

סמסטר אביב התשע"ז

מרצה: פרופ' בני קימלפלד

מתרגלים: שובל לגזיאל

דביר דוקאן

ליאת פיטרפרוינד

מערכות מסד נתונים

236363

מועד ב'

28 בספטמבר 2017

פירוט החלקים והניקוד:

חלק	נושא	מס' שאלות בחלק	מס' שאלות שיש לענות עליהן	ניקוד
1	תכן מסדי נתונים	2	2	30
2	שאליות במודל היחסים	2	2	34
3	מודלים לא יחסיים	4	3	36

**הנחיות לנבחנים**

1. כתבו את התשובות אך ורק בטופס הבחינה ובמקום המיועד להן. מחברת הטיוטה לא תיבדק.
2. כל חומר עזר הכתוב על נייר בלבד מותר בשימוש.
3. אין לקבל או להעביר חומר עזר כלשהו בזמן הבחינה.
4. **בבחינה בחירה בין שתי שאלות:** עליכם לענות על שאלה אחת מבין שאלות Neo4j או MongoDB, ציינו במפורש במה בחרתם, במקרה של סימון לא ברור התשובה הראשונה תיבדק.
5. יש להשתמש רק בסימנים או פונקציות שנלמדו בתרגול או בהרצאה, או שמופיעות בשקפים של הקורס. כל שימוש בסימון שאינו כזה מחייב הסבר מלא של משמעות הסימון.
6. משך הבחינה הינו שלוש שעות. תכננו את הזמן בהתאם.
7. בבחינה 8 שאלות בשלושה חלקים. נא וודאו שיש בידכם את כל הטופס.

בהצלחה!

**חלק 1- תכן מסדי נתונים – 30 נק' – כל השאלות בחלק זה הן חובה**

**1. צורות נורמליות 10 נק'**

נתונה הסכמה  $R(A_1, \dots, A_n)$

א. (5) לכל  $n > 2$  הראו דוגמא לזוג תלויות פונקציונליות מעל  $R$  כך שהסכמה המתקבלת הינה בצורה הנורמלית השלישית אך לא BCNF.

$$A_1, \dots, A_{n-1} \rightarrow A_n, A_2 \rightarrow A_1$$

ב. (5) הראו כי ל  $n=3$  לא ניתן לפתור את סעיף א' ע"י תלות אחת (במקום זוג תלויות).

There are two options, up to isomorphism: (a)  $A \rightarrow B$  and (b)  $AB \rightarrow C$ . For (a) the result is not 3NF, since A is not a key and B is not part of any admissible key: any key must contain A, hence not B. For (b) the result is BCNF.

## 2. תלויות פונקציונליות, 20 נק'

נתונה הסכמה  $R(A,B,C)$  וקבוצת התלויות  $F = \{A \rightarrow B, B \rightarrow C, BC \rightarrow A\}$

א. (5) הראה כיסוי מינימלי של  $F$  כפי שיכול להתקבל ע"י האלגוריתם שנלמד בכיתה.

$$A \rightarrow B, B \rightarrow C, B \rightarrow A$$

ב. (5) הראה כיסוי מינימלי של  $F$  המכיל את התלות  $A \rightarrow C$ . האם כיסוי זה יכול להתקבל ע"י האלגוריתם שנלמד בכיתה?

$A \rightarrow B, A \rightarrow C, B \rightarrow A$ . לא, כיוון שלא ניתן להגיע ל  $A \rightarrow C$ : אין תלות עם  $A$  בצד שמאל ו- $C$  בצד ימין.

ג. (10) הוכח כי בכל כיסוי מינימלי של  $F$ , כל תלות מכילה בדיוק תכונה אחת בכל צד.

צד ימין – תמיד נכון מהגדרת הכיסוי המינימלי. נניח בשלילה כי בכיסוי מינימלי כלשהו יש לנו תלות עם שני אטריביוטים בצד שמאל, נניח  $XY \rightarrow Z$ . התלות אינה טריוויאלית, ולכן נסתכל על שתי רשומות שמפרות את התלות: הן מסכימות על  $XY$  ולא על  $Z$ . מכאן, הן מפרות את אחת התלויות מהסעיפים הקודמים. אבל כזו הפרה יכולה להיות רק  $X \rightarrow Z$  או  $Y \rightarrow Z$  (כי למעלה כל הצדדים השמאליים בגודל אחד). מכאן נובע ש  $XY \rightarrow Z$  מכיל יתירות (אפשר לזרוק אטריביוט). ולכן, סתירה.

