

# מכללת הדסה, החוג למדעי המחשב

## אנליזה של ביג דאטה / חננאל פרל

סמסטר ב', תשפ"ד

תרגיל 6 / 24.05.2024

### תאריך הגשה:

יום ראשון 2.06.2024 בשעה 20:00

### מטרות התרגיל:

תרגול Neo4j – Graph DB

### הוראות:

נא לפתוח את סביבת המשחק "ארגז החול" של Neo4j

<https://sandbox.neo4j.com/>

בתוך הסביבה שנוצרה לפתוח את הפרויקט שנקרא: **Entity Resolution**

ואז נלחץ Open in Browser

שם אפשר לשחק בממשק ולהריץ את שאילתות.

ההגשה צריכה להיות של השאילתות בקובץ טקסט.

בנוסף יש להגיש קובץ וורד שמכיל את

- השאילתה כתובה בעימוד משמאל לימין בפונט **Consolas**
- צילום מסך של התוצאה בצורת גרף Graph כולל השורה העליונה שרואים השאילתה.
- צילום מסך של התוצאה כטבלה Table – כולל החלק התחתון שמראה כמה תוצאות היו.

דוגמא:

תשובה 1:

השאילתה:

**MATCH** (x) **RETURN** \* **LIMIT** 20

סה"כ ההגשה צריכה להכיל 2 קבצים:  
א. קובץ טקסט אחד (.txt) של השאילתות.  
ב. קובץ וורד אחד (.docx) עם השאילתות וצילומי המסך.

שאלה שאתם לא יודעים את התשובה עליה, עדיין לכתוב את הקו המפריד ומספר השאלה, אך להשאיר את השאילתה והתוצאה ריקים.  
השאלות לא בהכרח ממוינות לפי סדר קושי..

**בהצלחה!**

### 1) Entity Resolution:

- a. נא להחזיר (כגרף וכטבלה) את כל האנשים מאוהיו (Ohio) ואת הסרטים שצפו בהם שנוצרו אחרי 2010 ומכילים את המילה "and" (אותיות גדולות או קטנות או שילוב)
- b. נא להראות את רשימת הזאנרים בוורג'יניה (Virginia) ממיון מהכי נצפה שם עד הכי נמוך.
- c. נא להחזיר (כגרף וכטבלה) את כל הזוגות של משתמשים על IP זהה, שאינם נמצאים באותה מדינה.
- d. נא להראות (כגרף וכטבלה) את כל הקשרים באורך של 6, שיש בין האנשים ששמן הפרטי "Flo" לבין האנשים ששמן הפרטי "Cam" אך ורק קשרים מסוג WATCHED או HAS (ז"א שלא יופיעו (IpAddress

### 2) שאלה כללית – על הפרויקט: POLE Data Model - Crime Investigation

- לבחור את הפרויקט מהארגז חול - POLE Data Model - Crime Investigation
- יש לכתוב שתי (2) שאלות אינטלגנטיות ורלוונטיות על הדאטה.
- יש לכתוב בקובץ טקסט את השאלות שערונות על השאלות שיצרתם.
- את השאלות האלו יש לכתוב בעברית באותו קובץ וורד. ליד כל שאלה לרשום את השאלתה שעונה עליה.