

תרגיל 2 : אלגברה רלציונית וSQL

תאריך הגשה : 55:23, 10.04.22.

הוראות הגשה:

בתרגיל זה אתם נדרשים להגיש קובץ zip בודד שיכלול את הקבצים הבאים:

- ex2.pdf עם התשובות לשאלות בחלק של אלגברה רלציונית.
- q1.sql
- q2.sql
- q3.sql
- q4.sql
- q5.sql
- q6.sql
- q7.sql
- README שמכיל שורה בודדת ובו ה-login של הסטודנט שמגיש את התרגיל. אם התרגיל מוגש בזוגות, על שורה זאת להכיל את שני ה-login מופרדים בפסיק.

שימו לב:

- נא לקרוא על הדרישות המנהליות של הקורס בלינק באתר הקורס כדי למלא את ההוראות להגשה של קבצים סרוקים!
- תרגיל מוקלד יזכה ב-2 נקודות בonus!

הקדמה:

בתרגיל זה אנחנו נשתמש ביחסים ונתונים מתוך האתר <http://csrankings.org/>.

זהו אתר לדירוג מוסדות אקדמיים בתחום מדעי המחשב מסביב לעולם. האתר מדרג את המוסדות האקדמיים על סמך מספר הפרסומים של חברי סגל במוסד.

ניתן למצוא באתר מידע על כמות הפרסומים המדעיים של כל מוסד, לפי תחומים שונים של מדעי המחשב. ניתן לסנן את הנתונים לפי מדינות, לפי שנים ולפי תחומי המחקר. בנוסף ניתן לראות מידע על חברי סגל ספציפיים, וכן יש קישורים למידע מאתרים נוספים כמו dblp.

למשל, בצילום המסך הבא מוצג דירוג של המוסדות בישראל בתחום של Computer Vision בין השנים 1970-2021.

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. Click on a triangle (▶) to expand areas or institutions. Click on a name to go to a faculty member's home page. Click on a chart icon (the 📊 after a name or institution) to see the distribution of their publication areas as a bar chart. Click on a Google Scholar icon (🔍) to see publications, and click on the DBLP logo (📄) to go to a DBLP entry.

Applying to grad school? Read this first.

Rank institutions in by publications from to

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence ☐
- ▼ Computer vision ☒
- CVF
- CVPR ☒
- ECCV ☒
- ICCV ☒
- ▶ Machine learning & data mining ☐
- ▶ Natural language processing ☐
- ▶ The Web & information retrieval ☐

#	Institution	Count	Faculty
1	▶ Hebrew University of Jerusalem 📄 📊	77.8	9
2	▶ Technion 📄 📊	74.4	21
3	▶ Tel Aviv University 📄 📊	61.1	8
4	▶ Weizmann Institute of Science 📄 📊	50.5	8
5	▶ Ben-Gurion University of the Negev 📄 📊	12.7	4
5	▶ University of Haifa 📄 📊	12.7	4
7	▶ Bar-Ilan University 📄 📊	4.5	2
8	▶ Ariel University 📄 📊	2.5	1

אלגברה רלציונית (50 נקודות):

נתונים היחסים הבאים מתוך מסד נתונים של האתר CSRankings:

authors (name, conference, year, institution, count, adjustedcount)

conferences (conference, area, subarea)

institutions (institution, region, country)

הערות:

- בטבלה של מחברים (authors) יש את המידע על פרסומים של מחברים בכנסים שונים:
 - name – שם המחבר.
 - conference – שם הכנס שבו הוא פרסם.
 - year – השנה שבה פרסם המאמר בכנס.
 - institution – שם המוסד האקדמי של המחבר.
 - count – מספר המאמרים שהמחבר פרסם באותו הכנס.
 - adjustedcount – מספר הפרסומים היחסי של המחבר בכנס. למשל אם פרסם מאמר אחד והיה אחד משני כותבים הספירה היחסית תהיה 0.5. אם היו שלשה כותבים למאמר 0.33... וכו'.
- בטבלה של הכנסים (conferences) יש את המידע לגבי הכנסים:
 - conference – שם הכנס
 - area – תחום המחקר של הפרסומים בכנס
 - Subarea – תת-תחום המחקר.
- בטבלה של מוסדות (institutions) יש מידע על מוסדות אקדמיים:
 - institution – שם המוסד
 - region – אזור גיאוגרפי בעולם.
 - country – המדינה בה נמצא המוסד מיוצגת בתקציר ע"י שני אותיות. למשל ישראל היא il.

