

עקרונות תיאורטיים במסדי נתונים - 67713

16 ביולי 2013

מרצה: שוקי שגיב

איני לוקחת אחריות על מה שכתוב כאן, so tread lightly
אין המרצה קשור לסיכום זה בשום דרך.
הערות יתקבלו בברכה - noga.rotman@gmail.com אהבתם? יש עוד!
<http://bit.ly/integrali>

תוכן עניינים

3	מנהלות	0.1	
4	זמני ריצה של אופרטורים של אלגברת היחסים	1	
4	מוטיבציה	1.1	
4	כמה זמן לוקח למיין יחס לפי עמודות מסוימות?	1.1.1	
4	זמן חישוב של $\prod_X (R_i)$ (הטלה)	1.1.2	
4	זמן החישוב של איחוד $R_i \cup R_j$	1.1.3	
5	זמן החישוב של מכפלה קרטזית $R_i \times R_j$	1.1.4	
5	זמן חישוב של $R_i \bowtie R_j$	1.1.5	
6	מה קורה כאשר יש סדרת פעולות?	1.1.6	
6	צירוף טבעי מרובה $R_1 \bowtie \dots \bowtie R_k$	1.1.7	
8	יחסים שהסכמות שלהן חסרות מעגלים - <i>Acyclic Joins</i>	1.2	
8	הרדוקציה של <i>GYO</i>	1.2.1	
10	האלגוריתם לחישוב צירוף של יאנאקאקיס, <i>Yanakis 1981</i>	1.2.2	
14	<i>Top k</i>	2	
14	דוגמא	2.0.3	
15	האלגוריתם של <i>Lawler – Murty</i>	2.1	
16	הקשר בין זמן ריצה פולינומיאלי טוטלי לבין השהייה פולינומיאלית	2.1.1	
18	מציאת המסלול הקצר ביותר תחת אילוצים	2.1.2	
18	דוגמא להפעלת האלגוריתם - מכפלה קרטזית	2.1.3	
19	דוגמא להפעלת האלגוריתם - צירוף חסר מעגלים	2.1.4	
21	אלגוריתם הסף וגישות נוספות למציאת <i>Top – K</i>	2.2	
22	האלגוריתם של <i>Fagin</i>	2.2.1	
22	אלגוריתם הסף - <i>Threshold Algorithm</i>	2.2.2	
23	השוואה בין שני האלגוריתמים	2.2.3	
24	אלגוריתם הסף ללא גישות אקראיות (<i>NRA</i>)	2.2.4	
25	אלגוריתם משולב - <i>Combined Algorithm</i>	2.2.5	
26	<i>Top – K</i> מקורב	2.3	
28	אלגוריתמים לבדיקת שקילות והכלה של שאילתות	3	
28	שאילתות גימום	3.1	
28	הגדרות בסיסיות	3.1.1	
29	הגדרת שקילות והכלה, ומציאת תנאים מספקים	3.1.2	
32	שאילתות שיש בהן השוואות	3.2	
32	תכונות של יחס סדר מלא, צפוף ולא חסום	3.2.1	
33	בעיית הגרירה	3.2.2	
38	הכלה ושקילות בין שאילתות עם השוואות	3.2.3	
42	חוקים עם שלילה	3.3	
43	תנאי מספיק להכלה	3.3.1	
45	הכללות למקרה הכללי	3.4	
45	ריקנות	3.4.1	
46	בדיקת הכלה $Q_1 \supseteq Q_2$ בין שני חוקים עם שלילה וללא השוואות	3.4.2	

0.1 מנהלות

4/3/2013

הרצאה 1

בפידיאף ובפאוורפוינט באתר יהיה הסילבוס של הקורס. בכל אופן, בכמה מילים, בקורס הזה יש שלוש שעות הרצאה, למעשה שעתיים הרצאה ותרגיל, ותרגילים, שאינם חובה אבל יעזרו בהבנת החומר. היום מחקרים מראים כי אפילו מחשבה על תרגיל ולא הגעה לפתרון עצמו עוזרים בהבנה.

התרגילים הם בדרגות קושי שונות, ובסה"כ מכריחים את הפותר לעבור על החומר. אם כי, יש תרגילים יותר מעניינים. כדברי המרצה: "אני אומר שאני אבדוק אותם, אבל זה לא תמיד קורה". מידי פעם נפתור את התרגילים הקודמים בשיעור, ואז מי שפתר יוכל להציג את הפתרון על הלוח ולקבל קרדיט. כמו כן, מי שיגלה טעות בשקפים - יקבל גם כן בונוס.

אנא, לא להגיש תרגילים מועתקים.

כמו כן, המרצה אוהב לתת בבחינה תרגילים דומים או זהים לתרגילים שהיו בכיתה.

הציון נקבע ע"י בחינה של שלוש שעות.

המרצה יפתח מספר פורומים היום בערב - הראשון שימצא שגיאות במצגת, עליו לדווח בפורום המתאים, וכך

יהיה תיעוד. כמו כן, יהיה הפורום לשאלות ותשובות.

אם בכל אופן יש משהו שחייבים לדבר עליו באופן אישי ולא בפורום - לשלוח למייל של הקורס TheoryDB@cs.huji.ac.il.

שעות קבלה: בהפסקה, או בסוף השיעור (בשעה שבע). אם יהיה צורך, אפשר לקבוע מועד אחר דרך מייל

הקורס.

1 זמני ריצה של אופרטורים של אלגברת היחסים

למה? כי זהו דבר מהותי.
 "תשכחו מה שלמדתם בקורס של מסדי נתונים" - אנו מניחים שכל החישוב הוא בזיכרון הפנימי, ולכן כל ניתוח זמן הריצה כמו שלמדנו באלגוריתמים.
 זהו נושא "לחימום הקנה, לא יותר מזה".

1.1 מוטיבציה

נשתמש בפרמטרים הבאים - בהנתן יחס R_i :

- t_i יהיה מספר הרשומות.
- l_i יהיה אורך רשומה.
- n_i יהיה גודל היחס.

ולכן:

$$n_i = l_i \cdot t_i$$

1.1.1 כמה זמן לוקח למיין יחס לפי עמודות מסוימות?

צריך להתאים את מה שאנחנו מכירים לגבי מיון לדבר הספציפי עליו אנו מדברים.
 יש למיין t_i עצמים. כמה זמן זה לוקח? $t_i \cdot \log t_i$, בהנחה שהשוואה לוקחת יחידה אחת. זה הגיוני שזה יקח באמת יחידה אחת במקרה שלנו? לא, זוהי פונקציה של אורך רשומה אחת - $O(l_i)$, ולכן הזמן הכולל למיין יהיה:

$$O((t_i \cdot \log t_i) \cdot l_i) \leq O(n_i \cdot \log n_i)$$

כאשר אי השוויון למעלה נובע ישירות מהגדרות הפרמטרים. בהמשך, מכיוון וסימון זה קצר יותר, נשתמש בו.

1.1.2 זמן חישוב של $\prod_X(R_i)$ (הטלה)

עוברים פעם אחת על היחס, וכותבים התוצאה. על כן, זמן החישוב יהיה:

$$O(n_i)$$

נשים לב כי זמן ריצה של אלגוריתם הוא לפחות גודל האוטופוט, כי האלגוריתם צריך זמן לכתוב את האוטופוט.
 אם צריך לבטל כפילויות, אז צריך להוסיף מחיר של מיון, כלומר צריך להוסיף $O(n_i \cdot \log n_i)$.
 אפשר להשתמש ב-Hash לביטול כפילויות, ואז המחיר רק $O(n_i)$ - אם כי זו רק תחת ההנחות הבאות:

- פונקציית הערבול אידיאלית (אין התנגשויות).
- הפעלה של הפונקציה על רשומה לוקחת $O(l_i)$.

1.1.3 זמן החישוב של איחוד $R_i \cup R_j$

כותבים את הרשומות של כל יחס באוטופוט, ולכן המחיר הוא $O(n_i + n_j)$.
 אם רוצים לבטל כפילויות, אז צריך למיין (או לעשות שימוש ב-hash).
 מחיר למיון - $O((n_i + n_j) \cdot \log(n_i + n_j))$.

1.1.4 זמן החישוב של מכפלה קרטזית $R_i \times R_j$

1. בודקים אם אחד משני היחסים ריק; אם כן, התוצאה ריקה. מחיר הבדיקה קבוע.
2. אחרת, משרשרים כל רשומה של יחס אחד עם כל רשומה של היחס האחר, וכותבים רשומה זו באוטופוט. המחיר הוא:

$$O(t_i \cdot t_j \cdot (l_i + l_j))$$

האם מתקיים אי השיויון הבא?

$$t_i t_j (l_i + l_j) \leq t_i t_j l_i l_j$$

כן, אם $l_i, l_j \geq 2$. אחרת:

$$t_i t_j (l_i + l_j) \leq 2 t_i t_j l_i l_j$$

תרגיל 1: הוכיחו את הרשום מעלה.

על כן, נקבל כי:

$$O(t_i t_j (l_i + l_j)) \leq O(n_i n_j)$$

מסקנה 1.1 עבור הטלה, איחוד ומכפלה קרטזית, זמן החישוב הוא $O(\text{size of result})$, בהנחה שאין ביטול כפילויות.

1.1.5 זמן חישוב של $R_i \bowtie R_j$

עושים מכפלה קרטזית ותוך כדי יצירת התוצאה בודקים את התנאי של הצירוף. זמן לבדיקת התנאי של הצירוף:

$$O(l_i + l_j)$$

ולכן בסה"כ המחיר הוא:

$$O(n_i n_j)$$

זהו רק חסם עליון. יש מקרים בהם זה יקח פחות זמן, וכרגע זו הפעם הראשונה שאנחנו לא מצליחים לבטא את זמן הריצה כפונקציה של גודל הפלט.

אתגר: למצוא דרך לחישוב וניתוח מתאים, כך שיזמן הריצה ינתן כפונקציה של גודל האינפוט וגודל האוטופוט. אפשר לעשות זאת! כיצד?

- תחילה, ממיינים כל יחס על עמודות הצירוף.
- לאחר מכן עושים מעבר לפי המיון וצירוף כשצריך: עושים סריקה של היחסים הממויינים במטרה לגלות שני רצפים שאפשר לבצע ביניהם צירוף.

- שלב הסריקה דורש מעבר פעם אחת על כל יחס. על כן, מחיר של $O(n_i + n_j)$.

- צירוף של שני רצפים הוא גודל הרשומה המתקבלת, כלומר מכפלה קרטזית. לכן, כל שלב הצירוף מחירו O של גודל התוצאה המתקבלת ממנו. מסכמים על כל שלבי הצירוף, והמחיר הוא O של גודל התוצאה.

על כן, המחיר הכולל של המיון הסריקה, וביצוע הצירוף הוא:

$$O(n_i \cdot \log n_i + n_j \cdot \log n_j + r)$$

כאשר r הוא גודל התוצאה.

אם משתמשים בפונקציית $hash$ אידאלית במקום מיון, אז המחיר הוא:

$$O(n_i + n_j + r)$$

לא דיברנו על פעולת הבחירה - יש עוד לנתח את זמן החישוב שלה, אבל נשאיר זאת להמשך.

1.1.6 מה קורה כאשר יש סדרת פעולות?

נניח ויש לנו ביטוי מהצורה הבאה:

$$R_1 \times R_2 \times \dots \times R_k$$

מהו המחיר של החישוב?

נניח שלכל היחסים אותו גודל. אז המחיר הוא:

$$O(n^k)$$

כאשר n^k הוא גודל התוצאה.

הקלט: k יחסים

הפלט: מכפלה קרטזית שלהם.

זמן הריצה אקספוננציאלי ב- k , שהוא חלק מהקלט. עם זאת, הבעיה עצמה איננה קשה - הדבר היחיד שקשה בבעיה הוא גודל הפלט.

השאלה: האם זמן ריצה אקספוננציאלי בגודל הקלט מצדיק לקרוא לבעיה "קשה"? לא! ההגדרה של "בעיה קשה" כאשר הפלט הוא מאוד גדול, מצריכה הגדרה שונה. כיצד אם כך נאפיין "בעיות קשות" שהפלט שלהן הוא אקספוננציאלי בגודל הקלט? **גישא:** נחשב את זמן הריצה כפונקציה של גודל הקלט + הפלט.

1.2 הגדרה 1.2 זמן ריצה פולינומיאלי

טוטאלי פירושו שזמן הריצה הוא פולינום של גודל הקלט וגודל הפלט¹.

אם כך, חישוב $R_1 \times \dots \times R_k$ יכול להתבצע בזמן פולינומיאלי טוטאלי, כי זמן החישוב הוא:

$$O(k + r)$$

כאשר r הוא גודל התוצאה, ו- k הוא הזמן לבדוק אם אחד היחסים ריק (במקרה זה התוצאה ריקה).

1.1.7 צירוף טבעי מרובה $R_1 \bowtie \dots \bowtie R_k$

האם פעולה זו אפשר לעשות בזמן פולינומיאלי טוטאלי?

עבור $k = 2$ ראינו שאפשרי, וגם בזמן פולינומיאלי לא טוטאלי. אולם אם יש יותר משני יחסים, יתכן ונגיע לאקספוננט ב- k .

נשאל, האם אפשר לחשב התוצאה בזמן פולינומיאלי טוטאלי, כלומר בזמן פולינומיאלי בגודל הקלט והפלט? דוגמא - בשקופית 19: אפשר לגרום לתוצאת ביניים להיות גדולה מאוד, ורק בצעד האחרון "הכל מתרוקן". איך נכליל את השיטה הזו? שקופית 20: הצירוף של כל היחסים הוא ריק, אבל הצירוף של k מתוכם (כלומר כולם פרט לאחד) איננו ריק.

הדוגמא הזו היא "לא כל כך מעניינת", כי היא לא מייצרת תוצאת ביניים אקספוננציאלית.

תרגיל 2: יצר דוגמא של $k + 1$ יחסים, כאשר k יכול להיות גדול כרצוננו, כך שהצירוף של כל k יחסים מתוכם מייצר תוצאה שגודלה אקספוננציאלי ב- k ביחס לגודל הקלט, אבל הצירוף של כולם ריק (ככל שמצרפים מספר רשומות גדול יותר, מספר הרשומות גדל, מלבד הצירוף של כולם).

עובדה: הבעיה האם צירוף של k יחסים איננו ריק היא NP -שלמה. כלומר, אין לה אלגוריתם פולינומיאלי בגודל הקלט.

¹או $P(n, r)$ או $P(n + r)$

הערה 1.3 זוהי בעיית הכרעה: הפלט הוא בגודל קבוע - "כן" או "לא".**תרגיל 3 (קשה מהאחרים):** הוכח את העובדה מעלה.

נשים לב כי יש לנו כאן בעצם שני בעיות - בעיית אי הריקנות, כלומר האם תוצאת הצירוף איננה ריקה, וחישוב הצירוף. מה הקשר בין שתי הבעיות מבחינות הקושי - האם יתכן שאחת מהן תהיה קלה והאחרת קשה? בעיית חישוב הצירוף היא קלה אם אפשר לחשבה בזמן פולינומיאלי טוטאלי. אם יכולנו לפתור אותה בזמן כזה, בהכרח בעיית אי הריקנות היא קלה - כלומר, יהיה לה אלגוריתם לחישוב בזמן פולינומיאלי בלבד. דהיינו זוהי רדוקציה, ולכן מכיוון ובעיית אי הריקנות היא קשה, לא יתכן שהבעיה השנייה תהיה קלה. נראה כי זוהי אכן רדוקציה - בהנתן אלגוריתם פולינומיאלי טוטאלי לחישוב הצירוף $R_1 \bowtie \dots \bowtie R_k$, ניתן אלגוריתם פולינומיאלי לבעיית או הריקנות:

יהי $P(n, r)$ הפולינום שנותן חסם עליון על מספר הצעדים המדויק עבור קלט בגודל n ופלט בגודל r . נרץ את האלגוריתם $P(n, 0)$ צעדים - כלומר חסם עליון על מספר הצעדים שהאלגוריתם עושה כאשר הפלט הוא ריק. מספר צעדים זה הוא פולינומיאלי. מה יתכן ויקרה בסוף הריצה הזו:

- אם האלגוריתם עוצר, אז אנחנו יודעים מהי התוצאה, ואפשר לבדוק אם היא ריקה או לא. הבדיקה נעשית בגודל הפלט, שהוא פולינומיאלי.
 - אחרת, האלגוריתם לא עצר, אז התוצאה לא ריקה, שהרי מובטח שאם התוצאה היתה ריקה, הוא היה עוצר תוך מספר צעדים שכזה על הקלט.
- נשים ♡ כי אין צורך לדעת מהו הפולינום $P(n, 0)$, ומספיק לדעת שהוא קיים.

תרגיל 4: חשב בעזרת t_i, l_i, n_i את זמן הריצה של ביצוע $\sigma_C(R_i)$. חדי העין ישימו לב כי לא מספיק להגיד מהם הפרמטרים מעלה, צריך לדעת מהו הגודל של C , המורכב מהשוואות, $\wedge, \vee$. האם אפשר לדייק יותר? למשל, $O(\# \text{ of operators in } C)$.

מה ראינו עד כה?
מצד אחד, חישוב של מכפלה קרטזית הוא קל, כי יש אלגוריתם פולינומיאלי טוטאלי.
חישוב של צירוף במקרה הכללי הוא קשה, כי אין אלגוריתם פולינומיאלי טוטאלי.
האם תמיד חישוב של צירוף הוא בעיה קשה? לא! נדבר על מחלקה גדולה של צירופים, עבורם יש אלגוריתם פולינומיאלי טוטאלי.
דוגמא לצורך אינטואיציה - שקופית 28: כאן אפשר לחשב את הצירוף בלי בעיות, כלומר "שום דבר לא נופל בדרך" אם מחשבים משמאל לימין. עם זאת, המקרה הקשה הוא כאשר במקום היחס האחרון מחליפים עם היחס התחתון, ואז רק בסוף מתקבלת תוצאה ריקה.
נשים לב לדבר נוסף - אמרנו ש"שום דבר לא נופל בדרך": תמיד יש לנו שתי רשומות, שכל אחת מצטרפת למשהו בשלב הבא. תכונה נוספת: גם בדוגמא העליונה וגם בתחתונה, אם מסתכלים על זוגות של יחסים, אז תמיד יש להם אותה קבוצת ערכים בעמודה המשותפת, או שאין עמודות משותפות.
נגדיר את התכונות הללו באופן פורמלי. לפני כן נגדיר פעולה חדשה:

הגדרה 1.4 חצי צירוף של R לפי S , המסומנת ע"י $R \ltimes S$ היא פעולה שהתוצאה שלה היא:

$$R \ltimes S := \prod_R (R \bowtie S)$$

כאשר R מציין גם את היחס עצמו (קרי קבוצת הרשומות שלו), וגם את קבוצת האטריביוטים של היחס.

הערה 1.5 אלא אם נאמר אחרת, תוצאה של ביטוי אלגברי היא תמיד קבוצה, כלומר אין כפילויות. התוצאה של $R \ltimes S$ היא תת קבוצה של R שמכילה את כל הרשומות שאפשר לחבר לרשומה אחת לפחות ב- S . פעולה זו כאמור אינה סימטרית.

הגדרה 1.6 שני יחסים R ו- S הינם מח"צ (מצומצמי חצי צירוף), ובאנגלית - Semi Join Reduce, אם:

$$R \ltimes S = R$$

וגם

$$S \ltimes R = S$$

או באופן שקול:

$$\prod_{R \cap S} (R) = \prod_{R \cap S} (S)$$

כלומר, אי אפשר לצמצם את היחסים ע"י חצי צירוף.

הגדרה 1.7 צמצום R ע"י חצי צירוף עם S היא הפעולה בה R מקבל את הערך החדש $R \ltimes S$ (כלומר $R \leftarrow R \ltimes S$).

בדוגמא בשקופית 28 העליונה, כל זוג הם מצומצמי חצי צירוף. אותו דבר גם בדוגמא התחתונה, שכן מדובר בקבוצה - הסדר אינו משנה. ההבדל המהותי בין שתי הדוגמאות היא התוצאה (כלומר, אם מחליפים את היחס העליון הימני ביותר ביחס מתחתיו).

איך נביע את התכונה שדיברנו עליה לגבי הדוגמא העליונה - כל הרשומות משתתפות בתוצאה?

הגדרה 1.8 יחסים $R_1, \dots, R_k$ הינם מ"צ (מצומצי צירוף), ובאנגלית - Join Reduced, אם מתקיים לכל $1 \leq i \leq k$ השיויון הבא:

$$R_i = \prod_{R_i} (R_1 \bowtie R_2 \bowtie \dots \bowtie R_k)$$

כאשר R_i מציין גם את קבוצת הרשומות, וגם את הסכמה (קבוצת האטריביוטים).

הערה 1.9 שימו לב שתמיד מתקיימת ההכלה $\supseteq$.

תרגיל 5: הוכח את ההערה מעלה: תמיד יש הכלה בכיוון $\supseteq$.

נתונים $R_1, \dots, R_k$, וישנן התכונות מ"צ ומח"צ. מה היחס ביניהן? מ"צ $\Leftarrow$ מח"צ.

תרגיל 6: הוכח את הגרירה.

מה לגבי הגרירה בכיוון השני? ראינו דוגמא לכך שהיא לא בהכרח מתקיימת - הדוגמא התחתונה בשקופית 28: מח"צ אך לא מ"צ. נרצה להגדיר מחלקה שלמה של יחסים, עבורם הכיוון ההפוך **כן** מתקיים. אלו ידועים בשם:

1.2 יחסים שהסכמות שלהן חסרות מעגלים - *Acyclic Joins*

ליחסים אלה מספר תכונות מעניינות:

1. מח"צ $\Leftarrow$ מ"צ, כלומר, אם כל זוג יחסים הוא מח"צ, אז כל היחסים הם מ"צ.

2. יש אלגוריתם פולינומיאלי טוטאלי לחישוב הצירוף.

נראה בהמשך את האלגוריתם לביצוע צירוף של יחסים שכאלו.

הגדרה 1.10 יחסים הם חסרי מעגלים אם הרדוקציה של GYO מוחקת הכל.

1.2.1 הרדוקציה של GYO

הרדוקציה של GYO מתבצעת על סכמות. **תזכורת:** סכמה היא קב' האטריביוטים של היחס. ברדוקציה שני צעדים:

1. מחק אטריביוט שמופיע רק בסכמה אחת.

2. מחק סכמה שמוכלת בסכמה אחרת.

חוזרים על הצעדים האלה שוב ושוב, עד שאי אפשר למחוק עוד דבר. אם לבסוף לא נותר דבר - היחסים הם חסרי מעגלים, אחרת יש בסכמות מעגלים.

