקבל.

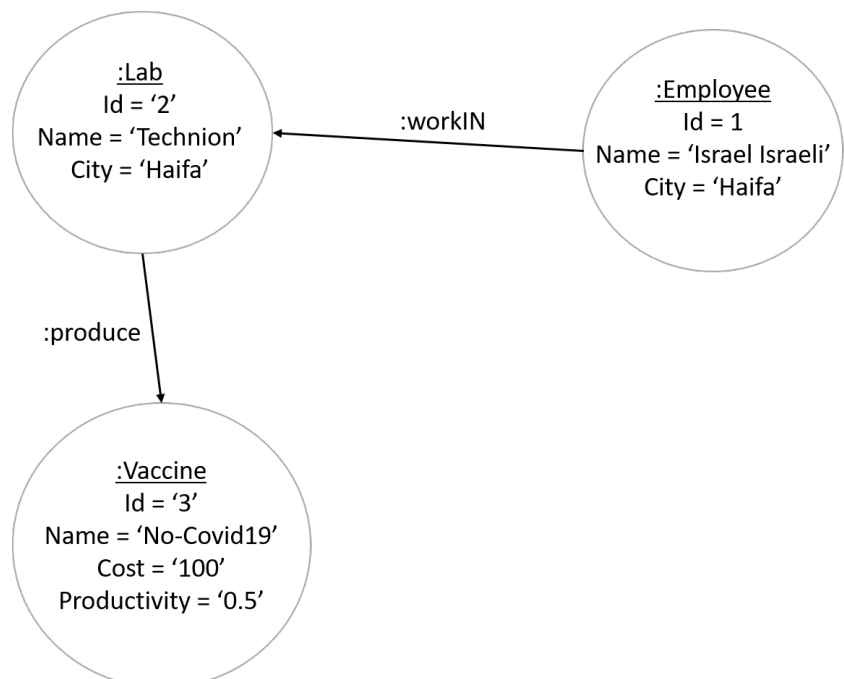
- (3) (15 נק') בהתאם למסמך ה-XML שהתקבל בסעיף הקודם, כתבו שאילתא שמחזירה את שמות הספרים שעבורם יש פקולטה שמחזיקה ביותר מעותק אחד (של אותה המהדורה).

שאלה 3 – Neo4j (35 נק')

נתון תיאור של גרף Neo4j המחזיק צמתים של עובדים, מעבדות ותרופות.
הגרף ממומש לפי הכללים הבאים:

- * כל צומת בגרף בתווית (label) אחת בדיוק מבין האפשרויות הבאות:
Employee, Lab, Vaccine
- * צומת מסוג Employee אשר מחזיק את התכונות הבאות:
Id – מזהה העובד.
Name – שם העובד.
City – העיר בה העובד גר.
- * צומת מסוג Lab אשר מחזיק את התכונות הבאות:
Id – מזהה המעבדה.
Name – שם המעבדה.
City – העיר בה המעבדה נמצאת.
- * צומת מסוג Vaccine אשר מחזיק את התכונות הבאות:
Id – מזהה התרופה.
Name – שם התרופה.
Cost – מחיר התרופה ב\$.
Productivity – יעילות התרופה (באחוזים).
- * צמתי עובדים יכולים להיות מקושרים לצומת מעבדה על ידי קשר בעל label מסוג WorkIN, אשר מסמל שהעובד עובד במעבדה אליה הוא מקושר.
עובד יכול לעבוד ביותר ממעבדה אחת.
- * צמתי מעבדה יכולים להיות מקושרים לצומת תרופה על ידי קשר בעל label מסוג Produce, אשר מסמל שהמעבדה מייצרת את התרופה אליה היא מקושרת.
מעבדה יכולה ליצור כמה תרופות.

דוגמא לגרף המקיים את הכללים:



- (1) (7 נק') כתבו שאילתה בשפת Cypher המחזירה את מזהה העובדים שעובדים בשתי מעבדות לפחות.

- (2) (8 נק') כתבו שאילתה בשפת Cypher המחזירה את שמות התרופות שכל המעבדות מייצרות והשם שלהן מתחיל ב-"Covid".

- (3) (10 נק') הממשלה החליטה לתת מענק לעובדים שרווח המעבדה בה הם עובדים קטן מ-\$15,000. הרווח מוגדר להיות סכום מחירי החיסונים שהמעבדה מייצרת. כתבו שאילתה בשפת Cypher המחזירה את שמות העובדים והעיר בה הם גרים אשר זכאים למענק.

- (4) (10 נק') כתבו שאילתה בשפת Cypher המחזירה את שמות העובדים כך שלמעבדה בה הם עובדים קיימת בדיוק תרופה אחת עם יעילות גבוהה מ-90% ויש לה שני עובדים לפחות.