שימו לב: הטבלה authors מכילה מידע שחוזר על עצמו, כמו המידע על המוסד האקדמי של המחבר, שחוזר שוב ושוב בכל רשומה של פרסום של המחבר. בהמשך הקורס נלמד איך לתכנן טבלאות בצורה שבה נמנע מחזרתיות של מידע (דבר בנקרא "צורה נורמלית" של טבלה). אבל בעולם האמיתי נתקלו לצערנו ברבה מידע שאינו מנורמל, וכך גם המידע הנ"ל.

כתבו את השאילתות הבאות באלגברה רלציונית:

1. החזר את שמות המחברים ששייכים לאוניברסיטה העברית (Hebrew University of Jerusalem).
 $\pi_{name} \sigma_{institution=Hebrew\ University} (authors)$
2. החזר את שמות המחברים ששייכים למוסד אקדמי שנמצא בישראל, ואת שם המוסד בו הם עובדים.
 $\pi_{name, institution} \sigma_{country=il} (authors \bowtie institutions)$
3. מצא את שמות המחברים ששייכים למוסד אקדמי שנמצא בישראל שקיבלו ציון יחסי (adjustedcount) של לפחות 2 בכנס כלשהו בתת-תחום של db (מסדי נתונים) או בתחום של ai (בינה מלאכותית).
החזר את שמות המחברים ושמות המוסדות האקדמיים שלהם הם שייכים. (שימו לב שdb הוא subarea ואילו ai הוא area)
 $\pi_{name, institution} \sigma_{country=il \wedge count \geq 2 \wedge \left(\begin{matrix} area=ai \\ \vee subarea=db \end{matrix} \right)} (conferences \bowtie authors \bowtie institutions)$

4. א. החזר את שמות המחברים מהאוניברסיטה העברית (Hebrew University of Jerusalem) שפרסמו מאמרים גם בכנס בתת התחום של 'vision' וגם בכנס בתת התחום של 'ml'.

$(\pi_{name, \sigma_{subarea=vision} \wedge institution=Hebrew\ University} conferences \bowtie authors)$

$\cap (\pi_{name, \sigma_{subarea=ml} \wedge institution=Hebrew\ University} conferences \bowtie authors)$

ב. מצא את שמות המחברים מהאוניברסיטה העברית (Hebrew University of Jerusalem) עבורם קיימת שנה שבו הם פרסמו גם בכנס בתת התחום של 'vision' וגם בכנס בתת התחום של 'ml'.

$(\pi_{year, name, \sigma_{subarea=vision} \wedge institution=Hebrew\ University} conferences \bowtie authors)$

$\cap (\pi_{year, name, \sigma_{subarea=ml} \wedge institution=Hebrew\ University} conferences \bowtie authors)$

5. החזר את שמות המחברים שפרסמו אך ורק בתחום systems, וכן אך ורק לפני שנת 1990.

$\pi_{name, \sigma_{area=systems}}(authors \bowtie conferences) - \pi_{name, \sigma_{year \geq 1990 \vee area \neq systems}}(authors \bowtie conferences)$

6. החזר את שמות המחברים שפרסמו מאמר בכל אחד מהכנסים בתחום של בינה מלאכותית (ai) שבו - Noam Nisan פרסם (אך לא דווקא אותה שנה שבה נעם ניסן פרסם).

$\pi_{name, conference} authors \div \pi_{conference, \sigma_{area=ai \wedge name=Noam\ Nisan}}(Conferences \bowtie authors)$

7. נאמר שחבר סגל מהאוניברסיטה העברית הוא שיאן הפרסומים בעברית בכנס focs בשנה x אם אין חבר סגל אחר מהאוניברסיטה העברית עם יותר פרסומים ב-focs (כלומר count גדול יותר) באותה שנה. שימו לב שיכולים להיות כמה שיאנים במקרה של שוויונות.