דוגמא בשקופית 38: בסוף התהליך הצלחנו למחוק את הכל - דהיינו היחסים חסרי מעגלים. בדוגמא בשקופית 39 אי אפשר למחוק כלום - אין אטריביוט שמופיע רק בסכמה אחת, ולכן הסכמות מעגליות. אפשר לגרום לסכמות מעגליות להיות לא מעגליות ע"י הוספת סכמה נוספת או אטריביוטים לסכמות קיימות. זהו דבר שאינו אינטואיטיבי! בשבוע שעבר ראינו חלק מהטענה:

11/03/2013
הרצאה 2

טענה 1.11 אם $k = 2$ אז מח"צ $\Leftrightarrow$ מ"צ.
אם $k > 2$ אז מח"צ $\Rightarrow$ מ"צ.

נוכיח את החלק הראשון של טענה, ובו בזמן ניתן דוגמא לסוג התשובה ששוקי מצפה לה משיעורי הבית. **הוכחה:** אם $k = 2$, והיחסים R_1, R_2 הם מח"צ, אז בצירוף אף רשימה אינה הולכת לאיבוד. כלומר, כל רשומה (של כל אחד מהיחסים) היא חלק מרשומה אחת לפחות של הצירוף. לכן, כאשר מטילים הצירוף על R_1 (או R_2), מקבלים את כל הרשומות המקוריות של היחס. צריך להשתמש גם בעובדה שציינו קודם: ההכלה $\supseteq$ שתמיד מתקיימת. ■

בהרצאה הקודמת ראינו את פעולת הצירוף $\times$, **עובדה:** R_i ו- R_j הם מח"צ אם ורק אם $R_i = R_i \times R_j$ וגם $R_j = R_j \times R_i$. עובדה זו נובעת מיידית מההגדרות: אם R_i, R_j הם מח"צ, אז חצאי הצירופים הנ"ל אינם משנים אותם. מצד שני, אם חצאי הצירופים אינם משנים דבר, אז לכל רשומה יחס אחד יש לפחות רשומה אחת תואמת בשני, ולכן הם מח"צ.

נזכור כי אנו רוצים למצוא אלגוריתם לחישוב $R_1 \bowtie \dots \bowtie T_k$ שרץ בזמן פולינומיאלי טוטאלי, כלומר פולינומיאלי בגודל הקלט והפלט:

$$P(n, r)$$

כאשר n הוא גודל הקלט, ו- r גודל הפלט. ראינו בהרצאה הקודמת את החשיבות במציאת אלגוריתם שכזה - הפרדה בין בעיות קלות וקשות "באמת", באשר הפלט הוא אקספוננציאלי בגודל הקלט, למשל - מכפלה קרטזית $R_1 \times \dots \times R_k$ אפשר לחשב בזמן פולינומיאלי טוטאלי. הראנו (באינדוקציה מראים לכל k יחסים):

$$R_i \times R_j = O(n_i \cdot n_j)$$

כאשר $n_i \cdot n_j$ הוא גודל התוצאה, n_i הגודל של R_i , ו- n_j הגודל של R_j . בפעם הקודמת השתמשנו בבעיית הכרעה (האם בהינתן k יחסים, האם הצירוף שלו ריק) שהיא NP -קשה ע"מ להראות שאלגוריתם כזה לא תמיד קיים.

דוגמא נוספת להפעלת האלגוריתם - שקופית 10: קו אופקי מציין מחיקת כל הסכמה, קו אלכסוני מציין מחיקת יחס.

תרגיל 1: הוכח שהתוצאה של GYO אינה תלויה בסדר המחיקות (כשיש כמה אפשרויות).

בגלל העובדה בתרגיל 1, המושג של יחסים חסרי מעגלים מוגדר היטב.

תרגיל 2: מצאו אלגוריתם יעיל ככל האפשר של הפעלת GYO . זכרו שהקלט הוא הסכמות, ולא היחסים עצמם.

כמובן ואפשר למצוא אלגוריתם פולינומיאלי שכזה (לראות אם אפשר להפעיל את אחד הכללים - פולינומיאלי, ועושים זאת בכמות לינארית בגודל הקלט).

היו $R_1, \dots, R_k$ יחסים חסרי מעגלים (מעטה ואילך, יחסים תמיד חסרי מעגלים, אלא אם כן נאמר במפורשות אחרת). נרצה להראות כי עבור היחסים הללו ניתן לחשב את צירופם בזמן פולינומיאלי טוטאלי. לשם כך, הדבר הראשון שעושים - בונים יער לפי הרדוקציה של GYO באופן הבא:

אם ברדוקציה R_j (או מה שנשאר ממנו) נמחק בגלל R_i (או מה שנשאר ממנו), R_j (היחס, לא הסכמות) הוא בן של R_i (היחס):

$$\begin{array}{c} R_i \\ | \\ R_j \end{array}$$

תרגיל 3: הראה כי הבניה מעלה מייצרת יער (כלומר אחד או יותר עצים שאינם מחוברים זה לזה).

מרגע זה ואילך נניח שנוצר עץ יחיד. אם יש יותר מאחד, מטפלים בכל אחד בנפרד, ואז עושים מכפלה קרטזית שלהם. השורשים ביער הם הסכמות שנמחקו כתוצאה מכך שכל האטריביוטים שלה נמחקו (ולא כתוצאה מהכלה).

1.2.2 האלגוריתם לחישוב צירוף של יאנאקאקיס, Yanakakis 1981

הערה 1.12 האלגוריתם הוא עבור $R_1, \dots, R_k$ שהם חסרי מעגלים ו- GYO מייצרת עבורם עץ יחיד.

באלגוריתם הזה יש שלושה צעדים:

1. חצי צירוף במעלה העץ.

2. חצי צירוף במורד העץ.

3. צירוף במעלה העץ.

נסביר כל צעד:

חצי צירוף במעלה העץ

1. מסמנים את העלים.

2. אם כל הילדים של R מסומנים ו- R עצמו לא מסומן, אז מצמצמים את R לפי חצי צירוף עם כ"א מהילדים, ואז מסמנים את R . חוזרים על צעד זה כל עוד אפשר.

למשל, אם נסמן את ילדי R ב- $S_1, \dots, S_k$:

$$R := R \ltimes S_1$$

$$R := R \ltimes S_2$$

⋮

$$R := R \ltimes S_k$$

כמה פעמים נעשית פעולת החצי צירוף? מספר היחסים פחות אחת: כל צלע מתאימה לפעולת חצי צירוף אחת.

חצי צירוף במורד העץ

1. מסמנים את שורש העץ.

2. אם האב R של S מסומן ו- S אינו מסומן, אז מבצעים צמצום של S לפי $R - S := S \ltimes R$, ומסמנים את S . חוזרים על 2 כל עוד אפשר.

צירוף במעלה העץ

1. כל עוד יש יותר מצומת אחת, נבחר עלה S עם אביו R . נחליף את R ב- $R \bowtie S$, ונמחק את S .

תרגיל 4: הביאו דוגמה מעניינת של האלגוריתם.

האם האלגוריתם נכון לכל אוסף של יחסים, גם לסכמות בעלות מעגלים? כן!

תרגיל 5: הוכיחו נכונות, כלומר הראו כי האלגוריתם מחשב את $R_1 \bowtie \dots \bowtie R_k$. הראו כי האלגוריתם נכון גם אם יש מעגלים בסכמות; במקרה זה, העץ נקבע באופן שרירותי.

טענה 1.13 בסיום שני השלבים הראשונים מתקיים לכל צומת R ובן S הם מח"צ, כלומר:

$$\prod_{R \cap S} (R) = \prod_{R \cap S} (S)$$

הוכחה "פיקטיבית":

הוכחה: בסיום שלב 1: $\subseteq$.

בסיום שלב 2: $\supseteq$.

ומהכלה דו כיוונית, התקבל שיוויון.

■

תרגיל 6: הראו שאם עושים תחילה חצי צירוף במורד העץ ואח"כ חצי צירוף במעלה העץ, אז הטענה מעלה אינה נכונה (לא נובע מבניית העץ עצמו - נכון לכל עץ).

נוכיח את הטענה "באמת": **הוכחה:** תחילה נראה שבסיום בשלב הראשון לכל אב R ובן S מתקיים:

$$\prod_{R \cap S} (R) \subseteq \prod_{R \cap S} (S)$$

ברור שההכלה מתקיימת מיד לאחר הפעולה $R := R \times S$. ההכלה ממשיכה להתקיים בהמשך (עד סיום השלב הראשון) כי R רק עשוי לאבד עוד רשומות בגלל חצאי צירוף עם בנים אחרים, ובנוסף, הבנים אינם משתנים יותר לאחר שמתחילים לצמצם את R .
עכשיו נראה שהשלב השני מייצר הכלה:

$$\prod_{R \cap S} (R) \supseteq \prod_{R \cap S} (S)$$

לכל אב R ובן S , והוא עושה זאת מבלי להרוס את ההכלות שנוצרו בשלב הראשון.
ההכלה הנ"ל מתקיימת מיד לאחר הפעולה $S := S \times R$. אם מתקיים:

$$\prod_{R \cap S} (R) \subseteq \prod_{R \cap S} (S)$$

אז מיד אחרי שעושים $S := S \times R$ חייב להתקיים שיוויון בין ההטלות, מכיוון שלא מוחקים מ- S רשומות שיש להן רשומות תואמות (אותם ערכים על האטריבייטס המשותפים) ב- R .
 S משתנה בשלב השני פעם אחת בלבד (שכן מדובר בעץ - יש לו אב יחיד)! עד ש- S משתנה, R רק יכול לאבד רשומות. לכן התכונה מהשלב הראשון נשמרת עד לביצוע $S := S \times R$, וכפי שהראנו, הביצוע הזה גורם לשיוויון בין ההטלות.

השיוויון ממשיך להשמר לאחר ביצוע $S := S \times R$, כי לא נוגעים יותר לא ב- R ולא ב- S .

עבור בונס, עברו על ההוכחה ומצאו חסרים!

לסיכום, הראנו שבסיום שני השלבים הראשונים, לכל אב R ובן S מתקיים ש:

$$\prod_{R \cap S} (R) = \prod_{R \cap S} (S)$$

■

תרגיל 7: הראו כי זה נכון תמיד (גם אם הסכמות מעגליות והעץ נקבע באופן שרירותי).

תכונת המופע הרציף

הגדרה 1.14 תכונת המופע הרציף פירושה שאם לשני יחסים בעץ יש אטריביוט משותף A , אז הוא מופיע בכל יחס על המסלול הלא מכוון הקצר ביותר (בכל יחס) המחבר אותם ל- lca שלהם, הוא האב הקדמון המשותף הנמוך ביותר בעץ.

דוגמא ציורית בשקופית 29.

תרגיל 8: הראו שתכונת המופע הרציף מתקיימת אם היחסים חסרי מעגלים והעץ נבנה לפי הרדוקציה של GJO .

תרגיל 9: הראו (תוך שימוש בתכונת המופע הרציף) שבמהלך כל השלב השלישי מתקיים:

$$\prod_{R \cap S} (R) = \prod_{R \cap S} (S)$$

בין R לבנו S .

שימו לב כי היחסים בצמותים משתנים במהלך שלב זה, וכזכור, בסוף שלב 2 שיויון זה מתקיים.

לכן במהלך כל הצירופים, אף רשומה אינה הולכת לאיבוד. על כן, תוצאות הביניים רק הולכות וגדלות, ואף אחת מהן אינה גדולה יותר מהפלט (התוצאה הסופית).

תרגיל 10: חשב את זמן הריצה כפולינום בקלט ובפלט. לא לשכוח את המחיר של שני השלבים הראשונים. הקלט הוא העץ (כלומר אין צורך לכלול את המחיר של יצירת העץ), וכמובן שהיחסים חסרי מעגלים.

פתרון תרגילים של הרצאה 2

תרגיל 3

צריך להראות שאם הרדוקציה של GJO מוחקת הכל, אז נוצר יער (קרי גרף חסר מעגלים), וזאת כאשר סכמה R היא בן של סכמה S אם R נמחקה בגלל הכלה ב- S .
נניח בשלילה שיש מעגל וניקח מעגל פשוט:

$$R_1, R_2, R_3, R_4$$

מבין כל הסכמות על המעגל נבחר את זו שנמחקה ראשונה. בה"כ נניח שזו R_1 . R_4 נמחקה אחרי R_1 , אבל לפי המעגל היא נמחקה בגלל הכלה ב- R_1 , בסתירה.

תרגיל 5

יש להראות כי האלגוריתם של ינקאקיס מחשב נכון את הצירוף אפילו אם קובעים את העץ בצורה שרירותית והיחסים מעגליים.

שני השלבים הראשונים מוחקים רק רשומות שאין להן למה להתחבר ביחס שכן, ולכן הן לא יכולות להיות חלק מהתוצאה של הצירוף. השלב השלישי מבצע צירוף, שהיא פעולה אסוציאטיבית וקומוטטיבית. לכן האלגוריתם נכון.

האלגוריתם נשאר נכון אפילו אם הופכים הסדר של שני השלבים הראשונים.

תרגיל 6

כזכור, בסוף שני השלבים הראשונים, כל אב ובן הם מח"צ, כלומר:

$$\prod_{R \cap S} (R) = \prod_{R \cap S} (S)$$

תכונה זו מתקיימת אפילו אם היחסים מעגליים והעץ נקבע שרירותית.

צריך להראות שהתכונה לא בהכרח מתקיימת אם הופכים את הסדר של שני השלבים הראשונים. "תפתרו את זה" - יש למצוא דוגמא נגדית.

תרגיל 7

ההוכחה שניתנה בכיתה תופסת גם למקרה הכללי, כי לא נעשה שימוש במבנה העץ או בכך שהיחסים חסרי מעגלים.

תרגיל 8

יש להראות כי תכונת המופע הרציף מתקיימת בעץ שנבנה לפי הרדוקציה של GYO עבור יחסים חסרי מעגלים. נזכיר כי התכונה אומרת שאם אטריביוט A מופיע ב- R וב- S , הוא מופיע בכל יחס על המסלול הלא מכוון הקצר ביותר בין שמחבר את R ו- S . נסמן את האב הקדמון המשותף הנמוך ביותר של שני היחסים ב- t . נפריד לשני מקרים:

- נניח תחילה ש- R אב קדמון של S , ו- A מופיע ב- R וב- S אבל לא בכל המסלול. תהי T הצומת הכי קרוב ל- S שאינו מכיל את A . יהי V הבן של T על המסלול מ- S ל- R . לפי הבחירה של V, T מכיל את A אבל T לא. כאשר V נמחק בגלל הכלה ב- R, T עדיין קיים, לכן לא יתכן ש- A נמחק מ- V לפני ש- V נמחק בגלל הכלה ב- T , כי A עדיין קיים ב- R . אבל T אינו מכיל את A , ולכן לא יתכן ש- V נמחק בגלל הכלה ב- T , בסתירה.
- את המקרה השני יש להשלים - אף אחד מ- R ו- S איננו אב קדמון של השני.

בתום שני השלבים הראשונים, כל אב ובן הם מח"צ. למה התכונה הזו חשובה? זה מבטיח שבהתחלה, מיד בתום השלב הראשון והשני, כאשר מחברים בין אב לבן ראשון, לכל רשומה ביחס אחד תהיה שורה ביחס האחר להצטרף אליה. רק אם היחסים חסרי מעגלים והעץ נבנה לפי GYO , התכונה הזו ממשיכה להתקיים בכל השלב השלישי. הוכחת טענה זו היא שאלה 3. הסיבה שהתכונה הזו נשמרת היא תכונת המופע הרציף. את הפתרון לשאלה זו יש להשלים.

תכונה זו מבטיחה כי גודל הפלט הסופי הוא חסם עליון על גודל כל תוצאת ביניים בעת ריצת האלגוריתם. בהתאם לכך יש לעשות את תרגיל 10 מההרצאה הקודמת גם כן.

תרגילים מהרצאה 3:

תרגיל 4: האם ניתן לחשב את $\prod_X (R_1 \bowtie \dots \bowtie R_n)$ כאשר הסכמות הן חסרות מעגלים באופן יעיל? "באופן יעיל" מצריך ביצוע הטלות מוקדם ככל האפשר. הראו במדויק איך לבצע הטלה "מוקדם ככל האפשר", כלומר מהן ההטלות הנדרשות בכל שלב במהלך השלב השלישי. אפשר להראות כי האלגוריתם רץ בזמן פולינומיאלי טוטאלי - מצאו חסם על גודל תוצאת ביניים שהוא פונקציה של הקלט והפלט, והראו בעזרתו שזמן הריצה הוא פולינומיאלי טוטאלי.

תרגיל 5: $X \subseteq R_i$. איך אז מחשבים את ההטלה של הצירוף הכי יעיל שאפשר, כאשר שוב הסכמות חסרות מעגלים?

מהו הנושא? ברצוננו למצוא את k התשובות הכי טובות.

2.0.3 דוגמא

נתון גרף G , צומת מקור s , צומת מטרה t , רוצים למצוא את k המסלולים הפשוטים (מסלול פשוט פירושו ללא מעגלים) הכי קצרים מ- s ל- t .

עד כה בדר"כ חיפשנו את המסלול הכי קצר (ואת בעיה זו פתרנו ע"י אלגוריתם דייקסטרה).

מספר המסלולים הפשוטים יכול להיות אקספוננציאלי בגודל הגרף.

אפשר למצוא את כל המסלולים הפשוטים ולדרג אותם - כל מסלול הוא לכל היותר כאורך הגרף. לכן למצוא

את כולם יהיה אקספוננציאלי בגודל הגרף, ואז לדרג.

את המסלול הכי טוב כאמור מוצא דייקסטרה. איך אפשר למצוא את השני הכי טוב?

רעיון: נניח שמצאנו את המסלול הכי קצר המורכב מכמה צלעות - נסמן ב- e_i . נוריד צלע אחת, ונריץ את

דייקסטרה. המסלול שהוא ימצא עכשיו הוא שונה מהמסלול שמצאנו בשלב הראשון, אולם לא מובטח שהוא השני.

לאחר מכן נוריד צלע אחרת, ושוב נריץ את דייקסטרה. נעשה כך לכל צלע בגרף. המסלול הכי קצר השני יהיה

הקצר מבין המסלולים הללו.

אבל, יש כאן עבודה מיותרת - יתכן ושני בדיקות כאלו יתנו את אותה תוצאה. אם כך עלינו לחלק את החישוב

למקרים זרים. כיצד נעשה זאת? נראה בעזרת דוגמא - באופן כללי יש צורך במספר מקרים השווה למספר

הצלעות. במקרה של 4 צלעות, נחלק לארבעה מקרים:

$$1. \neg e_1.$$

$$2. e_1, \neg e_2.$$

$$3. e_1, e_2, \neg e_3.$$

$$4. e_1, e_2, e_3, \neg e_4.$$

מצד אחד, ברור כי המקרים האלו זרים. מצד שני, החלוקה מכסה את כל המקרים, כי המסלולים פשוטים. לפיכך,

אם מסלול מכיל את e_1, e_2, e_3, e_4 הוא לא יכול להיות רישא ממש של מסלול פשוט מ- s ל- t .

זוהי שיטה כללית שנכליל לשיטה כללית למציאת $top\ k$, לא רק במקרה של מסלולים. נשים לב כי כאן אנחנו

צריכים אלגוריתם שיועד למצוא את המסלול הכי קצר תחת אילוצים. נחזור לכך בהמשך.

נמחיש את החלוקה בציור - שקופית 7: ככל שקרוב למרכז - יותר טוב. לכל נתח, מוצאים את הפתרון הכי טוב

תחת אילוצי ההכלה וההדחה. לאחר מציאת המסלול השני הכי טוב, נרצה לחלק את המרחב שוב למקרים זרים,

ושוב למצוא את הפתרון הכי טוב, תחת האילוצים החדשים.

נסמן לכל נתח את האילוצים ב- $[I_i, E_i]$, כאשר I_i קבוצת אילוצי הכלה, ו- E_i קבוצת אילוצי ההדחה.

נניח שהמסלול השני הכי טוב בא מנתח A_2 . נחלק את A_2 לנתחים זרים (כמספר הצלעות במסלול הקצר

השני) באופן הבא:

נניח כי A_2 מורכב מצלעות כלשהן e_1, e_2, e_3, e_4, e_5 , ובה"כ $E_2 = \{e'_3\}$, $I_2 = \{e_1, e_2\}$. המקרים החדשים

יהיו **תוספת** לאילוצים הקודמים בהם לא היה השני הטוב ביותר (דהיינו תוספת לנתחים A_1, A_3, A_4). נשים $\heartsuit$

כי כל מה שהיה ב- I_2, E_2 נשאר כחלק מהאילוצים החדשים, אחרת אנו עלולים לזלוג לנתחים שכנים. אזי מספר

המקרים החדשים יהיו כמספר הצלעות במסלול הקצר (שכעת מצאנו - כאן מדובר על המסלול הקצר השני) פחות

מספר הצלעות בהכלה באילוץ. בדוגמא שלנו, המקרים החדשים יהיו:

$$1. I = \{e_1, e_2\}, E = \{e'_3, e_3\}.$$

$$2. I = \{e_1, e_2, e_3\}, E = \{e'_3, e_4\}.$$

$$3. I = \{e_1, e_2, e_3, e_4\}, E = \{e'_3, e_5\}.$$

ונחפש את המסלול הכי קצר מבין כל הנתחים הללו והנתחים הקודמים.

תרגיל 1:

נתון גרף G , צומת מקור s , צומת מטרה t , קב' של אילוצי הכלה I וקבוצה של אילוצי הדחה E .

איך מוצאים מסלול הכי קצר שמקיים את האילוצים, כלומר מכיל את כל הצלעות של I ואינו מכיל אף צלע של E ?

2.1 האלגוריתם של Lawler – Murty

זוהי שיטה כללית המבצעת רדוקציה של מציאת k הפתרונות הכי טובים לבעיה של מציאת הפתרון הכי טוב תחת אילוצים.

זוהי שהיא שיטה כללית, ואיננה לא מכילה את כל הפרטים - יש להשלים בהנחן בעיה ספציפית: מהם האילוצים, איך מייצרים אותם, ולתת אלגוריתם שנותן את האלגוריתם הכי טוב תחת אילוצים. האלגוריתם משתמש בתור - *priority queue*, כאשר האיברים בתור הם מהצורה $[I, E, A]$, כאשר A הוא הפתרון הכי טוב שכבר מצאנו תחת האילוצים I, E .

בהתחלה, התור הוא ריק. האיבר הראשון שמוכנס לתור הוא $[\phi, \phi, A]$ - הפתרון הכי טוב ללא אילוצים. לכל פתרון יש משקל. העדיפות בתור נקבעת לפי משקל (נמוך יותר הוא טוב יותר). בתחילת איטרציה מוציאים את ראש התור - $[I, E, A]$ ומדפיסים את A . בעזרת A, E ו- I מבצעים חלוקה של הנתח של A לתתי נתחים המאופיינים ע"י קבוצת אילוצים חדשים: $[I_1, E_1], \dots, [I_n, E_n]$, המכילים את האילוצים המקוריים, כדי שלא נגלוש מהנתח לשכנים. עבור כל צמד אילוצים חדש $[I_i, E_i]$ מוצאים את הפתרון הכי טוב, ומכניסים את השלשה החדשה המבטאת את פתרון זה לתור. נשים לב כי שאר הפתרונות לנתחים האחרים כבר קיימים בתור, ולכן כשנוציא באיטרציה הבאה נוציא את הפתרון הבא הכי טוב (ויכול להיות שהוא כמובן לא זה שכרגע הכנסנו).

