הטכניון – מכון טכנולוגי לישראל הפקולטה למדעי המחשב

מרצה: פרופ' בני קימלפלד סמסטר אביב התשע"ז

מתרגלים: שובל לגזיאל

דביר דוקאן

ליאת פיטרפרוינד

מערכות מסד נתונים

236363

מועד א'

18 ביולי 2017

פירוט החלקים והניקוד:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| חלק | נושא | מס' שאלות בחלק | מס' שאלות שיש לענות עליהן | ניקוד |
| 1 | תכן מסדי נתונים | 2 | 2 | 30 |
| 2 | שאילתות 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 7 ו-8 ( 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.

1. משך הבחינה הינו שלוש שעות. תכננו את הזמן בהתאם.
2. 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. ERD, 10 נק'**

להלן חלק מדיאגרמת ה- ERD של מערכת מכרזים:

### Auction

#### Bid

### Customer

Containss

Item

**הסברים לדיאגרמה:**

* הישות **Customer** מייצגת את הלקוח. CId – מספר מזהה של לקוח, CName – שמו של הלקוח, CCard – מס' כרטיס אשראי של הלקוח, CAddr – כתובת של הלקוח.
* הישות **Auction** מייצגת מכרז. AId – מס' מזהה של המכרז, Type – סוג המכרז (מכירה פומבית או מכירה קבוצתית), Closed – האם המכרז סגור או לא (ערך בוליאני), CurPrice – המחיר הנוכחי (המינימלי שניתן להציע).
* הישות **Item** מייצגת פריט המוצע למכרז. IId – מס' מזהה של פריט, IDesс – תיאור מילולי של הפריט, Cost – עלות הפריט לספק, Quantity – כמות הפריטים במלאי.
* היחס **Bid** מתאר הגשת הצעה ע"י לקוח במכרז. Price – המחיר המוצע, NumPayments – מספר התשלומים, Win – האם ההצעה זוכה (בוליאני).
* היחס **Contains** מתאר שייכות של פריטים למכרזים.

א. (3 נק') בסעיף זה מטרתנו היא לתרגם את תרשים ה-ERD מהעמוד הקודם לסכמה רלציונית. מלאו את הפרטים הבאים.

טבלה Bid:

|  |  |
| --- | --- |
| שדות: |  |
| מפתח: |  |

טבלה Auction:

|  |  |
| --- | --- |
| שדות: |  |
| מפתח: |  |

טבלה Contains:

|  |  |
| --- | --- |
| שדות: |  |
| מפתח: |  |

ב. (2 נק') האם ייתכן שללקוח יהיה יותר מכרטיס אשראי אחד? יותר מכתובת אחת? הציע שינוי ל-ERD כך שהתשובה לכרטיס האשראי תהיה הפוכה וזו לכתובת תשאר ללא שינוי. ציירו את השינוי המוצע.   
**כן / לא ,** נימוק**:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

שינוי:

ג. (2 נק') האם פריט יכול להיות מוצע בו-זמנית ביותר ממכרז אחד? הצע שינוי ל-ERD כך שהתשובה תהיה הפוכה.   
**כן / לא** , נימוק: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

שינוי:

ד. (3 נק') מה הן הבעיות שעלולות להיווצר מהפיכת הלקוח לישות חלשה של המכרז )דרך הקשר Bid)?  
תשובה: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2. פירוקים ותלויות פונקציונליות, 20 נק'**

נתונה סכמה וקבוצת תלויות פונקציונליות

א. (6 נק') ציין שלושה מפתחות קבילים של R. (אין צורך לנמק.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ב. (5 נק') מצא כיסוי מינימלי ל- F. (אין צורך להציג את הדרך או להסביר.)

ג. (9 נק') נתון האלגוריתם הבא שכתב סטודנט שלומד בקורס ממ"ן שמטרתו לבדוק האם סכמה R עם קבוצת תלויות F היא בצורה הנורמלית 3NF:   
**שימו לב שבסעיף זה מדברים על סכמה כלשהי וקבוצות תלויות כלשהי ואין בהכרח קשר לסעיפים קודמים.**

Find a (minimal) key Z of R

For each **nontrivial** FD X ->Y in F {

If X is **not** a superkey then {

For each A in Y\X {

If A is not in Z then return **false**

}

}

}

Return **true**

1. אם האלגוריתם מחזיר **true** , האם הסכמה R **היא בהכרח ב3NF**?

