מכללת הדסה, החוג למדעי המחשב

אנליזה של ביג דאטה / חננאל פרל

סמסטר ב', תשפ"ד

24.05.2024 / **6** תרגיל

תאריך הגשה:

יום ראשון 2.06.2024 בשעה 20:00

מטרות התרגיל:

Neo4j – Graph DB תרגול

:הוראות

נא לפתוח את סביבת המשחק "ארגז החול" של Neo4j

https://sandbox.neo4j.com/

בתוך הסביבה שנוצרה לפתוח את הפרויקט שנקרא: Entity Resolution

Open in Browser ואז נלחץ

שם אפשר לשחק בממשק ולהריץ את שאילתות.

ההגשה צריכה להיות של השאילתות בקובץ טקסט.

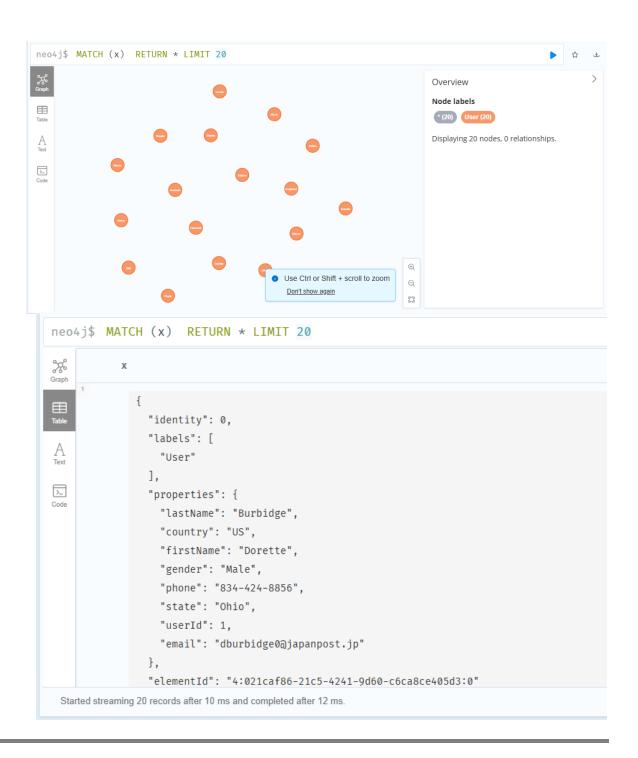
בנוסף יש להגיש קובץ וורד שמכיל את

- Consolas השאילתה כתובה בעימוד משמאל לימין בפונט
- 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:

:השאילתא



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!

השאלות NEO4J 6:

:Entity Resolution (1

- ואת הסרטים שצפו בהם שנוצרו אחרי (מגרף וכטבלה) את כל האנשים מאוהיו (Ohio) ואת הסרטים שצפו בהם שנוצרו אחרי .a 2010 ומכילים את המילה "and" (אותיות גדולות או קטנות או שילוב)
 - b. נא להראות את רשימת הזאנרים בווירג'יניה (Virginia) ממוין מהכי נצפה שם עד הכי נמוך.
- c. נא להחזיר (כגרף וכטבלה) את כל הזוגות של משתמשים על IP זהה, שאינם נמצאים באותה מדינה.
- "Flo" (כגרף וכטבלה) את כל הקשרים באורך של 6, שיש בין האנשים ששמם הפרטי .d או Cam (ז"א שלא יופיעו HAS או WATCHED לבין האנשים ששמם הפרטי "Cam" לבין האנשים ששמם הפרטי) (IpAddress
 - 2) שאלה כללית על הפרויקט: POLE Data Model Crime Investigation
 - POLE Data Model Crime Investigation לבחור את הפרויקט מהארגז חול
 - יש לכתוב שתי (2) שאלות אינטלגנטיות ורלוונטיות על הדאטה.
 - יש לכתוב בקובץ טקסט את השאילתות שעונות על השאלות שיצרתם.
 - את השאלות האלו יש לכתוב בעברית באותו קובץ וורד. ליד כל שאלה לרשום את השאילתה שעונה עליה.