הערה 2.1 יתכן גם של- I_i, E_i אין פתרון; במקרה זה, לא מכניסים דבר לתור.

פסודוקוד מובא בשקף 15, כאשר פתרון מצויין ע"י האות S ולא ע"י האות A כפי שמופיע בתיאור. הקלט הוא D , הנתונים, ו- P , הבעיה בה מדובר.

```

1: Queue ← an empty priority queue
2: S ← SolveOpt (D, P, φ, φ)
3: if S ≠ ⊥ then
4: Queue.insert ([φ, φ, A])
5: while Queue ≠ φ do
6: [I, E, A] ← Queue.removeTop ()
7: print (A)
8: create pairs of constraints [I1, E1], ..., [In, En] for the opritimization problems spawned by [I, E, A]
9: for i ← 1, ..., n do
10: Ai ← SolveOpt (D, P, Ii, Ei)
11: if Ai ≠ ⊥ then
12: Queue.insert ([Ii, Ei, A])

```

• שורות 1 – 4 מאתחלות את ה-*queue*.

• $solveOpt(D, P, I, E)$ מקבלת את הנתונים הבאים:

- D - הנתונים, *data*.

- P - הבעיה בה מדובר.

- I - אילוצי ההכלה.

- E - אילוצי ההדחה.

הפרוצדורה הזו מוצאת את הפתרון הכי טוב תחת אילוצים.

• הלולאה מתחילה בשורה 5, וממשיכה כל עוד התור אינו ריק.

- בהתחלה היא מוציאה לתוך משתנה מתאים את השלשה, ומדפיסה את הפתרון S .

- יוצרים זוגות של אילוצים.

- עבור כל אחד מהזוגות, מוצאים פתרון, אם אכן הפתרון הוא אינו ריק, מכניסים את הפתרון לתור.

החלוקה לתתי נתחים היא תמיד כך שתתי הנתחים זרים זה לזה ומכסים בדיוק את הנתח המקורי, למעט הפתרון הכי טוב בנתח זה. לכן, כל פתרון מתקבל בדיוק פעם אחת. האלגוריתם מאפשר למצוא את כל הפתרונות בסדר עולה לפי משקל.

בהנתן בעיה ספציפית P , צריך להשלים את הפרטים הבאים:

1. להחליט מהם האילוצים.
 2. לקבוע איך מייצרים תתי נתחים (שורה 8).
 3. לממש את הפרוצדורה $solveOpt$ למציאת פתרון הכי טוב תחת אילוצים.
- איך אנחנו יכולים לבטא בצורה הטובה ביותר שהאלגוריתם הזה יעיל יחסית למציאת כל הפתרונות? נניח שמספר הפתרונות הוא לכל היותר אקספוננציאלי בגודל הקלט. איזה מדד הכי מתאים לניתוח זמן הריצה? פולינומיאלי טוטאלי, שכן יכול להיות שהאופטמום יהיה מאוד גדול. כדי שהאלגוריתם ירוץ בזמן פולינומיאלי טוטאלי, צריך לדרוש:

1. $solveOpt$ ניתן לבצע לכל היותר בזמן פולינומיאלי בגודל הקלט*.
 2. את שורה 8 (ייצור תתי הנתחים) ניתן לבצע בזמן פולינומיאלי בגודל הקלט*.
- * פולינומיאלי בגודל הקלט המקורי D . זאת בהנחה שמספר הפתרונות הוא לכל היותר אקספוננציאלי בגודל הקלט, והגודל של כל פתרון הוא פולינומיאלי בגודל הקלט.

הערה 2.2 דרישה 1 למעשה גוררת שהגודל של כל פתרון הוא לכל היותר פולינומיאלי בגודל הקלט.

צריך לדרוש שמספר הפתרונות יהיה לכל היותר אקספוננציאלי בגודל הקלט, כדי שהפעולות על ה- $queue$, הלוחות לוגריתם בגודל ה- $queue$, יקחו יותר מפולינום בגודל הקלט, ואז הריצה כולה לא תהיה פולינומיאלי טוטאלית. למעשה קיבלנו כאן זמן ריצה שהוא טוב יותר מפולינומיאלי טוטאלי - האלגוריתם רץ עם השהייה פולינומיאלי. מה זה אומר? אנו רוצים לא לחכות יותר מידי זמן בין איטרציות. איטרציה אחת של האלגוריתם היא פולינומיאלי באורך הקלט - אף פעם לא צריך לחכות יותר מזמן פולינומיאלי עד שרואים את הפתרון הבא. זמן ריצה פולינומיאלי טוטאלי לא מבטיח לנו את זה: הוא יכול למשל רק בסוף להתחיל להדפיס, במקרה זה מחכים זמן אקספוננציאלי עד שאנו רואים את הפתרון הראשון. זה לא לגמרי מספיק - נגדיר את התכונה באופן מסודר:

הגדרה 2.3 זמן ריצה עם השהייה פולינומיאלי משמעו שיש פולינום $P(n)$ באורך הקלט n בלבד, החוסם את כח אחד מהזמנים הבאים:

- הזמן מתחילת עבודת האלגוריתם עד הדפסת הפתרון הראשון.
- הזמן בין הדפסה של שני פתרונות עוקבים.
- הזמן בין הדפסת הפתרון האחרון ועד עצירת האלגוריתם.

תרגיל 2: האם זמן ריצה עם השהייה פולינומיאלי גורר זמן ריצה פולינומיאלי טוטאלי?
תרגיל 3: האם הכיוון ההפוך נכון?
 (תרגילים 4+5 שייכים לחלק הקודם של החומר)

2.1.1 הקשר בין זמן ריצה פולינומיאלי טוטאלי לבין השהייה פולינומיאלי

ניזכר בהגדרה של זמן ריצה עם השהייה פולינומיאלי - נשים לב כי ההגדרה היא הרבה יותר חזקה מזו של ריצה פולינומיאלי טוטאלית, שכן ריצה מהסוג השני יכולה להדפיס את כל התוצאות אחרי זמן פולינומיאלי בגודל הקלט והפלט. ההבדל הזה חשוב במיוחד במקרה שאנו דנים בו, בו לא צריך את כל התוצאות אלא את חלקן, כמו בחיפוש ה- $top k$.

22/04/2013
הרצאה 4

חישוב $R_1 \bowtie \dots \bowtie R_k$ כאשר זו מכפלה קרטזית, כלומר אטריבוט מופיע ביחס אחד בלבד. האם אפשר לחשב עם השהייה פולינומיאלית? כן! נעשה זאת בשיטה של לולאות מקוננות, כפי שנראה בשקופית 3 - כל לולאה תרוץ על השורות של יחס מסויים (לולאה לכל יחס), וכל איטרציה לוקחת את הרשומה הבאה מאותו היחס, ובלולאה המקוננת האחרונה מדפיסים את התוצאה. נראה שהזמן הריצה של האלגוריתם הוא אכן בכל השהייה פולינומיאלית:

- הזמן בין שני פתרונות מורכב מהזמן הנדרש להצבעה על הרשומה הבאה בחלק מהלולאות או בכולן. המקרה הגרוע ביותר הוא כאשר יש להצביע על רשומה חדשה בכל לולאה. כל שינוי הצבעה הוא זמן קבוע, ולכן סה"כ $O(k)$ במקרה הגרוע ביותר. כמו כן, צריך לבצע פעולת הדפסה אחת, והזמן הוא $k \cdot t$, כאשר t הוא חסם על גודל הרשומה בכל אחד מהיחסים.

- הזמן מתחילת האלגוריתם ועד שהתוצאה הראשונה מודפסת בשלמותה חסום ע"י $O(k \cdot t)$ (מקדמים פעם אחת כל לולאה מקוננת, כמו במקרה הגרוע ביותר למעלה, ואז מדפיסים את התוצאה).

- הזמן מסיום הדפסת התוצאה האחרונה העצירה חסום ע"י $O(k)$ (עוברים על כל הלולאות ורואים שעברו על כל הרשומות).

לכן $O(k \cdot t)$ חוסם את כל הזמנים.

חישוב $R_1 \times \dots \times R_k$

נניח והחישוב נעשה ע"י כך שתחילה מחשבים את $R_1 \times R_2$ ורושמים התוצאה, לאחר מכן מחשבים את המכפלה הקרטזית של תוצאה זו עם R_3 , וכו'. במכפלה קרטזית, התוצאה גדולה לפחות כמו כ"א משני היחסים המקוריים, לכן תוצאות ביניים אינן גדולות מהתוצאה הסופית. כמו כן, בהרצאה 2 ראינו שאפשר לחשב מכפלה קרטזית של שני יחסים בזמן פולינומיאלי בגודל התוצאה. לפיכך, החישוב שתואר מעלה מתבצע בזמן פולינומיאלי טוטאלי. אבל, זמן החישוב איננו עם השהייה פולינומיאלית, כי את הרשומה הראשונה של התוצאה רואים רק לאחר חישוב $R_1 \times \dots \times R_{k-1}$, וזה לוקח זמן אקספוננציאלי בגודל הקלט, כי הגודל של $R_1 \times \dots \times R_{k-1}$ הוא n^{k-1} , כאשר n הוא הגודל של כ"א מהיחסים.

נביא פתרון לתרגילים 2 ו-3 מההרצאה הקודמת:

השהייה פולינומיאלית $\Leftarrow$ זמן פולינומיאלי טוטאלי

אם אלגוריתם רץ עם השהייה פולינומיאלית, אז זמן הריצה הכללי חסום ע"י (מספר הפתרונות + אחד) כפול $P(n)$, כאשר זה הפולינום חוסם ההשהייה, וזה למעשה פולינום בגודל הקלט והפלט - כלומר זמן ריצה פולינומיאלי טוטאלי.

נשים לב כי זהו מספר הפתרונות בלבד ולא הגודל שלהם - ההשהייה כוללת את הגודל להדפסה, ולכן לוקחת בחשבון את הגודל.

זמן פולינומיאלי טוטאלי $\nRightarrow$ השהייה פולינומיאלית

נראה באמצעות דוגמא נגדית - הקלט הוא נוסחה של 3SAT, הפלט הוא כל ההשמות האפשריות בסדר הבא: תחילה מופיעות כל ההשמות המספקות, ולאחר מכן את כל ההשמות הלא מספקות.

מהו האלגוריתם? מייצרים את כל ההשמות. לכל אחת בודקים אם היא מספקת את הנוסחה (נדרש זמן פולינומיאלי בגודל הנוסחה לכל השמה). לבסוף ממיינים ההשמות כך שתחילה מופיעות כל המספקות. סה"כ זמן פולינומיאלי בגודל הקלט והפלט.

אי אפשר לפתור זאת עם השהייה פולינומיאלית. נניח בשלילה שכן. נריץ את האלגוריתם ונעצור אותו מיד לאחר הדפסת ההשמה הראשונה. זה לוקח זמן פולינומיאלי בגודל הקלט, כי ההשהייה פולינומיאלית. עכשיו נבדוק בזמן פולינומיאלי בגודל הנוסחה, אם ההשמה שהודפסה מספקת או לא.

אם כן, הנוסחה ניתנת לסיפוק. אחרת לא, כי כל ההשמות המספקות חייבות להיות מודפסות תחילה. לכן, קיבלנו אלגוריתם פולינומיאלי לבעיה האם נוסחה 3SAT ניתנת לסיפוק, אולם זוהי בעיה NP-קשה, בסתירה.

למעשה, הניסוח המדויק של התוצאה שהוכחנו הוא: אם $P \neq NP$ אז יש בעיה שאפשר לפתור בזמן פולינומיאלי טוטאלי אבל לא בזמן ריצה עם השהייה פולינומיאלית.

בחזרה ל-Lawler & Murty

בפעם הקודמת תיארנו את האלגוריתם של Lawler ו-Murty.

הרעיון המרכזי ישנם אילוצי הכלה I והדחה E . מוצאים פתרון הכי טוב שמקיים את I ו- E , מכניסים אותו לתור, וכשהוא יוצא מהתור, מדפיסים אותו, ומחלקים את תת המרחב ממנו הגיע הפתרון, מלבד הפתרון עצמו, ע"י יצירת אילוצים חדשים $(I_1, E_1), \dots, (I_n, E_n)$. לכל צמד שכזה מוצאים את הפתרון הכי טוב, ומוסיפים אותו לתור. אז שוב מוציאים את ראש התור, וחוזר חלילה. נדרשים תנאים מסויים כדי שהתהליך כולו ירוץ עם השהייה פולינומיאלית, בהם דנו בפעם הקודמת.

2.1.2 מציאת המסלול הקצר ביותר תחת אילוצים

נתונים אילוצי הכלה I ואילוצי הדחה E , כאשר I, E הן קבוצות צלעות, s המקור, t המטרה, ויש למצוא את המסלול הפשוט הקצר ביותר מ- s ל- t המכיל את כל הצלעות של I ואינו מכיל את הצלעות של E . נשנה את הגרף באופן הבא:

1. נמחק כל הצלעות של E .

2. בנינו את I כך שהצלעות שלו הן תחילה של מסלול מ- s ל- t . נסמן ב- s' את הקודקוד בצלע האחרונה במסלול ב- I שמחובר במסלול לצלע שאיננה ב- I . נשים לב שלא נכון לשלב את המסלול מ- s ל- s' עם מסלול מ- s' ל- t , כי יכולות להיות כפילויות. מה נעשה? לכל צומת השייכת לצלע ב- I , נמחק את כל הצלעות שלה (מספיק למחוק את הצלעות היוצאות) חוץ מזו שב- I .

ואז נמצא את המסלול הכי קצר מ- s ל- t . דוגמא לכך - שקופית 17.

הערה 2.4 יש כאן חשיבות עצומה לדרך היצירה של אילוצי I : לולא ידענו כי זוהי ראשית של מסלול, הבעיה היא NP -קשה.

תרגיל 1: הוכח כי אם I הן צלעות כלשהן (כלומר לא בהכרח תחילה של מסלול פשוט מ- s ל- t), הבעיה של מציאת מסלול פשוט הכי קצר שמכיל את I היא NP -קשה. האם היא גם NP -שלמה?

2.1.3 דוגמא להפעלת האלגוריתם - מכפלה קרטזית

נתונה מכפלה קרטזית $R_1 \bowtie \dots \bowtie R_k$, כאשר כל אטריביוט מופיע ביחס אחד בלבד. יהיו $A_1, \dots, A_m$ אטריביוטים המופיעים ביחסים $R_1, \dots, R_k$. ראינו שניתן לחשבה עם השהייה פולינומיאלית. צריך להדפיס את כל הרשומות של המכפלה הקרטזית בסדר עולה לפי $SUM(A_1, \dots, A_m)$. הנחה: אף אחד מהיחסים אינו ריק (אחרת התוצאה תהיה ריקה). נביא פתרון לבעיה בעזרת האלגוריתם של Lawler & Murty. על מנת לממשו, יש למלא את הפרטים הדרושים:

מציאת הפתרון הטוב ביותר (ללא אילוצים) מציאת הרשומה הכי טובה, קרי בעלת הסכום הכי נמוך. מכל יחס שמכיל חלק מ- $\{A_1, \dots, A_m\}$ ניקח את הרשומה הכי טובה, כלומר זו שעבורה הסכום של האטריביוטים (מבין $\{A_1, \dots, A_m\}$) שהיא מכילה הוא הקטן ביותר. מכל יחס שאינו מכיל אף אחד מבין $A_1, \dots, A_m$ נבחר רשומה אחת באופן שרירותי. נחזיר השרשור של הרשומות שבחרנו.

מציאת האילוצים החדשים - במקרה הראשון (ההתחלה) נניח והפתרון הכי טוב הוא $t_1, t_2, \dots, t_k$. האילוצים החדשים יהיו:

$$\begin{aligned} I_1 &= \phi, E_1 = \{t_1\} \\ I_2 &= \{t_1\}, E_2 = \{t_2\} \\ I_3 &= \{t_1, t_2\}, E_3 = \{t_3\} \\ &\vdots \\ I_k &= \{t_1, t_2, \dots, t_{k-1}\}, E_k = \{t_k\} \end{aligned}$$

מציאת הפתרון הטוב ביותר תחת אילוצים נתונות קבוצת אילוצי הכלה I וקבוצת אילוצי הדחה E . אלו קבוצות של רשומות של היחסים $R_1, \dots, R_k$. צריך למצוא רשומה הכי טובה של המכפלה הקרטזית, כלומר זו שעבורה $SUM(A_1, \dots, A_m)$ הוא הכי קטן, שמקיימת את I ו- E ; כלומר מכילה את כל הרשומות של I ואינה מכילה אף רשומה של E . מוצאים את הרשומה הכי טובה שמקיימת I ו- E ע"י שינוי הקלט המקורי:

- מוחקים כל הרשומות של E .
- עבור יחס שרשומה שלו ב- I , מוחקים את כל הרשומות האחרות של יחס זה (למעשה מספיק למחוק את כל הרשומות היותר טובות).
- אם יחס נעשה ריק, אין פתרון. אחרת פותרים כמו קודם עבור המקרה שאין אילוצים.

תרגיל 2: הראו איך לייצר אילוצים מפתרון S שהתקבל עבור אילוצים I ו- E .
תרגיל 3: מה אם מדובר בפונקציה כלשהי f ולא SUM ? מה צריך לדרוש לגבי f על מנת שאפשר יהיה להכליל הפתרון שנתנו?
תרגיל 4: וודאו כי הפתרון שנתנו מקיים את כל הדרישות כך ש- $Lawler - Murty$ יהיה נכון וירוף עם השהייה פולינומיאלית.

2.1.4 דוגמא להפעלת האלגוריתם - צירוף חסר מעגלים

עכשיו נניח שמדובר בצירוף חסר מעגלים $R_1 \bowtie \dots \bowtie R_k$, וצריך למצוא את כל הרשומות של התוצאה בסדר עולה לפי $SUM(A_1, \dots, A_m)$. בעיה: כאן לא בהכרח כל רשומה מתחברת עם כל רשומה אחרת. נניח תחילה שיש יחס אחד שמכיל את כל $A_1, \dots, A_m$ (אבל הם גם יכולים להופיע ביחסים אחרים). האם הנחה זו עוזרת לנו? כן! כעת הצטמצמו לבעיה הבאה: תהי t_i רשומה של R_i . איך מוצאים רשומה כלשהי של הצירוף $R_1 \bowtie \dots \bowtie R_k$ כך ש- t_i היא חלק ממנה? מוחקים את כל הרשומות מ- R_i מלבד t_i , ועושים את שני השלבים הראשונים של האלגוריתם של $Yanakakis$, כלומר חצי צירוף במעלה העץ, ואחר כך חצי צירוף במורד העץ.

תרגיל 5: הוכח כי אחד היחסים ריק אמ"מ כולם ריקים בסוף שני השלבים.
תרגיל 6: הוכח כי יש רשומה של הצירוף שמכילה את t_i אמ"מ היחסים לא ריקים בתום שני השלבים.

נניח שלאחר השלבים האלה היחסים לא ריקים. איך נמצא את הרשומה המבוקשת?

תרגיל 7: תארו באופן מלא את האלגוריתם למציאת רשומה של הצירוף $R_1 \bowtie \dots \bowtie R_k$ שמכילה רשומה t_i של R_i . בהנחה שהיחסים חסרי מעגלים. על האלגוריתם לרוץ בזמן פולינומיאלי בגודל הקלט.

פתרון לשאלה 7 מעלה, מתוך שקופיות ההרצאה:

מניחים שלאחר שני השלבים הראשונים של האלגוריתם של יאנאקאקיס היחסים אינם ריקים. לכן היחס R_i מכיל בדיוק את הרשומה t_i . נסתכל על העץ של GYO , שבצמתים שלו מופיעים היחסים אחרי שני השלבים. נבחר את t_i מהצומת עבור R_i ונסמן צומת זו. כל עוד יש צומת מסומן שלא כל שכניו מסומנים, נבחר צומת כזה R_j . תהי t_j הרשימה שנבחרה (בשלב הקודם) עבור R_j . לכל שכן לא מסומן R_d של R_j , נבחר רשומה t_d של R_d ששווה ל- t_j על העמודות המשותפות לשתיהן (מדוע יש רשומה כזו t_d ? אחרת בשלב השלישי של הצירוף לא היה ל- t_j מה להתחבר, והתוצאה של הצירוף היתה ריקה, ובפרט לא היתה רשומה ש- t_i חלק ממנה), ונסמן את R_d . בצורה כזו בוחרים רשומה אחת לכל צומת. תכונת המופע הרציף מבטיחה שהצירוף של הרשומות שהנבחרו איננו ריק, ולכן הוא רשומה של $R_1 \bowtie \dots \bowtie R_k$ שמכילה את t_i . בצורה זו מוצאים רשומה של $R_1 \bowtie \dots \bowtie R_k$ המכילה את t_i בזמן פולינומיאלי בגודל הקלט. שימו לב כי אילו היינו מבצעים צירוף מלא של היחסים, אחרי שני השלבים הראשונים התוצאה היתה יכולה להיות אקספוננציאלית בגודל הקלט, ולכן זמן הריצה במקרה הגרוע הוא אקספוננציאלי. ברור שכל הרשומות של הצירוף (אילו היה מתבצע) מכילות את t_i , כי זו הרשומה היחידה ב- R_i . כמו כן, אי אפשר סתם לבחור רשומה מכל יחס, כי לא מובטח שאפשר לצרף אותן יחד.

דוגמא:

$$R_1(A, B), R_2(B, C)$$

נבחר $R_1(1, 1)$ ו- $R_2(0, 0)$; שתי הרשומות האלו לא מתחברות. שרטוט בשקופית 35.

תרגיל 8: איך מוצאים עם השהייה פולינומיאלית את כל הרשומות של צירוף חסר מעגלים $R_1 \bowtie \dots \bowtie R_k$ בסדר עולה לפי $SUM(A_1, \dots, A_m)$? לפתור זאת באופן מלא ומדויק לפחות עבור המקרה שבו יש יחס אחד שמכיל את כל האטריביוטים $A_1, \dots, A_m$.

תרגיל 9: יהי $R_1 \bowtie \dots \bowtie R_k$ צירוף (שהוא אינו בהכרח חסר מעגלים), אבל כזה שכשעושים אותו משמאל לימין, אז כל צירוף מתבצע על קבוצת אטריביוטים שהיא מפתח על של אחד משני היחסים. הוכיחו שאפשר לחשב הצירוף בזמן פולינומיאלי בגודל הקלט בלבד.

הערה: מפתח על היא קבוצת אטריביוטים (לאו דווקא מינימלית) שערכים עבורה קובעים רשומה באופן אחד ויחיד.