**חלק 2- שאלות מידע רלציוניות – 34 נק' - כל השאלות בחלק זה הן חובה**

3. SQL, RA, 20 נק'

בשירות הסרטים Netclicks רוצים לבנות מערכת המלצות הממליצה לצופה על סרטים שרלוונטיים עבורו.

נתונות הטבלאות הבאות

Costumers Table - CUST

<u>CustID</u>	Name	Age
---------------	------	-----

Movies Table – MOVIE

<u>MovieID</u>	Title	Genre
----------------	-------	-------

Views Table - VIEWED

<u>CustID</u>	<u>MovieID</u>	Rank
---------------	----------------	------

בטבלאות, VIEWED.CustID ו-VIEWED.MovieID הינם מפתחות זרים ל- CUST.CustID ו- MOVIES.MovieID, בהתאמה.

א. (5) כתבו שאלתת SQL **ללא קינון** המחזירה את הז'אנרים בעלי צפייה אחת לפחות באופן ממזין. המזין הוא בסדר יורד לפי הציון הממוצע של הסרטים בז'אנר.

```
Select movie.genre FROM
movie Join views
ON movie.movieid = viewed.movieid
Group By Genre
Order BY AVG(Rank) DESC
```

ב. (5) כתבו שאילתת SQL המחזירה, לכל ז'אנר, את שמות הסרטים בעלי הציון המקסימלי בז'אנר. התוצאה תכיל את האטריביוטים Genre ו-Title.

```
SELECT genre, title FROM
(movie JOIN viewed
ON movie.movieid = viewed.movieid) as b
where b.rank >= all
(SELECT rank FROM
movie JOIN views
ON movie.movieid = viewed.movieid
WHERE movie.genre = b.genre)
```

ג. (6) פתרו את סעיף ב' בעזרת RA (במקום SQL).

$$\begin{aligned}A &= movie \bowtie viewed \\ B &= \rho_{rank \rightarrow rank1}(A) \\ C &= \pi_{title, genre} \sigma_{(rank < rank1)} A \bowtie B \\ D &= \pi_{(title, genre)}(movie) \setminus C\end{aligned}$$

ד. (4) כתבו שאילתת RA המחזירה את שמות הסרטים שאף לקוח לא ראה.

$$\begin{aligned}A &= \pi_{(movieid)}(movie) \setminus \pi_{movieid}(viewed) \\ B &= \pi_{(title)} movie \bowtie A\end{aligned}$$

#### 4. RC , DATALOG , 14 נק'

בשירות הסרטים Netclicks רוצים לבנות מערכת המלצות הממליצה לצופה על סרטים שרלוונטיים עבורו.

הניחו כי קיימות במסד הנתונים הטבלאות מהשאלה הקודמת

עבור לקוח c נגדיר את ההגדרות הבאות:

סרט מוסכם בין שני לקוחות אם שניהם צפו בו ונתנו לו את הציון 5.

לקוח ממליץ ל-c - לקוח שאינו c וקיים לו ול-c סרט מוסכם .

סרט מומלץ ל-c - סרט ש-c לא ראה, אך קיים לקוח ממליץ שנתן לו את הציון 5.

ניתן להשתמש ביחסים  $R_{\neq}(x, y) = \{(x, y) | x \neq y\}$

א. (4) כתבו שאילתת RC המחזירה עבור כל לקוח c את כל הסרטים המומלצים לו.

$\{(c1, m2) | \exists c2, m [R_{\neq}(c1, c2) \wedge Viewed(c1, m, 5) \wedge Viewed(c2, m, 5) \wedge Viewed(c2, m2, 5) \wedge \neg \exists r Viewed(c1, m2, r)]\}$

ב. (5) נגדיר "ממליץ עקיף" ללקוח c באופן אינדוקטיבי, כלהלן.

a. כל ממליץ ל-c הוא גם ממליץ עקיף ל-c.

b. כל ממליץ לממליץ עקיף ל-c הוא גם ממליץ עקיף ל-c.