לכל האחד מהשנים 2000 עד 2020 החזירו שורה, או שורות, במקרה של מספר שיאנים, מהצורה (y, m) כאשר y הוא השנה ו-m הוא שם שיאן הפרסומים בעברית.

(רמז: נסו קודם למצא את הקבוצה המשלימה לקבוצת השיאנים ובעזרתה למצוא את הקבוצה המבוקשת)

נמצא קודם כל את כל המחברים מהעברית שפרסמו ב-focs ואת השנה בה הם פרסמו:

$A = (\pi_{name, year, count, \sigma_{institution=Hebrew\ University\ of\ Jerusalem \wedge conference=focs}} authors)$

קעת נמצא את כל המחברים הנ"ל שיש מחבר אחר מהעברית באותה השנה עם כמות פרסומים יותר גדולה. זו הקבוצה המשלימה לקבוצה אנחנו מחפשים, ונחזיר את השנה ושם המחבר:

$N = \pi_{year, name, \sigma_{count1 > count2}}(\rho_{a1(name1, year, count1)} A \bowtie \rho_{a2(name, year, count2)} A)$

עכשיו נוריד מקבוצת כל המחברים את המחברים שבאותה שנה לא היו שיאנים:

$P = \pi_{year, name}(A) - N$

ולבסוף נחזיר את השיאנים של שהשנים בין 2000 לשנה 2020:

$results = \sigma_{year \geq 2000 \wedge year \leq 2020} P$

שאלות SQL (50 נקודות):

בכל התשובות לשאלות בחלק זה:

- השתמשו ב- SELECT DISTINCT כדי למנוע כפילויות בתשובות (אם כפילויות עלולות להווצר בתשובה).
- שימו לב: בכל סעיף כתוב באיזה סדר למיין את התוצאות וכן את שמות העמודות בתוצאה.

באתר הקורס יש קובץ create.sql המכיל הגדרות עבור הטבלאות וקובץ drop.sql המכיל פקודות המוחקות את הטבלאות. כמו כן, נתונים הקבצים:

- generated-author-info.csv
- conferences.csv
- country-info.csv

<http://csrankings.org> הקבצים מכילים מידע על מחברים, פרסומים, כנסים ומוסדות אקדמיים. המידע משמש את האתר לדירוג מוסדות אקדמיים בתחום מדעי המחשב.

את המידע המלא ניתן למצוא בלינקים הבאים:

<https://raw.githubusercontent.com/cohensara/csrankings/main/conferences.csv>
<https://raw.githubusercontent.com/emeryberger/CSrankings/gh-pages/generated-author-info.csv>
<https://raw.githubusercontent.com/emeryberger/CSrankings/gh-pages/country-info.csv>

ניתן למצוא את הקבצים גם במערכת המחשבים במעבדה בתיקה:

~ db/data/ex2/

ניתן להעתיק אותם לתיקה שלכם.

על מנת לבדוק את התרגיל שלכם, יש ליצור את הטבלאות בעזרת create.sql, ולטעון לתוכן נתונים בעזרת הפקודות

```
cat generated-author-info.csv | psql -h dbcourse public -c "copy authors from STDIN DELIMITER ',' CSV HEADER"
```

```
cat conferences.csv | psql -h dbcourse public -c "copy conferences from STDIN DELIMITER ',' CSV HEADER"
```

```
cat country-info.csv | psql -h dbcourse public -c "copy institutions from STDIN DELIMITER ',' CSV HEADER"
```

כתבו את השאילתות הבאות בSQL. שם הקובץ שבו צריכה להופיע התשובה לכל שאלה נמצא בתחילת השאלה. שימו לב ששאלות 1-7 זהות לשאלות בחלק של אלגברה אבל בשאלה 4 יש רק את השאילתה מהסעיף השני. חזרנו שוב על הגדרות השאילתות והוספנו הגדרה מדויקת של העמודות שצריכות להופיע בתוצאה והסדר של מיון התוצאה.