29/4/2013
הרצאה 5

נתונה פונקציה $SUM(A_1, \dots, A_m)$ כאשר הם אטריביוטים. רוצים לחשב את הרשומות של צירוף חסר מעגלים בסדר עולה לפי פונקציית ההקבצה הנ"ל, כאשר זמן ריצת האלגוריתם יהיה בעל השהייה פולינומיאלית. בפעם הקודמת התחלנו לפתור את הבעיה עם הנחה מקלה: הנחנו כי קיים יחס אחד בו כל האטריביוטים המתבקשים קיימים, בעזרת האלגוריתם של לולר מורטי. אלגוריתם זה מצריך אילוצים: האילוצים הם I קבוצות של רשומות מהיחסים המשתתפים בתוצאה, שהיא רשומה של הצירוף, ו- E קב' רשומות שלא משתתפות בתוצאה.

איך מוצאים את הרשימה הכי טובה תחת אילוצים?
תחילה, מפעילים את האילוצים, כלומר מוחקים את הרשומות שנמצאות ב- E , וכמו כן, עבור כל רשומה של I , מוחקים את כל הרשומות האחרות מהיחס של רשומה זאת. אם כך, עשינו רדוקציה לבעיה של מציאת הפתרון הכי טוב, ללא אילוצים.

כאמור, בשלב זה אנו מניחים כי יש יחס אחד המכיל את כל האטריביוטים $A_1, \dots, A_m$, וכמובן שהם יכולים להופיע גם ביחסים אחרים.

לאחר מכן, מבצעים חצי צירוף במעלה העץ וחצי צירוף במורד העץ (שני הצעדים הראשונים באלגוריתם של יאנאקאקיס). זה מבטיח כי כל השורות שנשארו יהיו בצירוף. אם היחסים שנותרו ריקים, אין פתרון תחת האילוצים. אחרת, היחסים אינם ריקים (הוכחנו בשיעורי הבית כי אם יחס אחד כולם ריקים) בטוח יש פתרון. אם נפתור בצורה נאיבית, כלומר נחשב את כל הצירוף ואז נחשב את הטוב ביותר, נקבל זמן ריצה אקספוננציאלי, וכך לא נוכל לקבל אלגוריתם עם השהייה פולינומיאלית.

אזי, בוחרים מהיחס המכיל את כל $A_1, \dots, A_m$ את הרשומה הכי טובה (קרי, עם הסכום הכי קטן, כי רוצים סדר עולה). לרשומה הכי טובה שנבחרה מצרפים רשומה אחת מכל יחס אחר, לפי ההסבר מההרצאה הקודמת (ושאלה 7 בשיעורי הבית).

נניח ב.ה.כ כי השורש של העץ הוא היחס שמכיל את כל $A_1, \dots, A_m$ (אפשר להניח זאת כי ניתן לבנות את העץ כך שלא יהיה מכוון, ולעצים לא מכוונים אנו מכירים אלגוריתם שיכול להפוך כל קודקוד בעץ לשורש). מכל ילד לוקחים רשומה שמתחברת לזו שנבחרה עבור ההורה, וממשיכים באופן רקורסיבי. צריך להראות שהרשומות שנבחרו מצטרפות לרשומה יחידה של הפתרון. זה נובע משתי סיבות:

- בסוף שני השלבים של יאנאקאקיס, כפי שהוכחנו, כל אב ובן הם מ"צ, ועל כן הבניה מוגדרת היטב; כלומר בסופה בחרנו בדיוק רשומה אחת בכל יחס.
- הצירוף של הרשומות האלו איננו ריק (כלומר, מייצר בדיוק רשומה אחת של התוצאה) בגלל תכונה המופע הרציף (תכונה זו מראה גם שבאלגוריתם של יאנאקאקיס "שום דבר לא נופל בדרך" והתוצאות רק הולכות וגדלות).

פתרון לשאלה 2 משיעורי הבית מההרצאה הקודמת (שקופית 8):

בהנתן פתרון $t_1, \dots, t_k$ עבור אילוצים:

$$I = \{t_1, \dots, t_k\}_{0 \leq i \leq k-1}, E = \{\text{something (could be repetative)}\}$$

מייצרים חדשים עבור $k-1$ בעיות חדשות:

$$\begin{aligned} I_1 &= I, E_1 = E \cup \{t_{i+1}\} \\ I_2 &= I \cup \{t_{i+1}\}, E_2 = E \cup \{t_{i+2}\} \\ &\vdots \\ I_{k-1} &= I \cup \{t_{i+1}, \dots, t_{k-1}\}, E_{k-1} = E \cup \{t_k\} \end{aligned}$$

אם יש רשומה שמופיעה גם ב- I_j וגם ב- E_j ל- j כלשהוא, אז ל- $[I_j, E_j]$ אין פתרון.

אם כך, הראינו את כל מה שנחוץ ע"מ להפעיל את האלגוריתם של לולר מורטי. נותר להראות כי הוא יפעל בזמן עם השהייה פולינומיאלית. בשאלה 1 בשיעורי הבית היינו צריכים להוכיח כי אם סדר לצלעות אז הבעיה NP -קשה. נשים לב כי במקרה שדנו בו בשיעור זה, אין חשיבות לסדר בעת יצירת האילוצים, כלומר כאן אין בעיה כזו.

תרגיל 1: איזה רשומות אפשר למחוק מ- E_j בלי למחוק את האילוצים? אם מוחקים מ- E_j את כל הרשומות המיותרות, מכמה יחסים נשארות רשומות ב- E_j ?

בחזרה למקרה הכללי - אין יחס אחד המכיל את כל ה- A_i מה יש לשנות כדי להרכיב פתרון למקרה הכללי? האם יש משהו מהמקרה של פתרון תחת אילוצים שאפשר להשתמש בו במקרה הכללי? האם אנחנו עדיין יכולים לחשב רשומה אחת בלבד של התוצאה - זו שאנחנו מחפשים, כפי שעשינו במקרה הקודם? לא, אבל אנו יכולים להבטיח שמחשבים רק מספר פולינומיאלי של רשומות מהתוצאה. נביט במקרה של שני יחסים (שקופית 12) R_i, R_j , כאשר חלק מהאטריביוטים מ- A_i ב- R_i , וחלק ב- R_j , ונניח כי B הוא האטריביוט המשותף לשני היחסים. בשקופית מסומנות הרשומות שאנחנו צריכים. כלומר, לכל ערך שעליו מתבצע צירוף יש לבחור את הרשומה הכי טובה, ואז לבצע צירוף ולבחור את התוצאה הכי טובה מביניהן. נשים לב כי נעשה כאן צירוף על מפתח על, ובצורה כזו אנו מבטיחים כי מספר התוצאות יהיה פולינומיאלי בלבד (הפעולה ב- SQL היא $Group By$). נכתוב זאת בצורה מסודרת עבור האלגוריתם המלא.

תרגיל 2: לנסח במדויק את השלב השלישי של האלגוריתם של יאנאקקיס תוך שימוש ברעיון מעלה, ולהוכיח ששלב זה לוקח זמן פולינומיאלי בלבד בגודל הקלט.

בדרך זו פותרים את הבעיה של מציאת הפתרון הכי טוב תחת אילוצים בזמן פולינומיאלי. צריך להשתמש בפתרון התרגיל הבא (שניתן בפעם שעברה):
 $R_1, \dots, R_k$ יחסים כלשהם (לאו דווקא חסרי מעגלים). נתון שאפשר לשים סוגריים לצורך חישוב הצירוף של $R_1 \bowtie \dots \bowtie R_k$ כך שכל פעולת צירוף תתבצע בין שני יחסים שהאטריביוטים המשותפים ביניהם הם מפתח על באחד מהם (זו גרסה קצת יותר כללית לשאלה מההרצאה הקודמת). צריך להראות שאפשר לחשב הצירוף $R_1 \bowtie \dots \bowtie R_k$ בזמן פולינומיאלי באורך הקלט בלבד.
 אם אנחנו רוצים לייצר את הרשומות בסדר יורד לפי הסכום: כשבחרנו את הרשומה "הכי טובה", לבחור את הרשומה עם הסכום הכי גבוה, וזהו.

2.2 אלגוריתם הסף וגישות נוספות למציאת $Top - K$

הגדרת הבעיה נתונים אובייקטים, שלכל אחד מהם ערכים עבור קבוצה קבועה של אטריביוטים $A_1, \dots, A_n$. הערך של אטריביוט הוא תמיד בין 0 ל-1.
 עבור כל אטריביוט נתונה רשימה ממוינת (נניח באופן קבוע כי הרשומות מופיעות בסדר יורד, וככל שהערך יותר גבוה "זה יותר טוב") של זוגות מהצורה הבאה:

(object id, attribute value)

. דוגמא - מצגת חיצונית ושקף 18. כמו כן, קיימת פונקציה מונוטונית עולה f , המקבלת את ערכי האטריביוטים. כלומר, לכל i , אם $x_i \leq \bar{x}_i$ אז:

$$f(x_1, \dots, x_i, \dots, x_n) \leq f(x_1, \dots, \bar{x}_i, \dots, x_n)$$

רוצים למצוא את ה- $Top - K$ אובייקטים, כלומר את k האובייקטים עם הערכים הכי גבוהים עבור f .

הפעולות על הרשומות:

- גישה ממוינת (או סדרתית) - קיים מצביע אל המקום ה- i בכל רשימה, כאשר בהתחלה $i = 1$. גישה ממוינת פירושה קריאת המקום ה- i מכל רשימה, וקידום המצביע למקום ה- $(i + 1)$. תמיד בגישה ממוינת קוראים במקביל ערך אחד מכל רשימה, לכן בסה"כ n גישות.
- גישה אקראית - מקבלת id של אובייקט, ומחזירה את הערכים שלו עבור האטריביוטים החסרים. אפשר לבצע גישה אקראית רק אם מצאנו אטריביוט אחד לפחות של האובייקט בגישה ממוינת (אחרת לא יודעים שהוא קיים, ובפרט לא יודעים את ה- id שלו).

למה מפרידים בין שני סוגי הגישות? גישה ממויינת היא זולה, כי היא רק מצריכה קידום של המצביע. גישה אקראית יקרה יותר ומצריכה אינדקס, אחרת (ללא אינדקס) היא יקרה מידי (גישה דרך האינדקס תלויה בגודלו, ושנית מיקום המידע על הדיסק אולי לא המקום שהכי מהיר לקרוא ממנו כרגע). אנו מניחים כי יש גורם קבוע בו גישה אקראית יקרה מגישה ממויינת.

הערה 2.5 גישה מסוג כלשהוא פירושה קריאת ערך אחד מרשימה אחת.

דוגמא בשקופיות החיצוניות: האלגוריתם של פייגין.
באלגוריתמים שנתאר מבצעים בצעד אחד גישה ממויינת מכל אחת מהרשימות (אולם זה לא חייב להיות כך תמיד).

2.2.1 האלגוריתם של Fagin

1. בצע צעדים של גישות ממויינות עד שמתגלים k אובייקטים בשלמותם.

הערה 2.6 בכל ביצוע בודד של צעד 1, עושים גישה ממויינת אחת מכל רשימה.

2. בצע גישות אקראיות להשלמת הערכים החסרים באובייקטים שהתגלו באופן חלקי.

3. מבין האובייקטים שהתגלו, החזר את ה- $Top - K$.

תרגיל 3: הוכח נכונות של האלגוריתם של פייגין, כלומר שהוא באמת מחזיר את ה- $TopK$ מבין כל האובייקטים (לא רק מבין אלו שמצאנו).

בשקופיות 29-31 דוגמא להפעלת האלגוריתם.

הוכחת נכונות האלגוריתם (מתוך הרצאה 6)

הוכחה: עובדות המפתח לנכונות האלגוריתם הן:

- הרשימות של $A_1, \dots, A_n$ ממויינות.
- f מונוטונית.

הגדרה 2.7 יהי $\underline{x}_i$ הערך האחרון שקראנו בגישה ממויינת מרשימה L_i . ערך הסף מוגדר להיות $f(\underline{x}_1, \dots, \underline{x}_n)$.

עובדה: הערך של אובייקט שטרם התגלה הוא לכל היותר ערך הסף, בגלל המונוטוניות - דוגמא להמחשה בשקופית 7 בהרצאה 8.

אובייקט שטרם התגלה נמצא כולו מתחת למקום אליו הגענו בגישה ממויינת.
אובייקט שהתגלה בשלמותו בגישות ממויינות נמצא בכל אחת מהרשימות במקום הכי תחתון שאליו הגענו בגישות ממויינות או מעל למקום זה, לכן, הערך שלו הוא לפחות ערך הסף.
אובייקט שהתגלה באופן חלקי מושלם ע"י גישות אקראיות, ומוחזר כחלק מהתוצאה הסופית רק אם הוא גדול לפחות כמו אחד מאלה שהתגלו באופן מלא. לכן, אובייקט כזה גדול לפחות כמו אלה שלא התגלו בכלל.
אנו מחזירים את ה- $top - k$ מבין אלה שגילינו, ולפי הנאמר לעיל, מובטח שהם גדולים לפחות כמו אלה שלא גילינו בכלל.

■

2.2.2 אלגוריתם הסף - Threshold Algorithm

יהי x_i הערך האחרון שקראנו בגישה ממויינת מרשימה i . ערך הסף הוא: $f(x_1, \dots, x_n)$, כאשר הערכים הם האחרונים שנקראו מכל רשימה.
תיאור באנימציה - בשקופיות הנוספות.

האלגוריתם

• חזור:

- בצע גישה ממויינת מכל רשימה (ועדכן את ערך הסף).
- בדוק תנאי עצירה.
- השלם ערכים חסרים בגישות אקראיות.
- השאר את k האובייקטים הכי טובים.
- בדוק תנאי עצירה.

תנאי העצירה

יש k אובייקטים שלמים (כלומר כל האטריביוטים שלהם ידועים), וערך הפונקציה של כל אחד מהם גדול או שווה לערך הסף. אם התנאי מתקיים, החזר את k האובייקטים השלמים הטובים ביותר.

האלגוריתם נשאר נכון גם אם מוחקים את אחת מבדיקות העצירה בלולאה, אבל אז עושים קצת יותר גישות עד העצירה. בשקפים 36-40 מובאת דוגמא לפעולת האלגוריתם.

אם רוצים למזער את מספר הגישות עוד יותר, אז אפשר לבדוק את תנאי העצירה בכל פעם שמתקבל אובייקט שלם נוסף ע"י גישות אקראיות (הבדיקה של תנאי העצירה פעמיים כאן הוא מטעמי סיבוכיות). עם זאת, לא ברור אם גישה זו באמת תעבוד? נשאר לקורא השקדן להוכיח או להפריך.

תרגיל 4: הראו כי כל אחת משתי הבדיקות של תנאי העצירה איננה מיותרת, כלומר כי ישנם מקרים בהם האלגוריתם יעצור בבדיקה הראשונה, ומקרים בהם האלגוריתם יעצור בבדיקה השנייה.

תרגיל 5: הוכח נכונות אלגוריתם הסף.

2.2.3 השוואה בין שני האלגוריתמים

תרגיל 6: הראו כי האלגוריתם של פייגין עושה לפחות אותו מספר של גישות ממויינות כמו אלגוריתם הסף, והראו שלפעמים יותר.

מה לגבי גישות אקראיות? אי אפשר לדעת.

תרגיל 7: הראו דוגמא בה אלגוריתם הסף עושה יותר גישות אקראיות, ודוגמא בה האלגוריתם של פייגין עושה יותר גישות אקראיות (המצב ההפוך).

מה לגבי הזיכרון שכל אחד מהאלגוריתמים האלה דורש? אלגוריתם הסף מצריך זיכרון קבוע ל- k אובייקטים ועוד ___ אובייקטים נוספים, בעוד האלגוריתם של פייגין מצריך זיכרון לא חסום, כלומר בגודל הנתונים.

תרגיל 8: מצאו חסם מדויק על הזיכרון הנחוץ לאלגוריתם הסף.

נשים לב שאלגוריתם הסף משאיר כל פעם k אובייקטים. השאלה המעניינת:

תרגיל 9: האם אלגוריתם הסף יכול לגלות שוב אובייקט שזרק?

נותר לנו לדבר על עוד שתיים שלוש וריאציות של אלגוריתם הסף. בחצי השני של ההרצאות שנותרו לנו נעבור לנושא חדש, על כן יש לפתור בבית את כל התרגילים הקודמים.

פתרון לשאלה 5 הרצאה 4 - בהנתן יחסים חסרי מעגלים, אם בסוף שני השלבים הראשונים של חצי צירוף אחד היחסים ריק אז כולם ריקים ולהיפך

כיוון אחד טריוויאלי.

בכיוון השני - אם יחס אחד ריק, אז גם השכנים שלו ריקים, כי כל אב ובן הם מח"צ. באופן רקורסיבי מראים שכל היחסים ריקים.

נשים לב כי מח"צ היא איננה תכונה טרנזיטיבית! כלומר, אם R_1 ו- R_2 הם מח"צ, וגם R_2 ו- R_3 הם מח"צ, אז לא בהכרח R_1 ו- R_3 הם מח"צ.

תרגיל 10: הראו זאת.

אבל אם היחסים הם חסרי מעגלים, אז כל שניים הם מח"צ בגלל שתי סיבות:

1. כל אב ובן הם מח"צ.

2. תכונת המופע הרציף.

תכונת המופע הרציף היא זו שהופכת את מח"צ לטרנזיטיבית על המסלול המשותף המחבר כל שניים.

בחזרה לאלגוריתם הסף.

6/5/2013

הרצאה 6

תרגיל 1: כמה גישות אקראיות נוספות עושים לכל היותר אם מבטלים את הבדיקה הראשונה של תנאי העצירה?

נכונות אלגוריתם הסף

תחילה נניח שיש רק בדיקת עצירה אחת בסוף הלולאה.

האלגוריתם מחזיר k אובייקטים שגדולים לפחות כמו ערך הסף, ולכן גם גדולים לפחות כמו כל אובייקט שטרם התגלה.

כמו כן, האובייקטים המוחזרים גדולים לפחות כמו כל אובייקט שהתגלה אבל לא נשמר (כי לא היה מבין ה- $(top - k)$).

לכן האלגוריתם מחזיר את ה- $top - k$.

יש לציין כי בכל פעם שמשנים את הקבוצה של ה- $top - k$ ששומרים, אזי הערך המינימאלי של הפונקציה מעל קבוצה זו רק יכול לגדול. לכן, האיברים בקבוצה זו לא רק גדולים לפחות כמו אלה שהושמטו באיטרציה הנוכחית, אלא גם כמו אלה שהושמטו באיטרציות קודמות (בגלל הטרנזיטיביות).

תרגיל 2: הוכיחו את נכונות אלגוריתם הסף כשיש שתי בדיקות עצירה בלולאה.

תשובה לשאלה 6 מהרצאה 5 - הראו כי האלגוריתם של פייגין עושה לפחות אותו מספר של גישות ממויינות כמו אלגוריתם הסף, והראו שלפעמים יותר

כשהאלגוריתם של *Fagin* עוצר, יש k אובייקטים שהתגלו בשלמותם בגישות ממויינות. לכן, הערך של כל אחד הוא לפחות ערך הסף. כלומר, תנאי העצירה של אלגוריתם הסף מתקיים. לכן, מספר הגישות הממויינות של האלגוריתם של *Fagin* הוא לפחות כמו זה של אלגוריתם הסף.

תשובה חלקית לשאלה 7 מהרצאה 5 - דוגמא למקרה בו אלגוריתם הסף עושה פחות גישות אקראיות מאלגוריתם הסף

הדוגמא מופיעה בשקופית 17: כל האובייקטים הדרושים מופיעים בשלמותם לאחר 3 גישות ממויינות, ואין צורך בגישות אקראיות בכלל. לעומת זאת, אלגוריתם הסף כן מבצע גישות כאלה בדוגמא (כבר לאחר הגישה הממויינת הראשונה).

2.2.4 אלגוריתם הסף ללא גישות אקראיות (NRA)

שומרים את כל האובייקטים שמתגלים בגישות ממויינות.

הגדרה 2.8 יהי o אובייקט כלשהוא. הערך הגרוע ביותר של o הוא:

$$w(o) := f(v_1, v_2, \dots, v_n)$$

כאשר v_i הוא הערך שהתגלה ל- A_i , או אפס אם לא התגלה ערך ל- A_i .
הערך הטוב ביותר של o הוא:

$$B(o) := f(v_1, v_2, \dots, v_n)$$

כאשר v_i הוא הערך של o שהתגלה ל- A_i , או ערך הסף $\underline{x}_i$ אם לא התגלה ערך ל- A_i .

האלגוריתם

• חזור:

- בצע גישה ממויינת אחת מכל רשימה.
- בדוק תנאי עצירה.

• עצור

תנאי העצירה

יש לפחות k אובייקטים שהתגלו, ומתקיים הדבר הבא:
 ממיינים בסדר יורד את כל האובייקטים שהתגלו לפני ערכי ה- w שלהם. אם יש שניים עם אותו w , ממיינים אותם לפי ערך ה- B . אם גם ה- B אותו דבר, ממיינים אותם באופן שרירותי.
 יהי M ערך ה- w הכי קטן מבין ה- $top - k$ במיון זה. אם M גדול לפחות כמו ה- B של כל אלה שאינם בין ה- $top - k$ הנ"ל, החזר את ה- $top - k$ הנ"ל. אחרת, המשך לאיטרציה הבאה.

הערה 2.9 נשים לב שה- B של אלה שלא התגלו בכלל הוא בדיוק ערך הסף.

כשעוצרים, מחזירים את ה- $top - k$ לפי ערכי ה- w שלהם.

שאלה 3: האם הסדר הנכון בין ה- $top - k$ שהוחזרו הוא בהכרח כמו במיון שלהם לפי ערכי ה- w ?
שאלה 4: הוכח נכונות של אלגוריתם NRA .
שאלה 5: בתיאור של תנאי העצירה, נאמר שאם לשניים יש אותו w , אז ממיינים אותם לפי B . מדוע זה חשוב?
 האם זה נחוץ לנכונות? אם לא, למה צריך את זה?
שאלה 6: עבור כל אחת מהפונקציות MAX, MIN, SUM , האם תמיד אפשר לדעת מה הערך של כ"א מהאובייקטים המוחזרים?

2.2.5 אלגוריתם משולב - Combined Algorithm**סימונים והגדרות**

נסמן:

- C_R הוא המחיר של גישה אקראית ($Random$).
- C_S הוא המחיר של גישה ממויינת ($Sorted$).
- A היא משפחה של אלגוריתמים.
- D היא קב' של קלטים אפשריים.

Instance Optimality

הגדרה 2.10 אלגוריתם $b \in A$ הוא אופטימלי לכל מופע, אם לכל אלגוריתם $a \in A$ ולכל קלט $d \in D$ מתקיים:

$$cost(b, d) = O(cost(a, d))$$

כלומר:

$$cost(b, d) \leq c \cdot cost(a, d) + c'$$

כאשר c תלוי ב- n (מספר האטריביוטים), וכן תלוי גם ביחס C_R/C_S .