בנוסף, סרט מומלץ באופן עקיף ל-c - הינו סרט ש-c לא ראה, אך קיים ממליץ עקיף ל-c שנתן לסרט את הציון 5.

כתוב תכנית Datalog עם שלילה ברת ריבוד המחזירה, לכל לקוח, את כל הסרטים שאינם מומלצים באופן עקיף עבור הלקוח.

MovieID(m) :- Movie(m,t,g)

VM(c,m) :- Viewed(c,m,r)

Same(c,c) :- Cust(c,n,a)

Rec(c1,c2) :- Viewed(c1,m,5), Viewed(c1,m,5), Viewed(c2,m,5), !Same(c1,c2)

IndRec(c1,c2) :- Rec(c1,c2)

IndRec(c1,c2) :- IndRec(c1,c'), Rec(c',c2)

IndRecMov(c,m) :- IndRec(c,c2), Viewed(c2,m,5), ! VM(c,m)

Output(c,m) :- Cust(c,n,a), MovieID(m), ! IndRecMov(c,m)

ג. (2) הראו ריבוד של התכנית שכתבתם בסעיף הקודם.

Layer 0 (EDB's) = {Viewed, Movie}

Layer 1 = {same, VM}

Layer 2 = {Rec, IndRec, IndRecMov }

Layer 3 = { output}

ד. (3) האם ניתן היה לפתור את הסעיף הקודם בתוכנית ללא שלילות?

לא מכיוון שדאטה-לוג היא מונוטונית, כלומר הוספת מידע, תשאיר או תגדיל את סט הפתרונות של התוכנית. במקרה שלנו הוספת מידע יכולה לצמצם את הפתרון

ניקח שני משתמשים  $c_1, c_2$  שקיים ביניהם סרט מוסכם

נניח כי הפתרונות של התוכנית כלל בתוכו את הסרט  $\gamma$ , כיוון שמשתמש  $c_2$  לא אוהב את הסרט הזה

אם נוסיף רשומה שבה המשתמש  $c_2$  אוהב את סרט  $\gamma$

לכן אם נריץ שוב את השאילתה לאחר עדכון המסד נקבל שהסרט  $\gamma$  לא בסט הפתרונות

### חלק 3- מודלים לא רלציוניים – 36 נק' – בחלק זה שאלות 5 ו-6 הן חובה ועליכם לבחור בין שאלה 7 ל-8.

#### 5. XML , שאלת חובה 13 נק'

נתון מסמך ה DTD הבא:

```
<!DOCTYPE A [  
  <!ELEMENT A ((B+)|(E,F?))>  
  <!ATTLIST A ref IDREF #REQUIRED>  
  <!ELEMENT B EMPTY>  
  <!ATTLIST B serial CDATA #REQUIRED  
    ref CDATA #IMPLIED>  
  <!ELEMENT E (A)>  
  <!ELEMENT F EMPTY>  
  <!ATTLIST F serial ID #REQUIRED  
    ref IDREF #REQUIRED>  

```

א. (8) מהו מספר האלמנטים המינימלי במסמך XML שמציית לDTD הנתון. הצג מסמך XML עם מספר אלמנטים כזה והוכח כי זהו אכן מספר האלמנטים המינימלי.

**מספר אלמנטים לפי הגדרות הDTD:**

A מכיל בתוכו לפחות B אחד (ואין ב B שום אלמנט) או שיכיל E (وهוא מכיל A ואז מדובר ברקורסיה) ורשימת F (מגודל 0 ומעלה). A מכיל IDREF שהוא required והאלמנט היחיד במסמך שיש לו ID הוא F לכן למעשה המסמך חייב להכיל F כדי שיהיה חוקי.

```
<A ref="1">  
  <E>  
    <A ref="1">  
      <B serial="aaa"></B>  
    </A>  
  </E>  
  <F serial="1" ref="1"></F>  
</A>
```

**מס' אלמנטים במסמך זה:**

5



ב. (3) הצג מסמך XML שמציית ל DTD הנתון ואין בו אלמנטים מסוג B, או הוכח כי לא קיים מסמך כזה.