1. **(q1.sql)** חזרו את את שמות המחברים ששייכים לאוניברסיטה העברית (Hebrew University of Jerusalem). יש להחזיר טבלה עם העמודה name, ממיון לפי name.

```
select distinct name
from authors
where institution='Hebrew University of Jerusalem'
order by name;
```

2. **(q2.sql)** חזרו שמות המחברים ששייכים למוסד אקדמי שנמצא בישראל, ואת שם המוסד בו הם עובדים. יש להחזיר טבלה עם העמודות institution, name ממיון מיון ראשוני לפי institution ומיון שניוני לפי name.

```
select distinct institution, name
from authors natural join institutions
where country='il'
order by institution, name ;
```

3. **(q3.sql)** מצא את שמות המחברים ששייכים למוסד אקדמי שנמצא בישראל שקיבלו ציון יחסי (adjustedcount) של לפחות 2 בכנס כלשהו בתת התחום של db (מסדי נתונים) או בתחום של ai (בינה מלאכותית). חזרו את שמות המחברים ושמות המוסדות האקדמיים שלהם הם שייכים. (שימו לב שdb הוא subarea ואילו ai הוא area) יש להחזיר טבלה עם העמודות institution, name ממיון מיון ראשוני לפי institution ומיון שניוני לפי name.

```
select distinct institution, name
from conferences natural join authors natural join institutions
where country='il' and adjustedcount >= 2 and (area='ai' or subarea='db')
order by institution, name;
```

4. **(q4.sql)** החזר את שמות כל החוקרים מהאוניברסיטה העברית (Hebrew University of Jerusalem) שפרסמו באותה שנה גם בכנס בתת התחום של 'vision' וגם בכנס בתת התחום של 'ml'. יש להחזיר טבלה עם העמודות name, year, ממיון לפי name ומיון שניוני לפי year.

```
select distinct name, year
from authors natural join conferences
where institution='Hebrew University of Jerusalem' and subarea='vision'
intersect
select name, year
from authors natural join conferences
where institution='Hebrew University of Jerusalem' and subarea='ml'
order by name, year;
```

5. **(q5.sql)** החזר את שמות המחברים שפרסמו אך ורק בתחום systems, וכן אך ורק לפני שנת 1990. יש להחזיר טבלה עם העמודות name, ממיון לפי name.

```
select distinct name
from authors natural join conferences
where area='systems'
except
select name
from authors natural join conferences
where year >= 1990 or area != 'systems'
order by name;
```

6. **(q6.sql)** החזר את שמות המחברים שפרסמו מאמר בכל אחד מהכנסים בתחום של בינה מלאכותית (ai) שבו- Noam Nisan פרסם (אך לא דווקא אותה שנה שבה נעם ניסן פרסם). יש להחזיר טבלה עם העמודות name, ממיון לפי name.

```
select distinct name
from authors a1
where not exists
(
  select *
  from conferences natural join authors a2
  where a2.name='Noam Nisan' and area='ai' and
  conference not in(
    select conference
    from authors a3
    where a3.name=a1.name
  )
)
order by name;
```

7. **(q7.sql)** נאמר שחבר סגל מהאוניברסיטה העברית הוא שיאן הפרסומים בעברית בכנס *focs* בשנה x אם אין חבר סגל אחר מהאוניברסיטה העברית עם יותר פרסומים ב- *focs* (כלומר count גדול יותר) באותה שנה. שימו לב שיכולים להיות כמה שיאנים במקרה של שוויונות. לכל האחד מהשנים 2000 עד 2020 החזירו שורה, או שורות, במקרה של מספר שיאנים, מהצורה (y, מ), כאשר y הוא השנה ו-m הוא שם שיאן הפרסומים בעברית. יש להחזיר טבלה עם העמודות year, name, ממיון במיון ראשוני לפי year ומיון שניוני לפי name

```
select distinct year, name
from authors
where institution='Hebrew University of Jerusalem' and conference='focs' and year >= 2000 and year <= 2020
except
select a2.year, a2.name
from authors a1, authors a2
where a1.institution='Hebrew University of Jerusalem' and a2.institution='Hebrew University of Jerusalem'
and a1.conference='focs' and a2.conference='focs'
```

```
and a1.count>a2.count  
and a1.year=a2.year  
  
order by year,name;
```

בהצלחה!