הערה 2.11 אפשר גם להראות (זה לא קל) שאלגוריתם הסף הוא אופטימלי לכל מופע ביחס למשפחה גדולה של אלגוריתמים לחישוב $top - k$ שפועלים בעזרת גישות ממויינות ואקראיות.

הערה 2.12 אלגוריתם NRA הוא אופטימלי לכל מופע ביחס למשפחה של אלגוריתמים ל- $top - k$ שעושים רק גישות ממויינות.

האלגוריתם המשולב - CA אלגוריתם זה עושה פחות גישות אקראיות ככל שהיחס C_R/C_S גדול יותר, ומהווה ואריאציה של NRA .

האלגוריתם עושה גישות אקראיות בכל איטרציה שמספרה הוא כפולה של $\lfloor C_R/C_S \rfloor$. באיטרציה כזו נבחר, מבין כל האובייקטים שהתגלו רק באופן חלקי ושעבורם $B > M$, את זה שיש לו ערך B הכי גבוה, ונשלם אותו ע"י גישות אקראיות. כלומר, עושים גישות אקראיות לכל היותר לאובייקט אחד. כל השאר באלגוריתם CA הוא כמו ב- NRA . את הפעולה של ביצוע הגישות האקראיות מוסיפים לאלגוריתם NRA אחרי השורה של ביצוע הגישות הממויינות ולפני השורה של בדיקת תנאי העצירה. כאמור, עושים גישות אקראיות רק פעם אחת כל $\lfloor C_R/C_S \rfloor$ איטרציות, וגם אז עושים זאת רק על אובייקט אחד (אם הוא קיים) שנבחר כפי שתיארנו.

אפשר להראות כי CA הוא אופטימלי לכל מופע ביחס לאלגוריתמים שמוצאים $top - k$ ע"י גישות ממויינות ואקראיות, כך שהקבוע c תלוי ב- n בלבד ולא ב- C_R/C_S . למה משלימים את האובייקט עם ה- B הכי גבוה, ולא לפי w הכי קרוב למעלה לערך הסף? אם משלימים כך, מקטינים את הסיכוי שנעשה קריאה אקראית על אובייקט שיימחק, ולכן הקריאות היקרות יתבזזו (ה"פוטנציאל" של האובייקט הזה לא להמחק יותר גבוה מזה לפי w , מההגדרות).

שאלה 7: נניח שרק על תת קבוצה נתונה של רשימות אפשר לעשות גישות ממויינות. על האחרות - רק גישות אקראיות. תארו ואריאציה של אלגוריתם הסף המתאימה למקרה זה.

2.3 Top - K מקורב

הגדרה 2.13 קירוב θ ($\theta \geq 1$) של $Top - K$ הוא תת קבוצה T של k אובייקטים כך שלכל $o \in T$ ולכל $o' \notin T$ מתקיים:

$$\theta f(o) \geq f(o')$$

שאלה 8: תארו ואריאציה של אלגוריתם הסף שמוצאת קירוב θ של ה- $top - k$. כמובן שצריך לעצור מוקדם ככל האפשר.

פתרון תרגילים

שאלה 1 מהרצאה 5 - שימוש באלגוריתם של לולר מורטי

$$\begin{aligned} I, E, t_1 \bowtie t_2 \bowtie \dots \bowtie t_n \\ I = \{t_1, t_2\} \\ I_1 = \{t_1, t_2\}, E_1 = \{t_3\} \cup E \\ I_2 = \{t_1, t_2, t_3\}, E_2 = \{t_4\} \cup E \\ \vdots \end{aligned}$$

מה מיותר? אם E אומר שמישהו לא יכול להיות, אז יש שתי אפשרויות: או ש- I_1 מכילה את אותה הרשומה, ואז זה לא יכול להיות ואין פתרון, ואם ב- I_1 מופיעה רשומה של אותו היחס אבל שונה, אז ממילא יכולה להיות רק רשומה אחרת, ואז הרשומה ב- E מיותרת. לכן, רק מהיחסים שלא מוזכרים ב- I יכולות להיות רשומות שלא מיותרות. בהנחה שתמיד בונים את האילוצים ב- I מ-1 ל-2 וכן הלאה, אז יש $n - j$, אבל האם אפשר פחות מזה? כן! בגלל סדר הבניה, הרשומות של אילוצי ההדחה שכן יכולות להשפיע הן תמיד מיחס אחד בלבד R_{k+1} , ויכול להיות מספר כלשהוא של רשומות כאלה (הסבר קצת יותר מפורט מופיע בשקפים באתר, בסוף הרצאה 7).

שאלה 3 הרצאה 6 - אלגוריתם הסף ללא גישות אקראיות (NRA): האם הסדר הנכון בין ה- $top - k$ שהוחזרו הוא בהכרח כמו במיון שלהם לפי ערכי ה- w ?

לא - נראה ע"י דוגמא נגדית: נביט במקרה בו מחפשים $top - 2$, $f = AVG$, והרשימות הן:

L_1	L_2
$(a, 1.0)$	$(c, 0.9)$
$(b, 0.3)$	$(b, 0.3)$
$(c, 0.3)$	$(a, 0)$

	L_1	L_2	w
a	1	-	0.5
c	-	0.9	0.45
b	0.3	0.3	0.3

תנאי העצירה מתקיים באיטרציה השניה, ואז ל- a ערך יותר גרוע מזה של c , אבל הסדר האמיתי הפוך.

שאלה 1 הרצאה 6 - שיפור אלגוריתם הסף: כמה גישות אקראיות נוספות עושים לכל היותר אם מבטלים את הבדיקה הראשונה של תנאי העצירה?

כמה מרוויחים גישות אקראיות פחות נעשה בין האלגוריתם עם עצירה אחת לבין האלגוריתם עם שתי העצירות? יהי n מספר האטריביוטים (רשימות). כשיש שתי בדיקות עצירה, הראשונה היא לפני השלמת ערכים חסרים בגישות אקראיות, והשניה מיד לאחר השלמה. המספר הגדול ביותר של ערכים חסרים שיכול להתקבל בעת גישה ממוינת (אחת מכל רשימה) הוא כאשר קוראים n אובייקטים חדשים, כאשר לכל אחד חסרים $n - 1$ ערכים. לפיכך, ביצוע שתי בדיקות עצירה (לעומת אחת) חוסך לכל היותר $n(n - 1)$ גישות אקראיות.

שאלה 6 הרצאה 6 - עבור כל אחת מהפונקציות MAX , MIN , SUM , האם תמיד אפשר לדעת מה הערך של כ"א מהאובייקטים המוחזרים?

אם הפונקציה היתה SUM ולא AVG , היה יכול לקרות לנו אותו המצב מהדוגמא בשאלה 3. מה לגבי MIN ? גם לא! לגבי MAX דווקא אפשר: כל הערכים שלא ידועים ה- MAX שלהם לכל היותר הוא ערך הסף, וכל אלה שעדיין לא גילינו, הערך המקסימלי שלהם הוא ערך הסף. על כן, אפשר לדעת מהו הערך של האובייקטים המוחזרים מבלי למצוא את הערכים החסרים. כלומר - בדר"כ התשובה היא לא, אבל במקרה של MAX התשובה היא כן.

שאלה 7 הרצאה 6 - נניח שרק על תת קבוצה נתונה של רשימות אפשר לעשות גישות ממויינות. על האחרות - רק גישות אקראיות. תארו ואריאציה של אלגוריתם הסף המתאימה למקרה זה.

רק חלק מהעמודות מסודרות בסדר יורד, ולכן אפשר להשתמש בגישה ממויינת רק בחלק מהעמודות, ובחלק מבצעים רק גישה אקראית. השאלה היא, איך משנים את האלגוריתם כך שיתאים למצב הזה? עדיין מדובר על האלגוריתם הבסיסי. כל השינוי בהגדרה הוא בתנאי העצירה: ערך הסף החדש יהיה הערך האחרון בעמודות שבהן יש גישה ממויינת, ובשאר העמודות שמם 1. אלגוריתם זה יעבוד, אבל יקח לו הרבה זמן לעצור...

שאלה 8 הרצאה 6 - קירוב של $Top - K$: איך משנים את אלגוריתם הסף הרגיל כך שימצא מהר ככל האפשר קירוב θ בהנתן θ כלשהוא?

כדי שאלגוריתם הסף הרגיל ימצא קירוב θ מהר ככל האפשר, צריך לבצע את השינוי הבא: בעצ בדיקת תנאי העצירה, צריך לקחת ערך הסף חלקי θ .

3 אלגוריתמים לבדיקת שקילות והכלה של שאילות

3.1 שאילות גימום

13/05/2013
הרצאה 7

מתי שאילות שקולות? אם לכל מסד נתונים, השאילות יתנו בדיוק את אותה התוצאה.
בקורס במסדי נתונים דיברנו על תכונה זו קצת - היו זהויות שבעזרתן היינו יכולים במקרים מסויימים לקבוע שקילות, למשל היכולת להכניס פרוג'קשן או סלקשן. במסגרת קורס זה נדבר על אלגוריתמים כלליים, ולא על זהויות פרטניות.
נתמקד בתחילה בשאילות גימום - Conjunction queries - אפשר להראות שזה "אותו דבר" כמו שאילות SPJ - Select, Project, Join.

3.1.1 הגדרות בסיסיות

נכתוב את השאילות האלה בצורה שהיא שונה ממה שראינו בקורס מסדי נתונים - כחוקי הורן, כתיבה שנגזרת מלוגיקה מסדר ראשון, במקום לכתוב אותה באמצעות אופרטורים של אלגברה.

אטומים מה זה אטום?

הגדרה 3.1 אטום נכתב כך:

$$R(X, I, Y)$$

כאשר:

- R הוא שם של יחס (פרדיקט).
- בעמודות (X, I, Y) יש סימנים (*symbols*), כאשר סימן הוא משתנה או קבוע.
- כל העמודות של R מופיעות.
- אפשר לשים במקום המשתנים קבועים כלשהם.
- לכל פרדיקט (שם של יחס) יש מספר קבוע של עמודות. לפרדיקטים שונים יכול להיות מספר שונה של עמודות.

אפשר לחשוב על אטום כעל רשומה במסד הנתונים.
תחילה נדבר על סוג מיוחד של אטומים, והם אטומים קרקעיים:

הגדרה 3.2 אטום הוא קרקעי (ground atom) אם בכל העמודות יש קבוע.

הגדרה 3.3 מסד נתונים הוא קבוצה סופית של אטומים קרקעיים.

חוק הורן

הגדרה 3.4 בצורה הכללית, חוק הורן יכתב באופן הבא:

$$H : -B$$

כאשר H מורכב מאטום אחד, הנקרא הראש, ו- B היא קבוצה של אטומים (במשמעות של גימום) הנקראת הגוף, הסימן - : מסמל גרירה בכיוון שמאל.

דוגמא:

$$R(X, Y) : -T(X, Z), T(Z, Y)$$

כאשר הראש הוא $R(X, Y)$ במקרה שלנו, ו- $T(X, Z), T(Z, Y)$ הם הגוף. נשים לב כי בכל חוק הורן יופיעו פסיקים בין האטומים בגוף, גאשר הפסיק מסמן $\wedge$ (וגם).

הערה 3.5 אנחנו מניחים, אלא אם כן נאמר במפורש אחרת, כי המשתנים בראש מופיעים גם בגוף - הגבלה זו נקראת "תנאי בטיחות" (נחזור בהמשך ונבין מהיכן מגיע השם הזה).

הערה 3.6 נניח לשם פשטות (אלא אם כן נאמר אחרת) כי הפרדיקט בראש שונה מאלו שבגוף.

כדי שחוק שכזה יחשב משהו, צריך מסד נתונים - שהוא כאמור אוסף סופי של אטומים קרקעיים. כיצד, בהנתן מסד נתונים, מייצרים ממנו בעזרת החוק תוצאה? נציב עבור המשתנים קבועים, ונבדוק אם תוצאת ההצבה מופיעה במסד הנתונים, וממה שנקבל נבנה אטומים קרקעיים של R , והם יהיו התוצאה. בשביל הגדרה פורמלית, עלינו להכיר מספר מושגים נוספים.

הגדרה 3.7 מיפוי סימנים הוא פונקציה f הממפה משתנה לסימן וקבוע לעצמו.

דוגמא - שקופית 6:

$$f(x) = y, f(z) = w, f(u) = 5, f(2) = 2$$

הפעלת מיפוי סימנים על אטום מוגדרת באופן טבעי כהפעלת המיפוי על כל סימן בעמודות האטום. תחת אותו מיפוי הסימנים נקבל:

$$f(R(x, u, 3, z, 2)) = R(f(x), f(u), \dots) = R(y, 5, 3, w, 2)$$

הגדרה 3.8 מיפוי סימנים קרקעי הוא פונקציה שממפה את כל המשתנים לקבועים.

הגדרה 3.9 בהנתן מסד נתונים $\mathcal{D}$, התוצאה של חוק הורן $-B$ היא הקבוצה:

$$\{g(H) \mid g \text{ is a ground mapping s.t. } \forall b \in B \ g(b) \in \mathcal{D}\}$$

הערה 3.10 המסד $\mathcal{D}$ סופי ולכן יש בו מספר סופי של קבועים. מספיק לבדוק את כל המיפויים מסימני החוק לקבועים במסד. יש מס' סופי של מיפויים כאלה. לכן, התוצאה סופית וניתנת לחישוב בזמן סופי, כי מתקיים תנאי הבטיחות - כל משתנה בראש מופיע גם בגוף.

3.1.2 הגדרת שקילות והכלה, ומציאת תנאים מספקים

נרצה להגדיר שקילות והכלה באופן פורמלי. לשם כך, נביט בדוגמא:

$$Q_1 R(X, Y) : -T(X, Z), T(Z, Y), T(Z, W)$$

$$Q_2 R(X, Y) : -T(X, Z), T(Z, Y)$$

האטום השלישי ב- Q_1 לא מוסיף שום תנאי חדש - אם מצליחים למצוא אטום המקיים $T(X, Z), T(Z, Y)$ הוא גם יקיים את $T(Z, W)$. מדוע? כדי להסביר זאת, נגדיר סימון נוסף:

הגדרה 3.11 בהנתן שאילתא (למשל, חוק הורן) Q ומסד $\mathcal{D}$, נסמן ב- $Q(\mathcal{D})$ את התוצאה של Q עבור $\mathcal{D}$.

הגדרה 3.12 שאילות Q_1 ו- Q_2 הן שקולות אם לכל מסד נתונים $\mathcal{D}$ מתקיים $Q_1(\mathcal{D}) = Q_2(\mathcal{D})$. שקילות תסומן $Q_1 \equiv Q_2$.

הגדרה 3.13 שאילתא Q_1 מכילה את שאילתא Q_2 אם לכל מסד נתונים $\mathcal{D}$ מתקיים $Q_1(\mathcal{D}) \supseteq Q_2(\mathcal{D})$. הכלה תסומן $Q_1 \supseteq Q_2$.

עובדה שנובעת בקלות מההגדרות:

$$Q_1 \equiv Q_2 \Leftrightarrow Q_1 \supseteq Q_2 \wedge Q_2 \supseteq Q_1$$

נכתוב את הכלל השני מקודם בצורה שונה:

$$Q_2 R(U, V) : -T(V, U), T(W, U)$$

האם עדיין ברור שיש הכלה או שיוויון בין השאילות? אפשר לראות זאת אם ממפים את המשתנים של Q_2 ל- Q_1 . אבל, מיפוי זה צריך לענות על תנאים מסויימים:

- הראש של Q_2 ממופה לראש של Q_1 . מיפוי זה כבר מחליט לנו על חלק מהמשתנים.
 - את שאר המשתנים ניתן למפות למשתנה המתאים ב- Q_1 .
- לאחר מציאת מיפוי שכזה, אנו יכולים לראות את ההכלה. לפי האינטואיציה הזו נביע פורמלית תנאי מספק להכלה.

הגדרה 3.14 נסמן:

$$Q_1 : H_1 : -B_1$$

$$Q_2 : H_2 : -B_2$$

מיפוי h מהסימנים של Q_1 לסימנים של Q_2 נקרא הומומורפיזם אם:

$$1. h(H_1) = H_2$$

$$2. h(B_1) \subseteq B_2, \text{ כלומר לכל אטום } b \in B_1 \text{ מתקיים ש-} h(b) \text{ שייך ל-} B_2.$$

תנאי מספיק להכלה

למה 3.15 אם יש הומומורפיזם מ- Q_1 ל- Q_2 , אז $Q_1 \supseteq Q_2$.

נדגים את ההוכחה על הדוגמא שראינו:

$$Q_1 : R(X, Y) : -T(X, Z), T(Z, Y), T(Z, W)$$

$$Q_2 : R(V, U) : -T(V, W), T(W, U)$$

הפעם נחפש את ההומומורפיזם בכיוון ההפוך. נבנה אותו:
תחילה, נעביר את הראש: $X \rightarrow V, Y \rightarrow U$.
לאחר מכן, נעביר את Z ל- W . ולבסוף, נעביר את W ל- U .
בצורה כזו קיבלנו מיפוי סימנים h . אם מאמינים לטענה אז נקבל שקילות (כי הראנו הומומורפיזם בשני הכיוונים).
אבל, עלינו להראות שזהו אכן תנאי מספיק! איך מקיום הומומורפיזם אפשר להראות שיש הכלה? יש לקחת מסד $\mathcal{D}$ כלשהוא, למשל:

$$\mathcal{D} = \{T(1, 2), T(2, 3), \dots\}$$

ולהראות שכל תוצאה ש- Q_2 מייצרת, גם Q_1 מייצרת. מה למשל Q_2 שלנו יכול לייצר במסד הנתונים הזה? נביט למשל במיפוי הקרקעי g הממפה $U \rightarrow 3, W \rightarrow 2, V \rightarrow 1$. במקרה הזה, $Q_2(\mathcal{D}) \ni R(1, 3)$. כלומר, עלינו להראות כי $Q_1(\mathcal{D}) \ni R(1, 3)$.

איך נעשה זאת? נבצע הרכבה על המיפויים הללו - $g \circ h$, ואז נקבל תוצאה המקבלת ב- Q_1 , כנדרש.
וכעת, באופן פורמלי: **הוכחה:** יהי h הומומורפיזם מ- Q_1 ל- Q_2 . יהי $\mathcal{D}$ מסד נתונים, ונניח שהמיפוי הקרקעי g מייצר את $g(H_2)$ מ- Q_2 ומ- $\mathcal{D}$.
ההרכבה $g \circ h$ הינה מיפוי קרקעי המוגדר על סימני Q_1 ומתקיים ש:

$$1. \text{ כל אטום בגוף } B_1 \text{ ממופה ע"י } g \circ h \text{ לאטום קרקעי ב-} \mathcal{D}.$$

$$2. \text{ לפי הגדרת הומומורפיזם, } h(H_1) = H_2, \text{ ולכן } g \circ h(H_1) = g(H_2), \text{ כלומר } g \circ h \text{ מייצר מהראש של } Q_1 \text{ אותו אטום קרקעי של } \mathcal{D} \text{ כמו ש-} g \text{ מייצר מהראש של } Q_2.$$

לכן הראנו שבהנתן דרך לקבל אטום בתוצאה של הפעלת Q_2 על $\mathcal{D}$, אפשר גם לקבל אטום זה ע"י הפעלת Q_1 על $\mathcal{D}$.

לפיכך, $Q_1(\mathcal{D}) \supseteq Q_2(\mathcal{D})$, ומכיוון שזה נכון לכל $\mathcal{D}$, מתקיימת הכלה, כלומר $Q_1 \supseteq Q_2$. ■
איך מראים את הכיוון ההפוך, כלומר שאם יש הכלה, יש גם הומומורפיזם? זה "טיפה" יותר קשה. ניקח $\mathcal{D}$ אחד ספציפי שמזה שיש עליו הכלה, נובע שיש הומומורפיזם. איך נייצר $\mathcal{D}$ שכזה?
נביט שוב בשני הכללים ממקודם:

$$Q_1 : R(X, Y) : -T(X, Z), T(Z, Y), T(Z, W)$$

$$Q_2 : R(U, V) : -T(V, U), T(W, U)$$

ניקח את הגוף של זה שהוא מוכל, ונייצר ממנו **מסד קנוני**. מה זה אומר? נמפה כל משתנה לאיזשהו קבוע מיוחד, אותו נסמן באינדקס אפס, למשל V_0 הוא הקבוע המיוחד עבור V , W_0 הוא הקבוע המיוחד עבור W , וכו'. נעשה זאת לכל אטום בגוף. קיבלנו את המסד:

$$\mathcal{D} = \{T(V_0, W_0), T(W_0, U_0)\}$$

בהנתן המסד הקנוני, ברור מאלינו כי Q_2 מייצר ממנו את הראש הקנוני $R(V_0, U_0)$, זאת על ידי המיפוי הקנוני g , שהוא מיפוי קרקעי הממפה כל משתנה לקבוע היחודי שלו. עכשיו, נתון שיש הכלה, כלומר גם מ- Q_1 ניתן לייצר את הראש הקנוני, וזאת ע"י מיפוי קרקעי f הפועל באופן הבא:

$$X \rightarrow V_0, Z \rightarrow W_0, Y \rightarrow U_0$$

נשים לב כי g היא חח"ע ועל, ולכן קיימת לה פונקציה הפוכה g^{-1} . אזי, כפי שניתן לקרות בשקופית 18, קיבלנו את ההרכבה של $g^{-1} \circ f$, המהווה הומומורפיזם מ- Q_1 ל- Q_2 .
נותר אם כך רק לפרמל את ההוכחה: **הוכחה:** מהגוף של Q_2 מייצרים מסד קנוני ע"י כך שמחליפים כל משתנה בקבוע ייחודי X_0 ששונה מכל קבוע אחר (שכן ייתכן שקבוע כלשהו מופיע בחוק, לא ראינו זאת בדוגמאות אבל זה אפשרי). אם בגוף של Q_2 יש אטום מהצורה $T(X, I, Y)$ אז הוא עובר לאטום הקנוני $T(X_0, I, Y_0)$, כלומר הקבוע I נשאר כמו שהוא והקבועים הייחודיים X_0 ו- Y_0 שונים זה מזה וגם מ- I . מיפוי הסימנים g שממפה כל משתנה של Q_2 לקבוע הייחודי המתאים לו, מראה ש- Q_2 מייצר את $g(H_2)$ מהמסד הקנוני, שיסומן ע"י $\mathcal{D}$.

הערה 3.16 נשים לב כי חייבים להקפיד שכל הקבועים יהיו שונים זה מזה וגם מקבועים אחרים, על מנת ש- g תהיה חח"ע ועל כן g^{-1} מוגדרת היטב.