לא קיים. שורש המסמך לא יכול להכיל B – מכיוון שאין ID במסמך. לכן חייב להכיל E, E מכיל את A ולכן נקבל רקורסיה שנוכל לעצור אותה רק על ידי B - < לא קיים מסמך סופי שלא מכיל את B.

ג. (2) כתוב שאילתת Xpath (בגרסה 1.0 שנלמדה בכיתה בלבד) שמחזירה את כל האלמנטים f מסוג F כך שה-serial של f הוא תכונת ref של אלמנט כלשהו במסמך.

//F[@serial = //\*/@ref]

## 6. RDF , שאלת חובה 13 נק'

נתון גרף ה-RDF הבא:

yum:R	yum:serves	yum:AA
yum:S	yum:serves	yum:AA
yum:S	yum:serves	yum:BB
yum:R	yum:chef	dbr:LLL
yum:S	yum:chef	dbr:LLL
yum:T	yum:chef	dbr:MMM
yum:S	yum:manager	dbr:LLL
yum:T	yum:manager	dbr:MMM
yum:AA	yum:calories	680
yum:BB	yum:calories	820

א. (6 נק') לכל אחת מהשאלות הבאות, כתבו את התוצאות תחת הסמנטיקה הרגילה ללא RDFS.

```
SELECT DISTINCT ?r, ?q WHERE {
  {?r yum:manager ?p.
   ?s yum:chef ?p.}
 UNION
  {?s yum:serves ?q.}
}
```

Results:

{?r→yum:S} {?r→yum:T} {?q→yum:AA} {?q→yum:BB}

```
SELECT DISTINCT ?r, ?q WHERE {
  {?r yum:chef ?m.
   OPTIONAL {
     ?r yum:serves ?a.
     ?a yum:calories ?c.
     FILTER(?c>700) }
   }}.
  ?q yum:serves ?a.
}
```

Results:

{?r→yum:S , ?q→yum:S}

{?r→yum:R , ?q→yum:R}

{?r→yum:R , ?q→yum:S}

{?r→yum:T , ?q→yum:R}

ב. (4) הוסף מסמר מינימלי של שלשות לגרף RDF שהתוצאה של השאילתא הבאה תכלול את `yum:R` ו `yum:S` אבל לא את `yum:T`. הנח את הסמנטיקה הרגילה ללא RDFS.

```
SELECT DISTINCT ?r WHERE {  
  ?r yum:chef ?c?.  
  ?r rdf:type dbo:business  
}
```

Triples:

`yum:R` `rdf:type` `dbo:business`

`yum:S` `rdf:type` `dbo:business`

ג. (3) פתור את סעיף ב', כעת תחת סמנטיקת RDFS.

Triples:

`yum:serves` `rdfs:domain` `dbo:business`

## 7.7 Neo4j , 10 נק', שאלת בחירה מבין 7,8

נתון מסד נתונים גרפי Neo4j המייצג רשת חברתית, המכיל צמתים מהסוגים הבאים (לכל צומת label יחיד):

Person	Group
Name	Name
ID	

הקשתות בגרף הן מהסוגים הבאים (לכל קשת label יחיד):

MemberOf: מחבר בין Person לGroup

FriendOf: מבחר בין Person לPerson

נניח שבמסד הנתונים שלנו קשר של חברות הוא דו כיווני ולכן אם X חבר של Y תהיה קשת מא Y ל X ומא X ל Y.

שימו לב: תשובות ארוכות ומסורבלות יתר על המידה עלולות לגרום להורדת נקודות.

א. (4 נק') הסבר במילים מה מחזירה שאילתת Cypher הבאה:

```
Match (p:Person)-[t:MemberOf]-(g:Group)
WITH count(t) as r
MATCH (g2:Group)
return 1.0*r/count(g2)
```

הסבר:

מספר החברים הממוצע בקבוצות במסד הנתונים

ב. (3 נק') נניח שתוצאת השאילתא מא' היא 2, ובנוסף ידוע לכם כי במסד הנתונים קיימות שתי קבוצות בלבד.

