

מטלה 3:

שימו לב! יש להעלות כל הגשה פעם אחת בלבד!

יש להעלות קובץ word שמכיל את כל התשובות ביחד.

על מנת למנוע אי נעימויות – הקפידו לכתוב את מספרי הזהות של כל השותפים

בתחילת הקובץ

אחראית על התרגיל – אביגיל שטקל Avigail.st@gmail.com

mongoDB

עבור collection של ספרים (בכל מסמך קיימים נתונים על הספר- שם, מחברים, הוצאה, שנת הוצאה, גיל מיועד, דירוג וכן הספר עצמו)

Book_name, authors, publisher, publishing_year, for_age, rank, the_book

1. כיתבו שאילתא ששולפת את כל הספרים שמתקיים אחד (לפחות אחד) מהתנאים הבאים:

- a. הסופרת היא "Ruth Cohen" וגם הדירוג הוא 3 לפחות
- b. הספר מיועד לילדים בני 10 ומעלה ונכתב לפני שנת 2000

2. כתוב פונקציית mapReduce המחזירה כמה מילים קיימים בכל הספרים ששנת ההוצאה אינה 2000 לכל אורך של מילה לדוגמא- אם הספר מכיל את הטקסט הבא-

SUPPOSING that Truth is a woman--what then?

נקבל את התוצאה הבאה:

```
{_id:"1", _value:"1"}, {_id:"2", _value:"1"}, {_id:"4", _value:"3"}, {_id:"5", _value:"2"},  
{_id:"9", _value:"1"}
```

:Neo4j

נתון גרף המתאר מבנה של חברת הייטק

בגרף ישנם צמתים מהסוגים הבאים:

worker המכיל את שם העובד, ת"ז, כתובת ומשכורת

manager המכיל את שם המנהל, ת"ז, כתובת ומשכורת.

department המכיל את שם המחלקה

בנוסף, קיימים הקשרים הבאים:

work_in בין עובד/מנהל למחלקות שבהן הוא עובד המכיל תאריך התחלת עבודה.

friend_of בין 2 עובדים או 2 מנהלים.

כתוב שאילתא ב neo4j המחזירה רשימה ממוינת של שמות כל העובדים בחברה (כולל מנהלים) מתל אביב שמשכורתם לפחות 19999 וכל החברים שלהם עד רמה 3 (חברים, חברים של חברים, חברים של חברים של חברים) מקבלים גם כן משכורת שהיא לפחות 19999

:elasticSearch

עבור type של ספרים, בכל מסמך קיימים נתונים על הספר- שם, מחברים, ז'אנר הספר, הוצאה, שנת הוצאה ותקציר הספר

1. כתוב פקודת הוספה של מסמך לindex 'ספרים'.
2. כתוב שאילתא שתחזיר את הספרים שהוצאו משנת 1999 והלאה ושז'אנר הספר הוא- 'מדע בדיוני' ושבתקציר הספר יש את המילים "Fiction Science" ו-"reality" -

:X-pathk

נניח שקיים לנו קובץ XML המכיל מדינות. כל מדינה (country) מכילה attribute שם וקודקודים בנים המייצגים את כל הערים שבה. כל עיר (city) מכילה בתור attribute את מספר התושבים שגרים בעיר זאת(num).

כתוב שאילתא X-path שתחזיר את שמות כל המדינות שסך כל התושבים בהם גדול ממליון.

:Stream

בהנתן טקסט נרצה לדעת כמה מילים יש מאותו אורך (כמו שעשיתם בשאלה 1 סעיף 2) אבל רק מילים באורך של בן 5 ל 9 אותיות.
כלומר אם יש 3 מילים באורך של 8 אותיות, שתי מילים באורך של 4 אותיות ו5 מילים באורך 2 הפלט יהיה
{8_letters=3, 4_letters=3}
כל הקוד צריך להיות בjavastream

:SPARQL RDF

המר את הטבלה הבאה לטבלת RDF ורשום שאילתא SPARQL שמחזירה את שמות כל הנכדים של ראובן.

ID	name	age	Father_ID
111	ישראל ישראלי	15	444
222	פלוני אלמוני	2	333
333	ג'ון סמית	30	444
444	ראובן אריאל	81	555

:TF-IDF

דרג את המשפטים הבאים לפי מדד TF-IDF:

Q: What a day

1. This is a nice day.
2. What do you eat?
3. What happened? This is not my day!
4. This is a day for swimming.
5. What can I ask? Give me a day.