היות ויש הכלה, גם Q_1 מייצר את $g(H_2)$ מ- $\mathcal{D}$. לפיכך, יש מיפוי f מהסימנים של Q_1 לסימנים של $\mathcal{D}$ כך שלכל אטום b בגוף של Q_1 מתקיים ש- $f(b) \in \mathcal{D}$ ובנוסף $f(H_1) = g(H_2)$.
קל להראות שההרכבה $g^{-1} \circ f$ היא הומומורפיזם מ- Q_1 ל- Q_2 .
לפיכך, הראנו שאם יש הכלה $Q_1 \supseteq Q_2$ אז יש הומומורפיזם מ- Q_1 ל- Q_2 . ■

תרגיל 1: מצאו אלגוריתם למינימיזציה של חוק הורן, כלומר מחיקת אטומים מיותרים מהגוף. הערה: "מיותרים" פרושו שהשקילות נשמרת לאחר מחיקתם.

תרגיל 2: האם הגוף בחוק הורן יכול להיות ריק? אם כן, מה התוצאה של הפעלת החוק בהנתן מסד $\mathcal{D}$?

תרגיל 3: הראו כיצד מוצאים ביטוי שקול באלגברת היחסים לחוק הורן כללי מהצורה:

$$H : -B$$

בשקופית 1 נמצא סיכום גרפי יפה מהשיעור הקודם:

ראינו שאם קיים הומומורפיזם $h : Q_1 \rightarrow Q_2$ אז מתקיימת ההכלה $Q_1 \supseteq Q_2$, כי לכל g שמייצר אטום לתוצאה של Q_2 , $g \circ h$ ייצר אותו לתוצאה של Q_1 .
איך משנים את התמונה בשקופית 1 כדי לקבל את הכיוון ההפוך? שקופית 3 - בהנתן Q_2 , ע"י איזשהיא פונקציה g מייצרים מסד קנוני מהגוף של Q_2 . איך מייצרים מסד קנוני זה, דהיינו מיהי g ? g היא מיפוי חח"ע ועל שמייצר את המסד הקנוני ע"י כך שממפים כל משתנה לקבוע ייחודי.
לכן, g גם מייצרת את $g(H_2)$ לתוצאה של Q_2 בהנתן המסד הקנוני.
מההנחה כי $Q_1 \supseteq Q_2$, קיים מיפוי סימנים קרקעי f שמייצר את אותו אטום לתוצאה של Q_1 .
על כן, $g^{-1} \circ f$ הוא הומומורפיזם מ- Q_1 ל- Q_2 .
למעשה, הראנו שהכלה $Q_1 \supseteq Q_2$ שקולה לכך שעבור המסד הקנוני שנוצר מהגוף של Q_2 , Q_1 מסוגלת לייצר ממסד זה את הראש הקנוני של Q_2 .

קבוצה\איחוד של חוקי הורן תהי $U = \{Q_1, \dots, Q_n\}$ קבוצה של חוקי הורן (או איחוד).

הגדרה 3.17 התוצאה של U בהנתן מסד $\mathcal{D}$, סימון $U(\mathcal{D})$, היא איחוד התוצאות:

$$\bigcup_{i=1}^n Q_i(\mathcal{D})$$

אם כך, השאלה הטבעית היא - בהנתן שתי קבוצות כאלה, שכל אחת מציינת את האיחוד של השאליות של הקבוצה, מה תהיה ההגדרה של הכלה, האם היא משתנה באיזושהיא אופן? לא; כלומר, קבוצה אחת מכילה קבוצה שניה אם התוצאה של אחת לכל מסד מכילה את התוצאה של הקבוצה השניה:

הגדרה 3.18 יהיו U_1 ו- U_2 איחודים של חוקי הורן. U_1 מכיל את U_2 , סימון $U_1 \supseteq U_2$, אם לכל מסד D :

$$U_1(D) \supseteq U_2(D)$$

איך עכשיו נמצא כלל מספיק והכרחי להכלה?

משפט 3.19 $U_1 \supseteq U_2$ אם ורק אם לכל $Q_i \in U_2$ קיים $Q_j \in U_1$ כך ש: $Q_j \supseteq Q_i$.

הערה 3.20 בהתאם לפיתוח מהשיעור הקודם, אפשר לנסח את המשפט הזה באופן השקול הבא: אם ורק אם לכל $Q_i \in U_2$ יש $Q_j \in U_1$ כך שקיים הומומורפיזם:

$$h: Q_j \rightarrow Q_i$$

הוכחה: כיוון אחד להוכחת המשפט, אם, הוא די טריוויאלי: אם כל $Q_i \in U_2$ מוכל באחת השאליות של U_1 , אז ברור שהאיחוד של השאליות ב- U_2 מוכל בזה של U_1 .

כעת לכיוון השני: נתון כי $U_1 \supseteq U_2$, ויש להראות כי לכל $Q_i \in U_2$ קיימת שאלית $Q_j \in U_1$ שמכילה את Q_i . אנחנו צריכים למצוא ל- Q_i מסד מסויים, כך שממנו Q_i מייצר אטום אחד מסויים, ומתוך זה שנבחר את המסד בצורה מאוד מסויימת נקבל הכלה (כי מההכלה המקורית חייב להיות מישו שמיצר את אותו אטום לתוצאה, ואז מקבלים הומומורפיזם, דהיינו ההכלה המתבקשת). השאלה היא, מיהו המסד ומיהו האטום המסויים. תהי Q_i שאלית של U_2 . יהי g מיפוי סימנים שמייצר מהגוף של Q_i את המסד הקנוני, שיסומן ע"י $\mathcal{D}_i$. לכן:

$$g(H_i) \in Q_i(\mathcal{D}_i)$$

כאשר H_i הוא הראש של Q_i . לפיכך, מההנחה, יש שאלית Q_j ב- U_1 שגם מייצרת את $g(H_i)$ מ- $\mathcal{D}_i$. כמו בשקף 3, נובע מכך שיש הומומורפיזם מ- Q_j ל- Q_i ולכן $Q_j \supseteq Q_i$. מ.ש.ל. ■

3.2 שאילתות שיש בהן השוואות

השלב הבא הוא לדבר על שאילתות שיש בהם השוואות. לפני כן, צריך להגדיר באופן פורמלי מהו יחס הסדר.

3.2.1 תכונות של יחס סדר מלא, צפוף ולא חסום

יחס סדר הוא יחס המקיים את התכונות הבאות:

1. טרנזיטיביות, כלומר:

$$X < Y, Y < Z \Rightarrow X < Z$$

2. אי-רפלקסיביות, כלומר לכל X , $X \not< X$.

3. אי-סימטריות, כלומר אם $X < Y$ אז $X \not< Y$.

• יחס סדר הינו מלא, אם לכל X ו- Y שונים זה מזה, מתקיים $X < Y$ או $Y < X$.

• יחס סדר הינו צפוף, אם לכל $X < Y$ קיים Z כך ש- $X < Z < Y$.

איזה יחס סדר למשל לא מקיים את זה? הטבעיים.

• יחס סדר הוא לא חסום, אם לכל X קיימים Y ו- Z כך ש:

$$Y < X < Z$$

דוגמא ליחס סדר מלא, צפוף ולא חסום היא המספרים הממשיים (השלמים לא, שכן הם לא צפופים).

כדי שנוכל לדבר על הכלה בין שאילתות שיש בהם השוואות, יש לפתור בעיה שלא קשורה לשאליות בשום צורה כלשהיא, אלא ליחס הסדר עצמו.

3.2.2 בעיית הגרירה

נביט בדוגמא (כאשר מדובר ביחס שמקיים את כל התכונות שדנו בהן):

$$X < Y, 3 < W, W < Z \stackrel{?}{\models} Z < 3$$

כאשר $\models$ מסמל גרירה, וכ"א מהאטומים בצד שמאל הוא אטום של השוואה. השאלה היא - האם הגרירה מתקיימת?

ברור שאטומים של השוואות מורכבים ממשתנים ומקבועים ששייכים לתחום שעליו מוגדר יחס הסדר, שמקיים כל התכונות שהוזכרו קודם. בדוגמאות שלנו, התחום הוא המספרים הרציונליים. כאן בעצם נובעת השאלה - כיצד מגדירים את הגרירה? הרי אפשר להגדיר אותה בכל אופן כל שהגרירה מעלה, שלמרות שבמונחים "אינטואיטיביים" לא מתקיימת, תתקיים.

אין גרירה אם יש דוגמא נגדית!

בעיית הגרירה בצורתה הכללית:

$$\Gamma \models \gamma$$

כאשר Γ היא גימון של השוואות, ו- γ היא השוואה בודדת, כאשר כל השוואה היא מהצורה $S_i < S_j$, כאשר S_i, S_j הם סימנים, כלומר משתנים או קבועים. למעשה, נניח שכל השוואה (ב- Γ או ב- γ) היא מהצורה θ , כאשר θ היא אחת מפעולות ההשוואה הבאות:

$$<, \leq, =, \neq, >, \geq$$

הגדרה 3.21 דוגמא נגדית לגרירה $\Gamma \models \gamma$ היא מיפוי סימנים קרקעי לתחום שעליו מוגדר יחס הסדר, כך שכל השוואה של Γ מתקיימת ו- γ איננה מתקיימת.

נחזור לדוגמא הקודמת, ונשנה אותה קצת כדי שלא ישעמם לנו:

$$X < Y, 3 < W, W < Z \models Z < 5$$

נבנה לה דוגמא נגדית - נגדיר מיפוי:

$$X \rightarrow 1, Y \rightarrow 2, 3 \rightarrow 3, W \rightarrow 4, Z \rightarrow 5$$

אזי צד שמאל מתקיים, אבל ימין לא, כנדרש.

הערה 3.22 אם מיפוי אינו מקיים כל ההשוואות בצד שמאל, אז הוא אינו דוגמא נגדית.

יהיו $S_1, \dots, S_n$ הסימנים המופיעים ב- Γ וב- γ . נרצה להגדיר הרחבה של הסימנים הללו לסדר מלא.

הגדרה 3.23 הרחבה של Γ ליחס סדר מלא על סימני Γ ו- γ היא פרמוטציה:

$$S_{i_1} \theta_1 S_{i_2} \theta_2 \dots \theta_{n-1} S_{i_n}$$

כאשר כל θ_i הוא $<$ או $=$, המקיימת את הדברים הבאים:

1. אם S_{i_j} ו- S_{i_k} שני קבועים שונים מהתחום (כלומר $k \neq j$) כך ש- $S_{i_j} < S_{i_k}$, אז $j < k$, ולפחות אחד מה- θ_{i_l} ביניהם ($j \leq l < k$) הוא $<$.

2. אם $S_{i_j} < S_{i_k}$ היא השוואה מ- Γ , אז $j < k$ ולפחות אחד θ_{i_l} ביניהם הוא $<$.

3. אם $S_{i_j} = S_{i_k}$ ב- Γ , אז $j = k$, או שכל ה- θ_{i_l} ביניהם הם $=$.

4. אם $S_{i_j} \leq S_{i_k}$ ב- Γ אז $j \leq k$.

5. אם $S_{i_j} \neq S_{i_k}$ ב- Γ , אז $j \neq k$ ולפחות θ_{i_l} אחד ביניהם הוא $<$.

אם נסתכל על הדוגמה הקודמת - מיפוי שהגדרנו לא הגדיר יחס סדר מלא - למשל, לא הוגדר הקשר בין Y ל- W , ויש להגדיר זאת כדי שסדר היחס יהיה מלא.

הערה 3.24 סעיפים 5 - 2 מגדירים מתי השוואה כלשהיא $S_{i_j} < S_{i_k}$ מתקיימת בהרחבה, גם אם היא לא ב- Γ .

טענה 3.25 יש דוגמה נגדית לגרירה $\gamma \models \Gamma$ אם ורק אם יש הרחבה של Γ ליחס סדר מלא על סימני Γ ו- γ כך ש- γ איננו מתקיים.

תרגיל 1: להוכיח הטענה מעלה.

תרגיל 2: למצוא אלגוריתם יעיל למקרה שבו כל ההשוואות ב- Γ וב- γ הן $<$, ולהוכיח את נכונותו. אם מצליחים יותר - זה מבורך!

רמז: דרגת הקושי של המקרים:

המקרה הכי פשוט - כל הסימנים הם $<$

פחות פשוט - ב- γ יש $<, \leq$ וב- Γ יש רק $=$.

מקרה עוד פחות פשוט - ב- Γ יש $<, \leq, =$ וב- γ יש $<, \leq, =$.

והמקרה הכי קשה: בשני הצדדים יכול להיות $<, \leq, \neq$.

פתרון תרגילי בית מההרצאה 7

תרגיל 1: ביצוע מינימיזציה של חוק הורן רוצים לצמצם כמה שיותר אטומים בשאלתא (על מנת לצמצם את מספר פעולות ה- $Join$), כיצד עושים זאת?

הרעיון הבסיסי ביותר - נמחק אטום מהגוף, ונבדוק אם מה שמתקבל שקול לחוק המקורי. אם לא, נחזיר אותו, וננסה למחוק אטום אחר. אם כן, נמחק אטום זה ונחזור על התהליך (באופן רקורסיבי) עם מה שנשאר. ברור שמסיימים כשאין בגוף אף אטום שאפשר למחוק.

תרגיל 3: האם צריך לבדוק האפשרות למחוק אטום בגוף יותר מפעם אחת? משמעות השאלה היא, האם מחיקת אטום יכולה ליצור אפשרויות חדשות למחיקה שלא היו קודם.

תרגיל 4: האם יתכן שמספר האטומים שנשארים בסוף תלוי בסדר המחיקה?

תרגיל 5: נניח שמספר האטומים בסוף אינו תלוי בסדר המחיקה. לכן, השאלה המעניינת היא, האם התוצאה הסופית היא יחידה. מהי ההגדרה לכך שהתוצאה הסופית היא יחידה?

שאלה 2: מהי המשמעות של חוק עם גוף ריק? H :

גימון ריק הוא טאטולוגיה - $TRUE$. תנאי הבטיחות דורש במקרה הזה שהראש יהיה אטום קרקעי, ולכן החוק הזה מייצר את H , מבלי קשר למה מכיל D , גם אם D ריק.

דוגמא: $R(1, 7, 2)$ מייצר את $R(1, 7, 2)$.

שאלה 3: איך מתרגמים חוק הורן לאלגברת היחסים?

למשל נביט בחוק הבא:

$$R(X, Y) : -T(X, Z), T(Z, W), T(Z, W)$$

איך נתרגם אותו? נעשה מכפלה קרטזית של היחסים בגוף. בדוגמא הנ"ל:

$$T \times T \times T$$

נמספר את העמודות של המכפלה הקרטזית (לסימון נוח) משמאל לימין: $A_1, \dots, A_{n'}$. בדוגמא שלנו, $n' = 6$. נפעיל את המכפלה הקרטזית על כל ההשוואות הבאות:

1. אם עמודה A_i מתאימה לקבוע C , נבצע $\sigma_{A_i=C}$.

2. אם עמודות A_i, A_j כך ש- $j \neq i$ מתאימות לאותו משתנה, נבצע $\sigma_{A_i=A_j}$.

נסמן ע"י B' את התוצאה עד כה. יהיו $C_1, \dots, C_j$ הקבועים שמופיעים רק בראש החוק. $\{(C_1, \dots, C_j)\}$ הינו יחס שמכיל רשומה יחידה המורכבת מקבועים אלו. נבצע מכפלה קרטזית $B' \times \{(C_1, \dots, C_j)\}$. נסמן את התוצאה של פעולה זו ב- B , ואת עמודותיה ב- $A_1, \dots, A_n$ (משמאל לימין).

הטלה מוכללת מוגדרת על סדרה של עמודות (ולא על הקבוצה). היא מבצעת שינוי סדר העמודות וגם שכפול של עמודות. יהי $R(S_1, \dots, S_m)$ ראש החוק. לכל סימן S_j נבחר עמודה A_{i_j} מ- B שמתאימה ל- S_j . נבצע הטלה מוכללת של B על $\prod_{A_{i_1}, \dots, A_{i_m}}$.

הערה 3.26 הסימנים $S_1, \dots, S_m$ בראש החוק לא בהכרח שונים זה מזה. אפשר אולי לבחור את A_{i_j} ביותר מדרך אחת, אבל זה לא משנה.

בדוגמא שלנו, עבור X בוחרים את A_1 , ועבור Y את A_4 .

שאלה 6: דיברנו על מינימיזציה של אטום בודד. באופן דומה אפשר לדבר על מינימיזציה של איחוד של חוקים. מה זה אומר?

שאלה 7: בזמן מינימיזציה של חוק בודד, מוחקים אותו ובודקים האם שקילות לחוק לפני המחיקה. למשל, אם B' התוצאה של מחיקת אטום מ- B , אז יש לבדוק הכלה משני הכיוונים בין $B' : -B$ ל- H . איך עושים את כל אחת משתי בדיקות ההכלה האלה בצורה המהירה ביותר?

27/5/2013
הרצאה 9

ישנן שתי דרכים להראות כי יש דוגמא נגדית:

- ניצור הרחבה שמקיימת את Γ ולא את γ .
- נראה מיפוי מהתחום המהווה דוגמא נגדית.

מדוע שתי דרכים?

כי יש אינסוף מיפויים אפשריים מהתחום, ולכן הרבה יותר מעשי להשתמש בהרחבות לשם מציאת דוגמא נגדית (יש בסך הכל $n!2^{n-1}$ הרחבות). זו עדיין כמות גדולה מידי של אפשרויות, ולכן בתרגיל רצינו פתרון פולינומיאלי.

פתירת בעיית הגרירה - שאלה 2 מהרצאה 8

תחילה נטפל במקרה בו כל ההשוואות הן $<$.

נגדיר גרף שבו יש צומת לכל סימן שנמצא ב- Γ או ב- γ , וצלע מכוונת מ- S_i ל- S_j אם $S_i < S_j$ שייך ל- Γ . כעת, נניח כי γ הוא $S_i < S_j$. אזי, $\Gamma \models \gamma$ אם ורק אם בגרף יש מסלול מ- S_i ל- S_j . דוגמא - שקופית 4. כדי שהאלגוריתם יהיה נכון, צריך להוסיף:

- אם צד שמאל אינו ספיק, אז הגרירה מתקיימת בכל מקרה. צד שמאל איננו ספיק אם ורק אם בגרף יש מעגל.
- הוספת צלע מכוונת בכיוון המתאים בין כל שני קבועים בגרף.

למשל, בשקופית 4 כאשר מוסיפים את הצלע האדומה.

שאלה כאשר פותרים את בעיית הגרירה, האם ניתן לפתור את הבעיה עבור כל תחום במקביל?

תשובה אם נתונים קבועים, חייבים להתייחס לתחום ספציפי לשם בחירת יחס הסדר בין הקבועים. אם אין קבועים, ניתן לפתור באופן כללי.

הערה 3.27 במקרים מסויימים ייתכן שיהיה ניתן לקבוע שיש גרירה אפילו מבלי לדעת את היחס בין הקבועים. דוגמא - שקופית 8.

תרגיל 1: הוכיחו את נכונות האלגוריתם לאחר התיקונים.

הגדרה 3.28 נתון גרף מכוון. רכיב קשירות חזק (רק"ח) היא קבוצה מקסימלית של קודקודים בגרף, כך שכל שניים נמצאים על מעגל מכוון. במילים אחרות, בין כל שני קודקודים x ו- y באותו רק"ח, יש מסלול מכוון מ- x ל- y , ומסלול מכוון מ- y ל- x .

שאלה כיצד מייצרים אלגוריתם למקרה הכללי (השוואות $<, \leq, =, \neq$)?

תשובה מייצרים גרף רכיבי קשירות - אם בכל רק"ח נמזג את כל הקודקודים לצומת אחת ונשאיר רק צלעות בין הקודקודים שיוותרו (כלומר, בין רק"חים שונים), נקבל גרף חסר מעגלים.

הערה 3.29 יש אלגוריתם לינארי המוצא את הרק"חים של גרף מכוון.

הגדרה 3.30 אם γ היא השוואה, $\neg\gamma$ היא ההשוואה המשלימה.

דוגמאות

• אם γ היא $X \neq Y$, אז $\neg\gamma$ היא $X = Y$.

• אם γ היא $X < Y$, אז $\neg\gamma$ היא $X \geq Y$.

טענה 3.31 $\Gamma \models \gamma$ אם ורק אם $\Gamma \wedge \neg\gamma$ בלתי ספיקה.

תרגיל 2: הוכיחו את הטענה הנ"ל.

אלגוריתם פולינומיאלי לבדיקה האם $\Gamma \wedge \neg\gamma$ בלתי ספיק נבנה גרף מכוון. בגרף יהיה צומת לכל סימן ב- Γ או ב- γ . כמו כן, יהיו צלעות לפי ההשוואות ב- $\Gamma \wedge \neg\gamma$ ² כדלקמן³:

• אם יש השוואה $S_i < S_j$ או אם $S_i \leq S_j$ יש צלע אחת $S_i \rightarrow S_j$.

• אם $S_i = S_j$ יש שתי צלעות, צלע אחת בכל כיוון.

• בין כל שני קבועים המופיעים ב- Γ או ב- γ יש צלע בכיוון המתאים לפי היחס המוגדר על התחום

• אם $S_i \neq S_j$, אין צלע.

הערה 3.32 לא שמים בגרף צלעות עבור המקרה של $\neq$, כי צלעות מציינות נתונים, כלומר דברים שמתקיימים בהכרח. במקרה של $\neq$ היחס המדויק לא ידוע.

תחילה, נמצא את הרק"חים.

$\Gamma \wedge \neg\gamma$ הוא בלתי ספיק אם ורק אם יש קבועים שונים, או עם שני סימנים לאו דווקא שונים⁴ S_i ו- S_j כך שב- $\Gamma \wedge \neg\gamma$ מופיעה ההשוואה $S_i < S_j$ או ההשוואה $S_i \neq S_j$.

הערה 3.33 עבור $\leq, \geq$ אין בעיה בין אם הקודקודים באותו רק"ח או לא, ולכן אין צורך לבדוק את האפשרויות השונות. הרק"חים מכילים בהכרח את מה שחייב להיות שווה.

דוגמה בשקופית 17.

כיצד התגברנו על כמות המקרים האקספוננציאלית?

• אם מנסים את כל האפשרויות עבור $\neq$, מקבלים כמות אקספוננציאלית.

• אם יש מעגל בגרף אז יש סתירה; בניית הרק"חים הופכת את חיפוש המעגלים (מהם יש כמות אקספוננציאלית) לתהליך קל חישובית.

תרגיל 3: הוכיחו את נכונות האלגוריתם, כלומר $\Gamma \wedge \neg\gamma$ בלתי ספיק אם ורק אם קובע האלגוריתם. כיוון $\Rightarrow$ ברור, כיוון $\Leftarrow$ לבית.

²כלומר, לא רק לפי אלה ב- Γ

³שימו לב שכאן במצגת הוגדרו משקלות לכל צלע. מכיוון ומאוחר יותר יש התעלמות ממשקלות אלה, פשוט מחקתי אותן מההגדרה.