כעת ביצענו את הפעולה הבאה הבאה בCypher:

```
CREATE (g:Group {name:"MAMAN_BEST"})
```

**כתוב שאילתת Cypher** שתוסיף קשר של שייכות לקבוצה בין אדם ששמו לא ידוע, אבל ידוע כי הוא החבר היחיד של דני, לבין קבוצה MAMAN\_BEST.

שאילתא:

```
MATCH (p:Person)-[:FriendsOf]->(:Person{name:"DANI"}),  
(g:Group {name:"MAMAN_BEST"})  
CREATE (p)-[:MemberOf]->(g)
```

ג. (3 נק') מה תהיה תוצאת השאילתא מסעיף א' כעת.  
(לאחר הפעולה שביצענו בסעיף ב' ולאחר הרצת השאילתא החדשה שנתבקשתם לכתוב)

$$\frac{5}{3} = 1.67$$

## 8. MongoDB , 10 נק', שאלת בחירה מבין 7,8

עבור מסד הנתונים Netclicks סרט מיוצג בצורה הבאה, באוסף movies

```
{  
  Movie_id:  
  Genre:  
  Profit:  
}
```

א. (6) השלימו את התכנית הבאה כדי שתחזיר את הרווח הממוצע של הסרטים בכל ז'אנר לאוסף **avgProfitPerGenre**

```
var map = function()  
{  
  emit(this.genre, this.profit)  
  
};
```

```
var reduce = function (key, values)  
{  
  return Array.avg(values)  
  
}
```

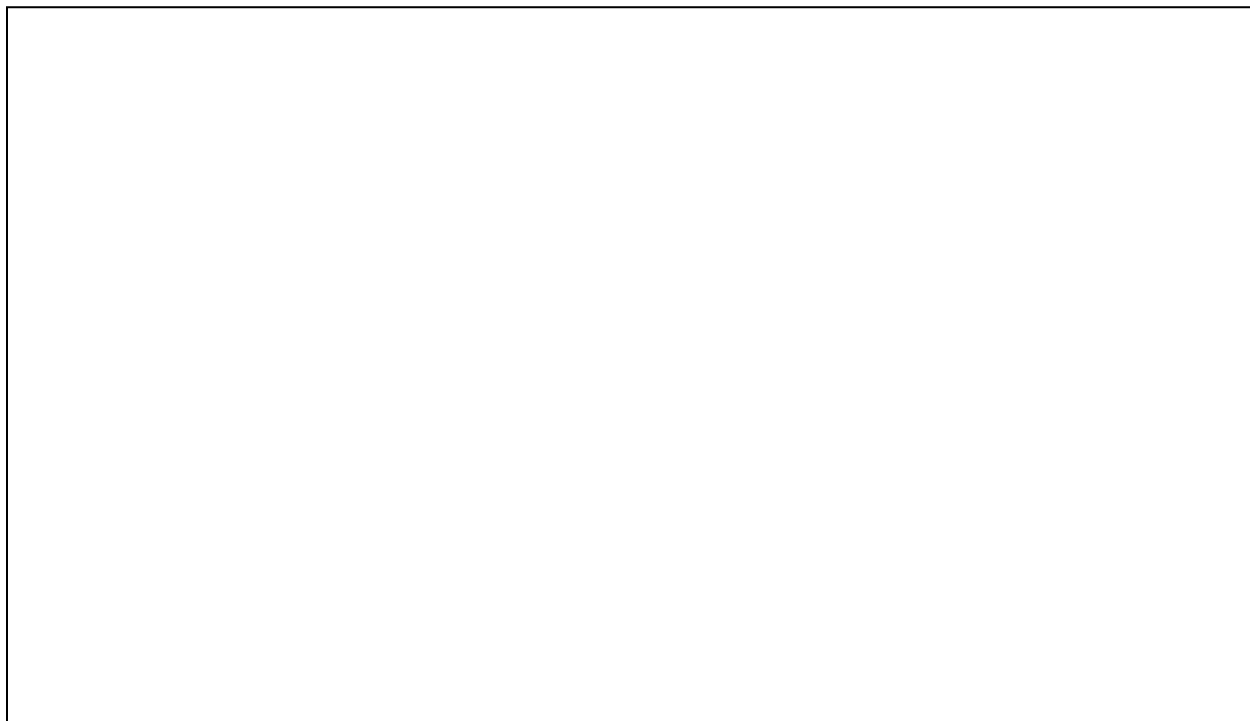
```
db.movies.mapReduce(  
  map, reduce,  
  {  
    query: {},  
    out : " avgProfitPerGenre"  
  }  
)
```

ב. (4) החזירו את הז'אנרים ממוינים לפי הרווח הממוצע בסדר יורד, ללא שימוש בתוצאה של סעיף א' (ניתן לפתור גם אם לא פתרתם את סעיף א'). חובה להשתמש בaggregate. ניתן להשתמש בavg \$ לחישוב ממוצע.