**כן / לא , הסבר בקצרה:**

1. אם האלגוריתם מחזיר **false**, האם הסכמה R **בהכרח לא** **ב3NF**?

**כן / לא , הסבר בקצרה:**

1. הצע שינוי קטן לאלגוריתם שיגרום לו להיות נכון. והסבר האם כעת האלגוריתם לאחר השינוי רץ בהכרח בזמן פולינומיאלי בגודל של R וF.

**השינוי המוצע לאלגוריתם:**

**הסבר זמן הריצה:**

האם האלגוריתם לאחר השינוי שהצעת רץ בזמן פולינומיאלי בגודל של R וF

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

**3. SQL, RA, 20 נק'**

באתר הקניות Babazon מנהלים את מסד הנתונים, המייצג לקוחות מוצרים ורכישות, על ידי הטבלאות הבאות:

**CUST (Costumers Table)**

|  |  |  |
| --- | --- | --- |
| Cust\_ID | Name | Age |

**PROD (Products Table)**

|  |  |
| --- | --- |
| Prod\_ID | Price |

**PURCH (Purchases Table)**

|  |  |
| --- | --- |
| Cust\_ID | Prod\_ID |

1. בסעיף זה אתם מתבקשים לכתוב שאילתות SQL המחזירות את *מספר הזיהוי של, וממוצע המחירים של המוצרים שרכש, כל לקוח שביצע יותר מ-5 רכישות.* עליכם לכתוב **שתי שאילתות** בצורות שונות כפי שיתואר בהמשך, אשר מחזירות את אותו הפלט.
   1. (4 נק') שאילתת SQL ללא having (ניתן להשתמש בשאילתות מקוננות)

* 1. (4 נק') שאילתת SQL עם having וללא קינון שאילתות

1. (4 נק') כתבו שאילתת SQL המחזירה לכל מזהה מוצר את שמות הלקוחות הצעירים ביותר שקנו מוצר זה )השאילתא תחזיר זוגות של מזהה מוצר ומזהה לקוח).

1. (2 נק') כתבו שאילתת RA המחזירה את שמות כל הלקוחות שקנו את כל המוצרים במסד הנתונים.
2. (6 נק') האם ניתן לבטא את פעולת החילוק באלגברה הרלציונית באמצעות הפעולות הטלה, בחירה, מכפלה קרטזית, איחוד ושינוי שם (π σ × ∪ ρ)? אם כן, הראה את הנוסחה המגדירה חילוק; אם לא, הוכח.

**4. RC , DATALOG , 14 נק'**

באתר הקניות Babazon רוצים לבנות מערכת המלצות הממליצה לגולש על מוצרים דומים למוצרים שהוא רכש או התעניין בהם, בהתאם למאפייני הגולש או מאפייני המוצר.

בשאלה זו עליכם להניח כי קיימות אותן טבלאות שהוצגו בשאלה קודמת.

עבור לקוח c שקנה מוצר p נגדיר *מוצר מומלץ ל-*c: זהו מוצר שונה מ p שנקנה על ידי לקוח אחר, בן גילו של c, אשר קנה גם הוא את p.

1. (5 נק') כתבו שאילתת RC המחזירה עבור מזהה לקוח c את כל מזהיי המוצרים המומלצים לו ומחיריהם.

ניתן להשתמש ביחס

שימו לב כי למעשה עליכם להחזיר שלשות מהצורה CustId, RecommededProductId, price)).

1. (5 נק') לצורך דירוג ההמלצות מוגדרת פונקציית הדירוג הבאה: שני משתמשים שונים הם **דומים** אם הם בני אותו גיל, וכמו כן, קיים ביניהם מסלול של משתמשים בני אותו גיל כך שכל שני משתמשים שכנים במסלול קנו אחד או יותר מוצרים זהים.

כתבו תוכנית בDatalog המייצרת פרידקט בשם Similar(id1,id2) כך שהוא מתקיים עבור שני משתמשים id1, id2 אם הם דומים.

1. (4 נק')

נתונים שני ה EDB הבאים:

Vertex

|  |
| --- |
| Vertex\_id |

Edge

|  |  |
| --- | --- |
| Vertex\_id1 | Vertex\_id2 |

ונתונה התוכנית הבאה למציאת צמתים שאין בינם מסלול בגרף

הראו שני ריבודים שונים לתוכנית כאשר ריבוד אחד הוא תקין (חוקי), והשני לא תקין.  
הסבירו את תשובתכם.