⁴כלומר, יתכן ש- $i = j$

פתרון תרגילים מהרצאה 8

תרגיל 7:

נתון חוק $H : -B$, ו- B' מתקבל מ- B על ידי מחיקת אטום אחד. צריך לבדוק האם $H : -B'$ שקול לחוק המקורי $H : -B$. כלומר, לבדוק הכלה בשני הכיוונים, כלומר צריכים לבדוק באם יש הומומורפיזמים בשני הכיוונים. כיוון אחד אין צורך לבדוק, כי ברור ש- B מוכל ב- B' , כי הגוף הוא שאילתות גימס: הומומורפיזם שמראה את ההכלה הוא מיפוי הזהות. לכן מספיק לבדוק הכלה בכיוון השני. בדיקת קיום הומומורפיזם היא אקספוננציאלית בגודל החוקים. למרות זאת, כדאי להשתמש בבדיקת הכלה לצורך אופטימיזציה, כי השאילתות מחושבות על מסד נתונים גדול. לעומת זאת, בדיקת הכלה מתבצעת על קלט קטן, שהוא רק גודל השאילתות.

שאלה 3:

"האם צריך לבדוק האפשרות למחוק אטום בגוף יותר מפעם אחת? משמעות השאלה היא, האם מחיקת אטום יכולה ליצור אפשרויות חדשות למחיקה שלא היו קודם." לא! נוכיח זאת: נסמן את החוק המקורי ב- $H : -B$. נתאר את המצב שלגבי צריך להוכיח שהוא בלתי אפשרי: מוחקים אטומים, ומקבלים $H : -B_1$ ששקול לחוק המקורי, ועכשיו מנסים למחוק אטום b מהגוף של B_1 :

$$H : -B_1 \setminus \{b\}$$

אבל אין שקילות. אולם, יש B_2 שהוא תת קבוצה של B_1 כך ש: $H : -B_2$ שקול לחוק המקורי, ו- B_2 מכיל את b . עכשיו אפשר למחוק את b , כלומר:

$$H : -B_2 \setminus \{b\}$$

שקול לחוק המקורי. אבל, במקרה זה נקבל שגם $H : -B_1 \setminus \{b\}$ שקול למקורי מטרנזיטיביות, ולכן אפשר למחוק את b מ- B_1 , בניגוד להנחתנו. הראנו שמספיק לנסות למחוק אטום מהגוף פעם אחת בלבד. במילים אחרות, אם לא ניתן למחוק אטום מהגוף תוך שמירה על שקילות, אז אי אפשר יהיה למחוק אותו בשלב מאוחר יותר (כלומר, אחרי שכמה אטומים כבר נמחקו).

שאלה 4 - האם יתכן שמספר האטומים שנשארים בסוף תלוי בסדר המחיקה?

לא! למעשה, צריך להוכיח שאין שני חוקים:

$$H_1 : -B_1, H_2 : -B_2$$

כך שהם שקולים, יש להם מספר שונה של אטומים בגוף ולאף אחד אין אטום מיותר בגוף (כלומר, כזה שאפשר למחוק תוך שמירה על שקילות). נסמן את מספר האטומים ב- B_1 ב- n_1 , ואת מספר האטומים ב- B_2 ב- n_2 . מכיוון והם שקולים, יש הומומורפיזמים בשני הכיוונים:

$$h_1 : (H_1 : -B_1) \rightarrow (H_2 : -B_2), h_2 : (H_2 : -B_2) \rightarrow (H_1 : -B_1)$$

הרכבה של הומומורפיזמים היא גם הומומורפיזם. נניח (בה"כ) שב- B_2 יש יותר אטומים מ- B_1 , כלומר $n_2 > n_1$. ההרכבה $h_1 \circ h_2$ היא הומומורפיזם מהחוק השני לעצמו. התמונה $h_1 \circ h_2 (B_2)$ היא תת קבוצה של B_2 , כי יש בה לכל היותר רק n_1 אטומים (כי זה מספר האטומים לכל היותר בתמונה של h_1). אם כך, $h_1 \circ h_2$ מראה שקילות בין $H_2 : -B_2$ ל- $H_2 : -h_1 \circ h_2 (B_2)$, אבל ב- $h_1 \circ h_2 (B_2)$ יש פחות אטומים מאשר ב- B_2 , סתירה.

הערה 3.34 $h_1 \circ h_2$ מראה הכלה רק בכיוון אחד (איזה?), אבל הכלה בכיוון השני תמיד מתקיימת (תרגיל 7).

שאלה 5 - נניח שמספר האטומים בסוף אינו תלוי בסדר המחיקה. מהי ההגדרה לכך שהתוצאה הסופית היא יחידה?

$$S(X, Y) : -R(X, Y)$$

$$S(W, V) : -R(W, V)$$

שני החוקים הנ"ל הם אותו דבר (עד כדי שינוי שמות המשתנים). באופן פורמלי, חוקים הם "אותו דבר" אם הם איזומורפיים:

הגדרה 3.35 שני חוקים הינם איזומורפיזם אם קיים הומומורפיזם חח"ע ועל מאחד לשני, כך שהמיפוי ההפוך הוא הומומורפיזם בכיוון השני.

אנו מניחים שגוף הוא קבוצת אטומים, כלומר אטום לא מופיע יותר מפעם אחת בגוף.

הערה 3.36 אם הומומורפיזם הוא חח"ע ועל, אין זה בהכרח אומר שהכיוון ההפוך (שמוגדר היטב) הוא גם הומומורפיזם. למשל:

$$\begin{aligned} R(X, Y) &: -T(X, Y) \\ R(X, Y) &: -T(X, Y), T(Y, X) \end{aligned}$$

פונקציית הזהות (על סימנים) היא הומומורפיזם מהחוק העליון לתחתון, אבל לא להיפך:

$$h(X) = X, h(Y) = Y$$

שאלה 4: הוכיחו שאם שני חוקים הם שקולים וללא אטומים מיותרים בגוף, אז הם בהכרח איזומורפיזם.

03/06/2013

הרצאה 10

תזכורת: איך בודקים אם גימון של השוואות הוא בלתי ספיק?
בונים גרף, כאשר הצמתים של הגרף הם הסימנים - כל סימן מופיע בגרף פעם אחת כצומת (גם אם הוא מופיע מספר פעמים בהשוואות, הוא יופיע פעם אחת בלבד בגרף), והצלעות מתאימות להשוואות בהתאם לסוג ההשוואה:

- אם ההשוואה מסוג $\neq$, אזי אין צלעות.
- אם ההשוואה מסוג $=$, אזי יש שתי צלעות בשני הכיוונים.
- לשאר הסימנים, תהיה צלע מהקטן לגדול.

באופן כללי, מה הזמן לבניית הגרף? לינארי בגודל האינפוט.
לאחר בניית הגרף, מוצאים רכיבי קשירות (האלגוריתם הכי טוב ימצא אותם בזמן לינארי בגודל הגרף - מספר הצלעות והצמתים).

מהו רכיב קשירות? תחילה, נזכור כי בגרף מכוון מדובר ברכיבי קשירות חזקים - רק"ח. זוהי קבוצה מקסימלית של קודקודים, כאשר עבור כל זוג קודקודים יש מסלול מאחד לשני. בעצם רכיבי קשירות מבטאים מחלקות קשירות, כאשר החלוקה מבוססת על היחס שהוגדר ⁵.

לבסוף, מבצעים בדיקה (שקופית 18 בהרצאה הקודמת) על כל אחד מהרק"חים: רק"ח איננו ספיק אם יש שני קבועים שונים באותו רכיב קשירות, או שיש S_i, S_j שהם לאו דווקא שונים, אבל באופן מפורש בהשוואות שניתנו $\neg \gamma \wedge \Gamma$ נתון כי $S_i < S_j$ או $S_i \neq S_j$.

3.2.3 הכלה ושקילות בין שאלות עם השוואות

הגדרה של הומומורפיזם כשיש השוואות בגוף נביט בשני החוקים הבאים:

$$\begin{aligned} Q_1 R(X, Y) &: -T(X, Y), X \leq Y \\ Q_2 R(X, Y) &: -T(X, Y), X < Y \end{aligned}$$

Q_1 מכיל את Q_2 .

השאלה היא, איך לקחת ובצורה הפשוטה ביותר להכליל הומומורפיזם למקרים בהם יש השוואות? באופן אינטואיטיבי, אפשר לבנות מיפוי סימנים (כמו שעשינו במקרה הקודם), אבל במקרה שכזה לא נקבל הומומורפיזם, למשל בדוגמה זו. כבר אנו רואים כי הומומורפיזם בצורה המחמירה ביותר הוא תנאי מספיק, אך לא הכרחי. איך בכל זאת אפשר לדעת "שזה בסדר"? ההשוואה התחתונה $X < Y$ היא יותר חזקה מ: $X \leq Y$ כי היא גוררת אותה. זוהי דרך "יותר הגיונית" להכליל הומומורפיזם.

⁵עוד על זה בקורס מבנים אלגבריים 1+2.

הגדרה 3.37 נסמן:

$$Q_1 H_1 : -B_1, \Gamma_1$$

$$Q_2 H_2 : -B_2, \Gamma_2$$

כאשר B_i היא קב' אטומים רגילים, Γ_i קב' השוואות, הקבוצות הן במשמעות של גימס, ואין בהן חזרות. מיפוי h מהסימנים של Q_1 אל הסימנים של Q_2 הוא הומומורפיזם, אם:

1. h מיפוי סימנים (כלומר, מעביר קבוע לעצמו).

$$2. h(H_1) = H_2$$

$$3. h(B_1) \subseteq B_2$$

$$4. \Gamma_2 \models h(\Gamma_1)$$

הערה 3.38 תנאים 1 – 3 הם ההגדרה של הומומורפיזם כשאין השוואות.

הערה 3.39 תנאי 4 אומר ש- $\Gamma_2 \models \gamma$ לכל $\gamma \in h(\Gamma_1)$, וברור ש- $h(\Gamma_1)$ מתקבל מהשוואות של Γ_1 ע"י הפעלת h על הסימנים של Γ_1 .

לעומת הפעם הקודמת, יש דבר שדורש שינוי - הגדרת התוצאה. ברגע שנגדיר זאת, ההגדרה של הכלה תשאר כפי שהייתה.

הגדרה 3.40 בהנתן מסד $\mathcal{D}$ המוגדר מעל תחום עם יחס סדר $<_{\mathcal{D}}$, התוצאה של שאלתא:

$$QH : -B, \Gamma$$

שתסומן ע"י $Q(\mathcal{D})$ היא הקבוצה:

$$\{g(H) \mid g \text{ is a ground mapping s.t. } \forall b \in B \ g(b) \in \mathcal{D}, \forall \gamma \in \Gamma \ g(\gamma) \text{ is maintained in } <_{\mathcal{D}}\}$$

הערה 3.41 ההשוואה $g(\gamma)$ יכולה להיות $=, \neq, \leq, <$.

הגדרה של הכלה היא כמו במקרה שאין השוואות, כלומר $Q_1 \supseteq Q_2$ אם לכל מסד $\mathcal{D}$ מתקיים $Q_1(\mathcal{D}) \supseteq Q_2(\mathcal{D})$.

האם זה תנאי מספיק להכלה? כן! הוכחה: נוכיח שקיום הומומורפיזם הוא תנאי מספיק להכלה $Q_1 \supseteq Q_2$. ההוכחה בצורה - שקופית 7: נסמן את ההומומורפיזם מ- Q_1 ל- Q_2 ב- h . יהי $\mathcal{D}$ מסד כלשהוא, ונניח שהמיפוי הקרקעי g מייצר את האטום $g(H_2)$ לתוצאה מ- Q_2 ומ- $\mathcal{D}$. נשים לב כי g מעבירה את B_2 ל- $\mathcal{D}$, ואת Γ_2 להשוואה שמתקיימת ב- $<_{\mathcal{D}}$.

ההרכבה $g \circ h$ מעבירה את הראש של Q_1 לאטום ש- Q_2 מייצר לתוצאה. כמו כן, היא מעבירה כל אטום בגוף של Q_1 לאטום ב- $\mathcal{D}$.

לגבי ההשוואות, נתון $\Gamma_2 \models h(\Gamma_1)$, ומכיוון g מיפוי סימנים, דהיינו מעביר כל קבוע לעצמו, גם $g(\Gamma_2) \models g \circ h(\Gamma_1)$. מכיוון ש- $g(\Gamma_2)$ מתקיימת ב- $<_{\mathcal{D}}$ (מההגדרה של g), גם $g \circ h(\Gamma_1)$ מתקיימת ב- $<_{\mathcal{D}}$, כי אחרת זו היתה דוגמא נגדית לגרירה.

לפיכך, הראינו ש- $g \circ h$ מייצר עבור Q_1 את אותו אטום ש- g מייצר עבור Q_2 מייצר מ- $\mathcal{D}$. הראנו את זה עבור g כלשהו, לכן Q_1 מייצר מ- $\mathcal{D}$ כל אטום ש- Q_2 מייצר מ- $\mathcal{D}$.

זה נכון לכל $\mathcal{D}$, ולכן $Q_1 \supseteq Q_2$. ■

דוגמא

$$Q_1 R(X, Y) : -T(X, Z), T(Z, Y), X < Y$$

$$Q_2 R(X, Y) : -T(X, Z), T(Z, Y), X \leq Z, Z < Y$$

האם יש הומומורפיזם, ואם כן, באיזה כיוון?

החוק התחתון דורש שתתקיים השוואה לכל אחד משני האטומים האלה, שזה יותר מגביל מהתנאי העליון. ככל שיש יותר תנאים, אפשר לייצר פחות אטומים לתוצאה. לכן החוק התחתון בוודאי לא מכיל את העליון.

על מנת להוכיח זאת באופן פורמלי בבחינה, יש להראות דוגמא נגדית. בדברים שנגיד היום בהרצאה תתחבא שיטה קלה למציאת דוגמא שכזו.
האם החוק העליון מכיל את התחתון? פשוט ננסה לבנות הומומורפיזם: האטומים עוברים לעצמם (שקופית 11), ויש לבדוק אם:

$$X \leq Z, Z < Y \models X < Y$$

גרירה זו מתקיימת, ולכן יש הומומורפיזם מהחוק העליון לתחתון.

האם זהו תנאי הכרחי?

נסתכל על (שקופית 12):

$$Q_1 P(W) : -T(W), R(X, Y), X \leq Y$$

$$Q_2 P(W) : -T(W), R(X, Y), R(Y, X)$$

האם יש פה הומומורפיזם באיזהו כיוון? אם היה הומומורפיזם מ- Q_2 ל- Q_1 , אז היה צריך להתקיים (תנאי 4):

$$X \leq Y \models h(\phi) = h(true) = true$$

(גימזם ריק הוא $true$), וזה תמיד נכון, אז תנאי זה מתקיים. אולם, אם ננסה לבנות הומומורפיזם בין השניים, אין לאן למפות את $R(Y, X)$, ולכן אין הומומורפיזם כזה.
אם מנסים בכיוון ההפוך, אז היינו מקבלים את התנאי (שכן גימזם ריק הוא $true$):

$$true \models X \leq Y$$

ואין כאן גרירה.

עולה השאלה, האם יש הכלה בכל זאת באיזהו כיוון? Q_2 לא מכיל את Q_1 , שכן קל למצוא דוגמא בה Q_1 מייצר משהו לתוצאה, ו- Q_2 לא, למשל:

$$\mathcal{D} = \{T(1), R(1, 2)\}$$

שכן, המיפוי הקרקעי g הבא:

$$g(X) \rightarrow 1, g(Y) \rightarrow 2, g(W) \rightarrow 1$$

מייצר את האטום $P(1)$ לתוצאה. אבל, מכיוון וב- $\mathcal{D}$ אין שני אינסטנסים של $R(X, Y)$ או אינסטנס בו $X = Y$, התוצאה של Q_2 על $\mathcal{D}$ היא ריקה.
בכיוון השני יש הכלה! ההוכחה לפי מקרים, בשקופית 15 והלאה.

הערה 3.42 אילו ב- Q_1 היה $X < Y$ ולא $X \leq Y$ לא היתה הכלה.

עולה מכאן השאלה, מהו תנאי מספיק והכרחי להכלה?

תנאי הכרחי ומספיק להכלה $Q_1 \supseteq Q_2$

טענה 3.43 נסמן:

$$Q_1 : H_1 : -B_1, \Gamma_1$$

$$Q_2 : H_2 : -B_2, \Gamma_2$$

$Q_1 \supseteq Q_2$ אם"מ לכל הרחבה $\bar{\Gamma}_2$ של Γ_2 ליחס סדר מלא מעל כל הסימנים של Q_2 יש הומומורפיזם מ- Q_1 ל:

$$\bar{Q}_2 : H_2 : -B_2, \bar{\Gamma}_2$$

דוגמא להרחבות אפשריות:

$$\begin{aligned} Q_2 : P(W) : -T(W), R(X, Y), R(Y, X) \\ W < X < Y \\ W = X < Y \\ W < X = Y \\ W = X = Y \\ \vdots \end{aligned}$$

כאשר לכל הרחבה אפשרית צריך להראות הומומורפיזם.

בשקופית 19 בודקים שתיים מההרחבות האפשריות.

בשקופית 20 - יש הכלה? Q_2 תמיד מייצר כלום, קבוצה ריקה, ולכן מוכלת בכל דבר, ועל כן $Q_1 \supseteq Q_2$. נשים לב כי כאן התנאי מתקיים באופן ריק, כי אין הרחבה של $X < X$ (שהרי לפי ההגדרה, הרחבה צריכה לקיים את כל ההשוואות, ולא ניתן לקיים את $X < X$). **הוכחה:** נראה שהתנאי הכרחי. נניח ש- $Q_1 \supseteq Q_2$, צריך להראות שלכל הרחבה אפשרית $\bar{\Gamma}_2$ של Γ_2 ליחס סדר מלא על כל סימני Q_2 , יש הומומורפיזם מ- Q_1 ל-:

$$\bar{Q}_2 H_2 : -B_2, \bar{\Gamma}_2$$

בהנתן הרחבה $\bar{\Gamma}_2$ נייצר מסד קנוני מ- B_2 ע"י העברה של כל משתנה V לקבוע ייחודי V_0 . ההרחבה $\bar{\Gamma}_2$ עשויה לקבוע שיש קבועים ששווים זה לזה. עובדה זו לא מהווה בעיה, אבל צריך לזכור זאת. יהי g המיפוי שמעביר כל משתנה לקבוע ייחודי, ו- $\bar{\Gamma}_2$ הוא יחס הסדר המוגדר על המסד הקנוני. בשרטוט בשקופית 24 יש לשנות - Q_2 הוא למעשה $\bar{Q}_2$, ולחוק Q_1 יש את Γ_1 . מההגדרה של הרחבה ואופן בניית המסד הקנוני, ברור ש- $g(\bar{\Gamma}_2)$ מתקיים במסד הקנוני.

כמו כן, אם $g(\bar{\Gamma}_2)$ מתקיים, אז בוודאי גם $g(\Gamma_2)$ מתקיים, שכן $\bar{\Gamma}_2$ היא הרחבה של Γ_2 . כלומר, g מראה ש- Q_2 מייצר את הראש הקנוני של עצמו מהמסד הקנוני. בגלל שיש הכלה, גם Q_1 צריך לייצר אותו דבר, ולכן קיים מיפוי הסימנים f , שחייב לקיים את כל ההשוואות $f(\Gamma_1)$ לפי יחס הסדר $g(\bar{\Gamma}_2)$ המוגדר על המסד הקנוני.

לכן, $f \circ g^{-1}$ הוא הומומורפיזם מ- Q_1 ל- $\bar{Q}_2$. ההסקנה האחרונה - הגרירה $g(\bar{\Gamma}_2) \models f(\Gamma_1)$ טריוויאלית אם $\bar{\Gamma}_2$ כוללת גם את כל הקבועים שמופיעים ב- Γ_1 . אחרת, לא טריוויאלי להוכיח כי הגרירה מתקיימת, אבל זה אפשרי. ■

בשל הבעיות שהופיעו בהוכחה, ולמען הפשוטות והאחידות, נשתמש בתנאי הכרחי ומספיק הבא:
 $Q_1 \supseteq Q_2$ אמ"מ קיים הומומורפיזם לכל הרחבה $\bar{\Gamma}_2$ של Γ_2 ליחס סדר מלא על כל סימני Q_2 והקבועים של Q_1 (כולל אלה שלא מופיעים ב- Q_2). במקרה זה, הגרירה $g(\bar{\Gamma}_2) \models f(\Gamma_1)$ נובעת באופן טריוויאלי מכך שההשוואות $f(\Gamma_1)$ מתקיימות לפי יחס הסדר $g(\bar{\Gamma}_2)$. מכיון ש- g חח"ע ועל, נובע ש- $\bar{\Gamma}_2 \models g^{-1} \circ f(\Gamma_1)$.

תרגיל 1: להוכיח שהתנאי גם מספיק.

ראינו בהרצאה קודמת את ההגדרה של איחוד של שאילות U , וההגדרה של הכלה כשאין השוואות. אי אפשר להגיד שהתנאי הזה נשאר כמו שהוא עבור הכלה במקרים בהם יש השוואות: כאן, בניגוד לפעם הקודמת, לא חייבים הכלה ממש באחד האיברים, כאן ניתן לייצר את התוצאה גם עם פיצול התנאי.

תרגיל 2: מהו תנאי הכרחי ומספיק ל- $U_1 \subseteq U_2$, כאשר בחוקים יש השוואות. הוכיחו התשובה.

תשובה לשאלה 2 מההרצאה הקודמת:

יהיו:

$$U = \{Q_1, \dots, Q_n\}, \quad U' = \{Q'_1, \dots, Q'_m\}$$

ונסמן:

$$Q_i : H_i : -B_i, \Gamma_i, \quad Q'_j : H'_j : -B'_j, \Gamma'_j$$

תיקון (שקופית 4):

תנאי הכרחי ומספיק להכלה בין איחודים $U \supseteq U'$

לכל $Q'_j \in U'$ ולכל הרחבה של Γ'_j ליחס סדר מלא $\bar{\Gamma}'_j$ על כל סימני Q'_j , קיימת $Q_i \in U$ שעבורה יש הומומורפיזם מ- Q_i לשאילתא:

$$\bar{Q}'_j : H'_j : -B'_j, \bar{\Gamma}'_j$$

הערה 3.44 כאשר מדובר בהכלות של איחודים, כאשר בודקים הכלה $Q_1 \supseteq Q_2$, ההרחבה $\bar{\Gamma}_2$ ליחס סדר מלא צריכה להיות מעל כל הסימנים של Q_2 והקבועים של Q_1 (כולל אלו שלא מופיעים ב- Q_2). לדוגמא:

$$U = \{T(X) : -T(X), X \leq 1; T(X) : -T(X), X > 1\}$$

$$U' = \{T(X) - T(X)\}$$

ההרחבות של החוק היחיד ב- U' חייבות לכלול את הקבוע 1 כדי להראות $U \supseteq U'$.