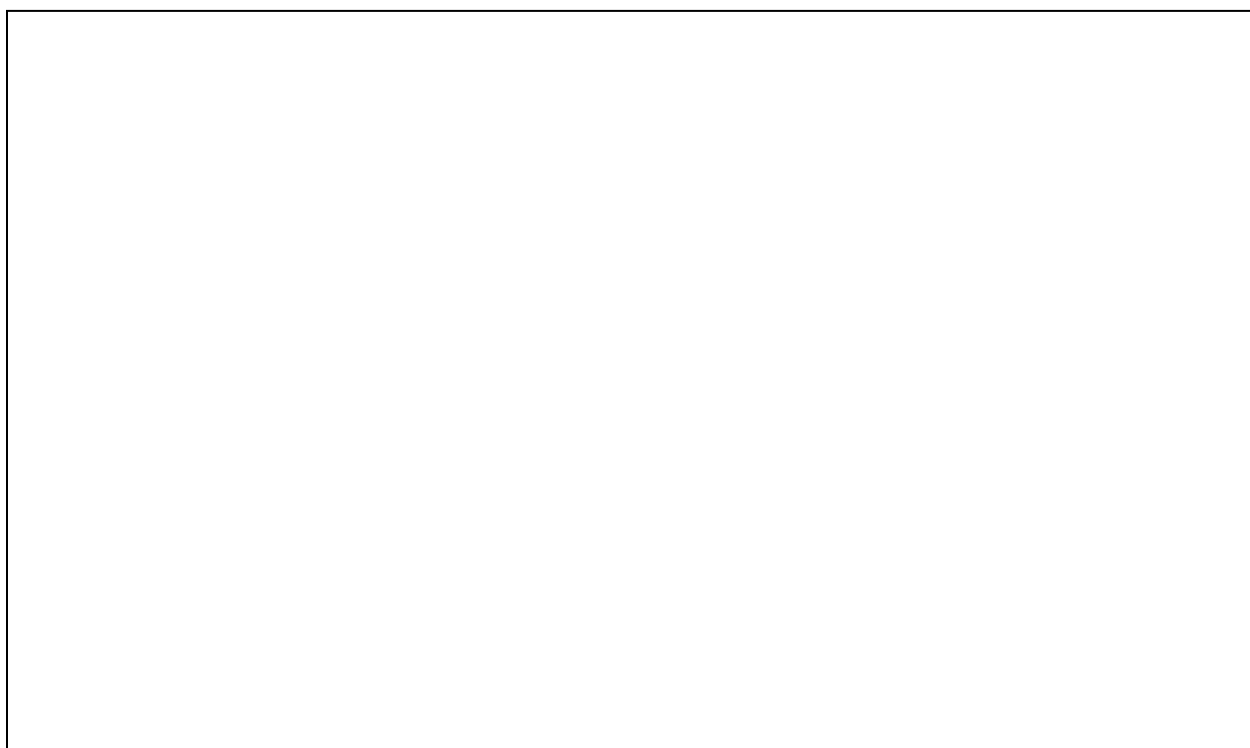
```
db.movies.aggregate(  
  {  
    $group :  
    {  
      _id: '$genre',  
      avg_profit:{$avg:'$profit'}  
    }  
  },  
  { $sort:{avg_profit:-1}  
})
```

אם אתם משתמשים בדף זה, ציינו זאת ליד השאלה/השאלות המקוריות, וציינו כאן את מספר/י השאלה/השאלות.

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_



שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_





שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_

שאלה: \_\_\_\_\_ סעיף: \_\_\_\_\_