**ריבוד תקין:**

**הסבר מדוע הריבוד תקין:**

**ריבוד לא תקין:**

**הסבר מדוע הריבוד אינו תקין:**

**חלק 3- מודלים לא רלציוניים – 36 נק' – בחלק זה שאלות 5 ו-6 הן חובה ועליכם לבחור בין שאלה 7 ל8.**

**5. XML , שאלת חובה 13 נק'**

נתון מסמך הDTD הבא:

<!DOCTYPE courses [

<!ELEMENT courses (faculty\*)>

<!ELEMENT faculty (course+)>

<!ATTLIST faculty name CDATA #REQUIRED>

<!ELEMENT course (prereq?,lecturer+)>

<!ATTLIST course number ID #REQUIRED

name CDATA #IMPLIED

grad (yes|no) "no"

>

<!ELEMENT prereq EMPTY>

<!ATTLIST prereq prenum IDREF #REQUIRED>

<!ELEMENT lecturer (#PCDATA)>

]>

א. ענו על הסעיפים הבאים בכן או לא, והסבירו בקצרה. שני הסעיפים מתייחסים למסמכי XML המצייתים לDTD הנתון.

a. (1 נק') האם ייתכן קורס שהוא קדם (prerequisite) לעצמו?

**כן / לא  
הסבר:**

b. (2 נק') האם האטריביוט prenumיכול להחזיק מצביע לאלמנט שהוא לא מסוג קורס (course)?

**כן / לא  
הסבר:**

ב. כתבו שאילתות XPath שמחזירות את המידע הבא מכל מסמך המציית לDTD הנתון. (שימו לב כי עליכם להשתמש בXPath1.0 בלבד, שזו הגרסה שנלמדה בכיתה).

a. (2 נק') כל הפקולטות שיש בהן לפחות קורס מוסמכים (@grad=”yes”) אחד.

**שאילתת XPath:**

b. (2 נק') כל הקורסים שהם דרישת קדם (@prereq) לקורסים מהפקולטה שלהם.

**שאילתת XPath:**

ג. (6 נק') לכל אחת משאילתות ה-XPath הבאות כתבו מסמךXML שמציית לDTD ויש לו תשובה לא ריקה עבור השאילתא, או הסבירו מדוע לא קיים כזה מסמך.   
שימו לב כי ייתכן שתוכלו להשתמש שוב בתשובתם באותו מסמך עבור יותר משאילתא אחת.

1. //faculty/course[@grad="yes"][../\*/@grad="no"]
2. //faculty[count(course)>1 and count(course[lecturer])=1]
3. //faculty[count(course)>1 and count(./\*[@name])=0]

**6. RDF , שאלת חובה 13 נק'**

נתון גרף הRDF הבא:

|  |  |  |
| --- | --- | --- |
| dbr:C1 | dbp:mayor | dbr:M1 |
| dbr:M1 | dbp:birthPlace | dbr:C22 |
| dbr:C22 | dbp:mayor | dbr:M22 |
| dbr:M22 | dbp:birthPlace | dbr:C1 |
| dbr:M22 | ***rdf:type*** | dbo:person |
| dbr:C333 | dbp:mayor | dbr:M333 |
| dbr:M333 | dbp:birthPlace | dbr:C333 |
| dbo:mayor | ***rdfs:range*** | dbo:official |
| dbo: official | ***rdfs:subClassOf*** | dbo:person |

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

a. שאילתא:

|  |
| --- |
| SELECT ?c, ?p WHERE {  { ?c dbp:mayor ?m. }  **UNION** {  ?p dbo:birthPlace ?c.  ?p rdf:type dbo:person. }  } |

תוצאות:

b. שאילתא:

|  |
| --- |
| SELECT ?c, ?m WHERE {  { ?c dbp:mayor ?m. }  **MINUS** {  ?p dbo:birthPlace ?c.  ?p rdf:type dbo:person. }  } |

תוצאות:

c. שאילתא:

|  |
| --- |
| SELECT ?c ?m WHERE {  { ?c dbo:mayor ?m.  OPTIONAL {  ?p dbp:birthPlace ?c.  ?p rdf:type dbo:person.  }  }  MINUS {  ?c dbp:mayor ?p.  }  } |

תוצאות:

ב. (3 נק') כתבו מה התוצאה של השאילתא האחרונה מסעיף קודם, אבל **הפעם תחת סמנטיקת** **RDFS**.