נוכיח כי התנאי הוא אכן הכרחי: **הוכחה:** תהי $\bar{\Gamma}'_j$ הרחבה של Γ'_j ליחס סדר מלא על כל סימני Q'_j וכל הקבועים ב- U (כולל אלה שאינם ב- Q'_j).

נייצר מסד קנוני מהאטומים הרגילים בגוף של Q'_j עם יחס הסדר המלא המוגדר ע"י $\bar{\Gamma}'_j$. Q'_j מייצר מהמסד הקנוני את הראש הקנוני של עצמו. מכיוון ש- $U \supseteq U'$, יש חוק $Q_i \in U$ המייצר את אותו ראש קנוני מהמסד הקנוני (אחרת זו דוגמא נגדית להכלה). שאר ההוכחה היא כמו בשקפים 27 – 23 בהרצאה 10. ■

תרגיל 1: לשנות את ההוכחה בשקפים 10 – 8 משבוע שעבר להוכחה שהתנאי המספיק והכרחי להכלה הוא אומנם מספיק. לעשות זאת לשני מקרים:
1. חוקים בודדים עם השוואות.
2. איחוד של חוקים עם השוואות.

3.3 חוקים עם שלילה

תחילה, אינטואיציה:

$$R(X, Y) : -T(X, Z), T(Z, Y), -T(Y, Z), \mathcal{D} = \{T(1, 2), T(2, 3)\}$$

דוגמא - שקופית 12: מיפוי סימנים קרקעי g מהחוק מעלה למסד, מתקבל $R(1, 3)$. כלומר, דורשים שהאטומים החיוביים במסד, והשליליים לא במסד (במיפוי זה, הדרישה היא $T(3, 2)$ לא במסד, והיא מתקיימת). אם היינו כותבים במקום האטום בשלילה את האטום $-T(W, V)$, ישנם משתנים שלא מופיעים כלל באף אטום חיובי, אם כך זה תמיד יוכל להתקיים. על כן, אנו דורשים בטיחות:

תנאי בטיחות למקרה הכללי ביותר (כולל השוואות ושלילה):

כל משתנה שמופיע בחוק חייב להופיע באטום חיובי בגוף.

הערה 3.45 הכוונה לאטום רגיל ולא להשוואה.

דוגמא לחוק אסור נוסף - שקופית 13, שכן הוא לא מקיים את תנאי הבטיחות.

יהי:

$$Q : H : -B, N$$

כאשר B קב' של אטומים רגילים חיוביים ו- N קבוצה של אטומים רגילים שליליים (כלומר, כל אחד מופיע עם שלילה לפניו).

הגדרה 3.46 בהנתן מסד $\mathcal{D}$, התוצאה של:

$$Q \ H : -B, N$$

שתסומן $Q(\mathcal{D})$, היא:

$$\{g(H) \mid g \text{ is a ground mapping s.t. } \forall b \in B \ g(b) \in \mathcal{D} \text{ and } \forall n \in N \ g(n) \notin \mathcal{D}\}$$

3.3.1 תנאי מספיק להכלה

נסמן:

$$Q_1 : H_1 : -B_1, N_1 \quad Q_2 : H_2 : -B_2, N_2$$

תנאי מספיק להכלה $Q_1 \supseteq Q_2$:

קיום הומומורפיזם $h : Q_1 \rightarrow Q_2$.

הגדרה 3.47 $h : Q_1 \rightarrow Q_2$ הוא הומומורפיזם אם הוא מקיים:

1. h הוא מיפוי סימנים מ- Q_1 ל- Q_2 .

2. $h(H_1) = H_2$.

3. $h(B_1) \subseteq B_2$.

4. $h(N_1) \subseteq N_2$.

הערה 3.48 אם N_1 ריק, אזי תנאי (4) מתקיים באופן ריק.

מקרה מיוחד אין פרדיקט (שם של יחס) שמופיע בגוף של חוק גם באופן חיובי וגם באופן שלילי.

טענה 3.49 במקרה זה, תנאי (4) הוא הכרחי.

הוכחה: נתחיל בצעד הראשון במקרה הרגיל- נייצר מסד קנוני (שקופית 19) מהאטומים ב- B_2 ע"י פונקציה g . אזי $g(N_2)$ יהיה מחוץ למסד, ו- $g(H_2)$ יהיה בתוצאה, ולכן החוק Q_2 מייצר מהמסד הקנוני את הראש הקנוני של עצמו. זאת בשל המקרה המיוחד: אף פרדיקט לא מופיע גם באטום חיובי וגם באטום שלילי. אזי מההכלה, קיימת f הממפה את B_1 למסד הקנוני, ויוצרת מ- H_1 את אותו אטום לתוצאה, ואת N_1 "החוצה", אבל זה לא אומר שהתנאי של ה- h מתקיים, כלומר לא ניתן להסיק דבר בנוגע לקשר בין N_1 ל- N_2 ! שלושת התנאים הראשונים להיות $h = g^{-1} \circ f$ הומומורפיזם מתקיימים (כפי שראינו בהוכחות מסוג זה), אך אנו צריכים לדון בתנאי הרביעי. היינו רוצים להסיק ש: $f(N_1) \subseteq g(N_2)$, שכן אז נוכל להפעיל את g^{-1} , ולקבל:

$$h(N_1) = g^{-1} \circ f(N_1) \subseteq g(N_2)$$

נניח $f(N_1) \not\subseteq g(N_2)$. נבחר באקראי אטום $n_1 \in N_1$ כך ש: $f(n_1) \notin g(N_2)$. נרצה להראות כי במקרה זה לא קיימת הכלה, כלומר צריך למצוא דוגמא נגדית. איך נעשה זאת? נתחיל בניטרול f - נוסיף למסד הקנוני את $f(n_1)$. מה יקרה? המיפוי הזה מייצר עדיין תוצאה עבור Q_2 , אבל לא מייצר תוצאה עבור Q_1 . תהליך זה נקרא חסימת f . כל מיפוי שלא מקיים את $f(N_1) \subseteq g(N_2)$ ניתן לחסום באותו האופן - לא תהיה תוצאה מ- Q_1 , בעוד ש- Q_2 עדיין מייצר אותה.

אם לא נותר מיפוי שלא הצלחנו לחסום, מצאנו דוגמא נגדית, בסתירה להכלה! לכן, קיים מיפוי f כך ש: $f(N_1) \subseteq g(N_2)$, ואז התנאי הרביעי מתקיים.

ובזאת הראנו קיום הומומורפיזם. ■

הערה 3.50 נשים לב כי לאחר שהוספנו את התוצאה $f(n_1)$ למסד הקנוני בנסיון לחסום את הפונקציה הבאה, B_1 עדיין ימופה על ידי כל פונקציה אחרת למסד הקנוני המקורי, שכן האטומים שמוסיפים למסד הקנוני הם של פרדיקטים שאינם מופיעים באופן חיובי ב- B_1 , זאת עקב האפיון המיוחד - פרדיקט לא מופיע בגוף גם באטום שלילי וגם באטום חיובי. ללא התנאי הזה, ההגדרה הזו הופכת לרקורסיה: B_1 יכול היה להתמפות לאחת התוצאות שהוספנו בדרך... שהרי כל הדברים שהוספנו למסד הקנוני שייכים ל- $g(N_1)$. לכן, Q_1 יכול לייצר את הראש הקנוני רק ע"י מיפוי האטומים הרגילים בגוף שלו (קרי, אלו ב- B_1) לחלק המקורי של המסד הקנוני (קרי $g(B_2)$). כלומר, זה מתקיים גם לאחר הוספת אטומים למסד הקנוני עקב חסימות. כמו כן, למרות התוספות למסד הקנוני, Q_2 ממשיך לייצר את הראש הקנוני של עצמו מהמסד הקנוני עם התוספות. זאת מכיוון ואנו מניחים כי האטומים שמוסיפים אינם ב- $g(N_2)$.

הערה 3.51 נשים לב כי השימוש באפיון של התנאי המיוחד הוא רק לגבי Q_1 , למעט השימוש בעובדה ש- Q_2 אינו ריק; אם אותו פרדיקט מופיע גם באופן חיובי וגם באופן שלילי בגוף, אז יתכן שהחוק ריק, כלומר מייצר תוצאה ריקה לכל מסד. לכן, תחילה צריך לבדוק ש- Q_2 אינו ריק. אם הוא ריק, אז יש הכלה. אחרת, בודקים אם יש הכלה כפי שתיארנו, בהנחה שב- Q_1 אין פרדיקט שמופיע גם באופן חיובי וגם באופן שלילי.

תרגיל 2: להוכיח שהתנאי של קיום הומומורפיזם הוא אמנם מספיק.
תרגיל 3: למצוא אלגוריתם לבדיקת הכלה של חוקים עם שלילה במקרה הכללי.
תרגיל 4: להכליל את המקרה המיוחד ל:
 1. השוואות.
 2. איחוד.
 3. שניהם.
תרגיל 5: להכליל את האלגוריתם למקרה הכללי כנ"ל.

פתרון שאלות

שאלה 6 הרצאה 8 - מינימיזציה של איחוד של חוקים

בהנתן איחוד של חוקים:

$$U = \{Q_1, \dots, Q_n\}$$

אם $i \neq j$, אז אפשר למחוק את Q_i . לפיכך, מינימיזציה של חוקים (ללא השוואות וללא שלילה) מתבצעת ע"י כך שכל עוד יש חוק Q_i שמוכל בחוק אחר Q_j (קרי $i \neq j$), מוחקים את Q_i . כשאין עוד יותר חוקים שאפשר למחוק, עושים מינימיזציה של כל חוק בנפרד. איך נעשה מינימיזציה?

• כדי לבדוק שחוק אחד מוכל בחוק אחר - $n(n-1)$.

• לאחר מחיקת חוקים מיותרים - מכל חוק מוחקים את האטומים המיותרים בגוף.

השאלה המעניינת - שקופית 30: האם שני איחודים שקולים ומינימליים יכולים להיות בעלי מספר חוקים שונה זה מזה?
 יהיו:

$$U = \{Q_1, \dots, Q_n\}, U' = \{Q'_1, \dots, Q'_m\}$$

איחודים שקולים, וכן כל אחד מינימלי. האם יתכן ש- $n \neq m$?
 לכל שאילתא ב- U יש שאילתא ב- U' שמכילה אותה. בה"כ $Q_1 \subseteq Q'_1$. האם יתכן ש- $Q'_1 \subseteq Q_i$ כך ש- $i \neq 1$? לא! אחרת, Q_1 חוק מיותר ב- U , בסתירה לכך ש- U מינימאלי.
 לפיכך, לכל שאילתא ב- U יש שאילתא אחת ויחידה כך ש- $Q_i \equiv Q'_i$, ולהפך. בפרט, נובע ש- $n = m$.

הערה 3.52 צריך להכליל את זה להשוואות, והמושג של מינימיזציה מעורפל יותר במקרה זה, ולא ברור מה אפשר להגיד עליו: אטום יכול להיות מיותר במקרה אחד, ולא מיותר במקרה שני.

17/6/2013

הרצאה 12

חזרה קצרה:

עבור חוקים שלא מכילים השוואות, תנאי הכרחי ומספק להכלה בין שני חוקים הוא קיום הומומורפיזם מהמכיל למוכל.

כאשר יש השוואות, ראינו הגדרה מתאימה להומומורפיזם. במקרה זה קיום הומומורפיזם הוא מספיק, אך לא הכרחי. תנאי הכרחי ומספיק היה אם לכל הרחבה $\bar{\Gamma}_2$ של Γ_2 ליחס סדר מלא מעל כל הסימנים של Q_2 והקבועים של Γ_1 יש הומומורפיזם מ- Q_1 ל:

$$\bar{Q}_2 : H_2 : -B_2, \bar{\Gamma}_2$$

התוספת שהכנסנו הם "הקבועים של Γ_1 " - זה נכון לחוקים בודדים, אך ללא זה ההוכחה קצת יותר ארוכה (יש להוסיף את ההשוואות המתאימות). נשים לב כי אם יש קבוע ב- B_1 שלא נמצא ב- B_2 , לא תיתכן הכלה - אין למה למפות את האטום שמכיל את אותו הקבוע.

למען הפשטות והאחידות תמיד נכלול את התוספת הזו.

בשבוע שעבר התחלנו לדבר על איחודים עם השוואות. שם היינו חייבים לכלול את התוספת, וראינו באותה הרצאה מדוע. בתרגיל 1 היינו צריכים להוכיח שהתנאי הנ"ל הוא גם מספיק.

לאחר מכן דיברנו על חוקים עם שלילה. ניסחנו את תנאי הבטיחות למקרה הכללי. הבאנו תנאי מספיק להכלה, המותאם למקרה שבו יש שלילה, והגדרנו הומומורפיזם במקרה שאין בו השוואות אך יש בו שלילה. דנו במקרה המיוחד בו אין פרדיקט שמופיע בגוף של חוק גם באופן חיובי וגם באופן שלילי - במקרה זה התנאי גם הכרחי. הוכחנו זאת.

מספיק להכיל את המקרה המיוחד רק על Q_1 ולא על Q_2 : ב- Q_2 השתמשנו בזה רק כדי להראות שהוא מייצר משהו מהמסד הקנוני של עצמו. אם כך, נוכל לוותר על הדרישה הזו אם נדע לבדוק אם חוק מייצר משהו או אם הוא ריק.

3.4 הכללות למקרה הכללי

3.4.1 ריקנות

כשבדקים הכלה, צריכים תחילה להביא בחשבון האפשרות שהחוק הוא ריק (כלומר, לכל מסד נתונים, התוצאה של הפעלת החוק היא הקבוצה הריקה). דוגמא לחוק מעניין ובטוח - שקופית 3:

$$P(7) : -R(1, 2)$$

אם במסד מופיע האטום $R(1, 2)$, אז תוצאת החוק ריקה. אחרת, תוצאת החוק היא $\{P(7)\}$. לכן, חוק זה לא ריק.

מתי חוק עם השוואות ושלילה הוא ריק? ברור שאם בגוף של החוק אותו אטום מופיע גם באופן חיובי וגם באופן שלילי, אז החוק ריק. כלומר, הופעה בגוף של אותו אטום גם באופן חיובי וגם באופן שלילי היא תנאי מספיק לכך שהחוק ריק. תחילה נדון בחוק עם שלילה וללא השוואות, במטרה למצוא תנאי הכרחי.

שלילה בלי השוואות

ננסה את הטכניקות הישנות: ניקח חוק, ונראה אם הוא מייצר משהו מהמסד הקנוני. אם הוא מייצר את ראש הקנוני של עצמו מהמסד הקנוני, אז הוא לא ריק. נשים לב כי חוק מייצר את הראש הקנוני של עצמו מהמסד הקנוני של עצמו אמ"מ אין בו אטום שמופיע בגוף גם באופן חיובי וגם באופן שלילי. לכן התנאי במקרה זה גם הכרחי.

המקרה הכללי כלומר, נדון במקרה בו בגוף החוק יש גם השוואות וגם אטומים רגילים וגם שלילה. ניקח דוגמאות פשוטות: תחילה, אם מתקיימים אחד משני התנאים שהזכרנו (הופעת אטום ושלילתו, וכן הופעת השוואה שאינה ספיקה), החוק ריק. אחרת, איך עדיין יכול להיות שהחוק יהיה ריק? דוגמא בשקופית 7:

$$P(X) : -R(X, Y), \neg R(Y, X), X = Y$$

שני התנאים שהזכרנו לא מתקיימים, אך ברור כי החוק ריק! מה עשינו במקרה זה? "החלפנו" משתנים לפי ההשוואות, ואז האטום הראשון למעשה כן מופיע גם בחיוב וגם בשלילה, כלומר התנאי ריק. נשים לב כי התנאי השני לא יכול "פתאום" להתקיים: החלפה של משתנים לא יכולה לגרום לגימום של השוואות להפוך לבלתי ספיק.

תנאי הכרחי ומספיק לריקנות במקרה הכללי: תחילה, בודקים אם ההשוואות ניתנות סיפוק. אם לא, החוק ריק. אם כן, צריך להשוות סימנים לפי השיויונות שנגררים מההשוואות. עושים זאת באופן הבא: מייצרים גרף מכיוון מהשוואות, כפי שהוסבר בעבר, ומחלקים אותו לרק"חיס. בכל רק"ח בוחרים סימן באופן הבא:

אם ברק"ח יש קבוע, בוחרים אותו. שימו לב שברק"ח יש לכל היותר קבוע אחד, כי הנחנו שההשוואות ניתנות לסיפוק. אם ברק"ח אין קבוע, בוחרים באופן שרירותי את אחד המשתנים של הרק"ח. הסימן שנבחר מחליף בכל החוק את כל המופעים של המשתנים ששייכים לרק"ח. עושים זאת לכל רק"ח.

משפט 3.53 חוק הינו ריק אמ"מ:

1. ההשוואות אינן ספיקות, או:

2. לאחר כל ההחלפות יש אטום רגיל שמופיע בגוף גם באופן חיובי וגם באופן שלילי.

הוכחה: תחילה נוכיח שהתנאי מספיק.

אם ההשוואות אינן ספיקות, ברור שהחוק ריק. אחרת, נבצע ההחלפות כדרש. ההחלפות שומרות על שקילות, כי הן מתבצעות לפי שיויונות שנגררים מההשוואות. לכן, אם בסיום ההחלפות יש אטום רגיל שמופיע גם באופן חיובי וגם באופן שלילי, אז החוק המקורי ריק.

עכשיו נוכיח שהתנאי הכרחי. נניח שהתנאי אינו מתקיים, ונראה שהחוק אינו ריק. מכיוון שהתנאי אינו מתקיים, ההשוואות ספיקות. נבצע החלפת סימנים, כפי שתואר קודם. כאמור, החלפה זו שומרת על שקילות לחוק המקורי. מכיוון שההשוואות ספיקות, אפשר להרחיב אותן לסדר מלא על כל סימני החוק. נייצר מסד קנוני מהאטומים החיוביים בחוק, וההרחבה שביצענו לסדר מלא תגדיר את יחס הסדר על המסד הקנוני. מכיוון ואין אטום המופיע בגוף גם באופן חיובי וגם באופן שלילי, החוק (לאחר ההחלפות) מייצר את הראש הקנוני של עצמו מהמסד הקנוני. לכן, חוק זה אינו ריק, ובגלל השקילות כך גם החוק המקורי. ■

הערה 3.54 ההרחבה לסדר מלא חייבת לקיים שאין שני סימנים שונים המשווים זה לזה. ברור שאין הרחבה כזו, כי כבר החלפנו סימנים לפי כל השיויונות הנגררים מההשוואות. לכן, מיפוי הסימנים, שממפה כל משתנה X לקבוע הייחודי המתאים לו X_0 , מקיים את כל ההשוואות. לפיכך, מיפוי זה מייצר את הראש הקנוני מהמסד הקנוני.

3.4.2 בדיקת הכלה $Q_1 \supseteq Q_2$ בין שני חוקים עם שלילה וללא השוואות

אחת השאלות שהיתה בבית דיברה על בדיקת הכלה בין שני חוקים Q_1, Q_2 בלי השוואות אולם עם שלילה במקרה הכללי, כלומר פרדיקט יכול להופיע בגוף באופן חיובי וגם באופן שלילי ב- Q_1 . כרגיל, נסמן:

$$Q_1 : H_1 : -B_1, N_1$$

$$Q_2 : H_2 : -B_2, N_2$$

כאשר B_i היא קבוצת אטומים חיוביים, ו- N_i היא קבוצת אטומים שליליים. נרצה לבדוק אם $Q_1 \supseteq Q_2$. שקופיות 17 והלאה מתארות אלגוריתם למקרה זה:

תחילה בודקים אם Q_2 ריק. אם כן, יש הכלה. אחרת מבצעים את התהליך הבא: מייצרים מסד קנוני מ- B_2 ; נסמן את המיפוי היוצר אותו ב- g . יהי f מיפוי קרקעי כלשהו, שמראה ש- Q_1 מייצר את הראש הקנוני של Q_2 . לכל חסימת f ע"י אטום בודד, מייצרים בן שמכיל את B_2 והאטום החוסם. זכרו שאטום חוסם מתקבל מ- N_i כך ש- $n_i \in N_i$ ו- $f(n_i) \notin g(N_2)$. ממשיכים באופן רקורסיבי. כלומר, לוקחים עלה, מוצאים מיפוי f שמראה ש- Q_1 מייצר את הראש הקנוני של Q_2 מהמסד הנוכחי, ומייצרים בן לכל אטום שחוסם את f . נשים לב כי המסד הנוכחי מורכב מ- $g(B_2)$ ומהאטומים החוסמים שהצטרפו לאורך המסלול מהשורש לעלה. f יכול למפות אטומים של B_i לאטומים שהצטרפו עקב חסימות. השאלה היא, מתי מפסיקים את התהליך הזה? כלומר, מתי עלה הוא עלה ולא ניתן לייצר לו בנים? שתי אפשרויות -

1. אין מיפוי f שמראה ש- Q_1 מייצר את הראש הקנוני של Q_2 מקבוצת האטומים בצומת. לכן, קבוצה זו היא דוגמא נגדית להכלה.

2. יש מיפוי שמראה זאת, אבל אי אפשר לחסום אותו. במקרה זה, הצומת איננה דוגמא נגדית.

האטומים שמוסיפים מורכבים תמיד מפרדיקטים וסימנים המופיעים בחוק המקורי Q_1 . לכן, יש מספר סופי של אטומים שניתן להוסיף. לכן, התהליך מסתיים בזמן סופי. אם בסוף התהליך יש עלה שהוא דוגמא נגדית (כלומר, תנאי 1 מעלה) אז אין הכלה, אחרת יש הכלה.

פתרון שאלות מהרצאות קודמות

נמצאות באופן מסודר בשקופיות הקורס.

על המבחן

לא ידרש לדעת רדוקציות לבעיות NP קשות או שלמות, וכו'.
השאלה שהיתה בתחילת החומר: בעיית הריקנות למקרה הכללי - רדוקציה מ- $3SAT$ לבעיה של $JOIN$ לא ריק. מה נשים ביחס? כל ההשמות המספקות.

כדי לקבל קרדיט מלא, צריכה התשובה להופיע באופן ברור, ולא צריכים להופיע בתשובה דברים לא נכונים.
מכיוון ורשומים הרבה סטודנטים, יתכן והמתכונת תשתנה מפורמט של שאלות פתוחות עם סעיפים לפורמט דומה לזה של דיבי.
בשנים קודמות - 6 שאלות ללא בחירה עם מספר סעיפים - חלקם מורכבים, חלקם פשוטים, כמו הגדרה למשל. להשתדל לא לענות על מה שלא נשאלנו!
יכולות להופיע וריאציות של דברים שראינו. חלק מהשאלות יהיו ספציפיות על חומר מההרצאה, ואפשר לעבור אם יודעים את כל החומר מההרצאה. כדי לקבל 100 צריך להתאמץ יותר, ולדעת יותר מחזרה מדויקת על מה שנאמר בכיתה.