תוצאות:

ג. (4 נק') הסבר במילים מה מחזירה שאילתת הSPARQL הבאה:

SELECT DISTINCT ?c ?g ?p WHERE {

?c rdf:type dbo:City.

?c owl:sameAs ?d.

GRAPH ?g { ?d ?p ?o}

}

הסבר:

**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):

|  |  |  |
| --- | --- | --- |
| **Group** | **Post** | **Person** |
| Name | Title | Name |
|  | Text | ID |

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

Likes: מחבר בין Person לPost  
Posted: מחבר בין Person לPost  
BelongsTo: מחבר בין Person לGroup  
FriendOf: מבחר בין Person לPerson

נניח שבמסד הנתונים שלנו קשר של חברות הוא סימטרי: אם X חבר של Y אז יש קשת מX לY ומY לX.

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

1. (5 נק') כתבו שאילתת Cypher שמחזירה:

את שמות כל האנשים שעשו Like **לכל** הפוסטים שפורסמו על ידי מיכל.

1. (5 נק') נגדיר **מרחק בין אנשים** כ*אורך המסלול (כמספר הקשתות)* ***המינימלי*** *ביניהם המכיל אך ורק צלעות מסוג FriendsOf .*  
   כתבו שאילתת Cypher שמחזירה את המרחק הגדול ביותר בין שני אנשים ששייכים לקבוצה בשם MAMAN. (ניתן להניח כי קיימים לפחות שני אנשים ששייכים לקבוצה MAMAN אשר המרחק ביניהם גדול מ-0.)

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

*מנועי חיפוש* ממלאים את המשימה הבאה: בהינתן אוסף מסמכים ומילות מפתח לחיפוש, החזר את המסמכים הרלוונטיים ביותר למילות המפתח.

מנועי חיפוש משתמשים ב״אינדקס הפוך״ כחלק מתהליך החזרת תשובה לשאילתה. אינדקס הפוך בסיסי יראה מהצורה הבאה:

**{**

\_id **:** /\*a token string\*/**,**

Value**:** **{**

Docs**:** **[**

**{**

Doc\_id**:** /\*document id\*/

Appearances**:**/\*number of appearances of

token in the document\*/

**}**

**]**

**}**

**}**

כאשר token הוא מילה ו doc\_id הוא מערך שמכיל את המזהים של המסמכים שמכילים את המילה token.

תחת מסד הנתונים moogle נתון לכם אוסף (collection) בשם **intermediate**, המכיל חישוב ביניים שבוצע עבורכם (בדומה לword count שביצעתם בשיעורי הבית) המחזיק לכל מילה ומסמך, את מספר המופעים של המילה במסמך.

כל אלמנט ב-**intermediate** בנוי בצורה הבאה:

**{**

\_id**:**

**{**

Token **:** /\*a string token\*/

Doc\_id**:** /\*the document id\*/

**}**

value**:** /\*number of appearances of the word in the document\*/

**}**

1. (7 נק') עליכם ליצור את האוסף moogle\_ii שהוא אינדקס הפוך באמצעות map-reduce על ידי שימוש באוסף intermediate הנתון.

**var map = function()**

**{**

**}**

**var reduce = function (key, values)**

**{**

**}**

**db.intermediate.mapReduce(**

**)**

1. (3נק') את סעיף זה ניתן לפתור ללא תלות בפתרון של סעיף א'.

מה מחזירה השאילתה הבאה?

**שימו לב:**

Split(delimiter) מפצלת מחרוזת לפי הdelimiter למערך של מחרוזות.

$in : array בודק האם ערך נתון נמצא במערך.

**var** query **=** "........"//some input query for the system

db**.**moogle\_ii**.**aggregate**(**

**{**

$match **:** **{**\_id **:** **{** $in**:** query**.**split**(**" "**)}}**

**},**

**{**

$unwind**:** '$value.Docs'

**},**

**{**

$group**:** **{**

\_id**:** "$value.Docs.Doc\_id"**,**

Total **:{**$sum **:**"$value.Docs.Appearances" **}**

**}**

**},**

**{**

$sort**:{**Total **:** **-**1**}**

**})**

**תשובה:**

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:\_\_\_\_\_\_\_\_