

קורס מערכות מידע

החוג למדעי המחשב
מכללת תל חי

מצגת מספר 5

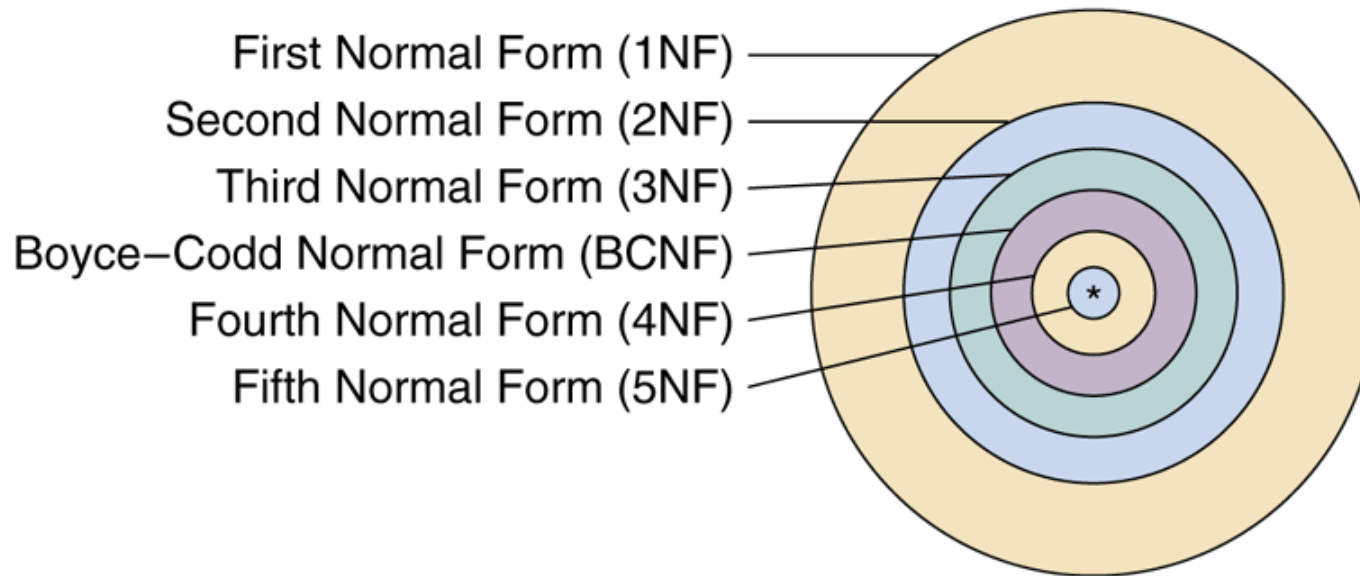
המודל הרלציוני ושיטת הנרמול

גב' אביבה עבדל

נרמול טבלאות

תהליך שיטתי ופורמאלי למניעת אנומליות הנובעות מכפילויות נתונים.

הנרמול מכיל ששה חוקים לארכיטקטורה תקינה של מסד נתונים טבלאי, חוקים אלו הבנויים אחד על גבי השני קובעים מבנה אופטימאלי של טבלאות ושדות בבסיס הנתונים, כך שהמידע יאוחסן באופן האמין, היעיל והחסכוני ביותר, ובכדי למנוע תופעות לא רצויות בתפעול בסיס הנתונים.



נרמול טבלאות

הכלל העקרוני והעיקרי של חוקי הנרמול
קובע שכל טבלה תתייחס לחטיבת מידע
אחת בלבד, וכל הנתונים שלה יתארו אך
ורק את אותה חטיבת מידע.

נרמול טבלאות

כאשר מעצבים סכמת בסיס נתונים רלציוני מעוניינים להשיג את המטרות הבאות:

✓ **פשטות ומובנות של רלציות –**

(כל רלציה צריכה להיות פשוטה להבנה, השדות שהיא מורכבת מהם צריך לתאר ישות אחת ולא לכלול שדות המתארים ישויות שונות)

✓ **מניעת כפילויות ובעיות עדכון –**

(משמעות של כפילות שאותם נתונים נשמרים ביותר ממקום אחד בבסיס הנתונים, בכך נגרם בזבוז שטח אחסון)

✓ **מניעת ערכים ריקים בשדות –**

(ערכים ריקים גורמים לבזבוז שטח אחסון ועלולים להקשות על הבנת הרלציה)

מתי סכמה איננה טובה?

דוגמה:

Student	Depr	Head
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

המידע לגבי ראש מחלקה מופיע הרבה פעמים – פעם אחת עבור
כל סטודנט השייך למחלקה

מדוע שכפול מידע אינו טוב?

ראשית, בזבוז של מקום בזיכרון
שנית, כאשר אותו מידע מופיע במספר מקומות במסד
הנתונים, צריכה להיות מודעות לכך בזמן ביצוע עדכונים

חייבים לעדכן את המידע בכל המקומות בהם הוא מופיע

דוגמה לעדכון שגוי

מעדכנים את ראש המחלקה של CS
רק ברשומה אחת, וכתוצאה מכך יש
נתונים סותרים לגבי ראש המחלקה
של CS

S	D	H
Levy	CS	Katz
Cohen	Math	Bush
Barak	CS	Rubin

כאשר מידע משוכפל במספר מקומות במסד הנתונים,
הניסיון מראה שלאורך זמן יש סיכוי גדול שהמידע יעודכן
רק בחלק מהמקומות, וכתוצאה מכך יהיו במסד הנתונים
פריטי מידע הסותרים זה את זה

מדוע נוצר מצב של שכפול מידע?

Student
Department
Head

S	D	H
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

למחלקה יש רק מנהל אחד אבל יש הרבה סטודנטים
לכן, המידע על מנהל המחלקה נשמר עבור כל סטודנט השייך
למחלקה

בעיות שנובעות מביטול שורות

דוגמה:

Student	Depr	Head
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

אם נרצה לבטל את הרשומה:

Cohen, Math, Bosh

התוצאה:

Student	Depr	Head
Levy	CS	Rubin
Barak	CS	Rubin

איבוד מידע



בעיות שנובעות מהוספת שורות

שאלה:

נרצה להוסיף ראש מחלקה של physics ואין לו עדיין סטודנטים, להיכן נוסיף אותו?

Student	Depr	Head
Levy	CS	Rubin
Cohen	Math	Bush
Barak	CS	Rubin

מה יהיה
המפתח של
הישות???

מפתח של הסכימה הוא : S , D

אם נפתחה מחלקה חדשה ויש לה ראש מחלקה אבל עדיין אין סטודנט נוצרת בעיה משום שהערך של התכונה S הוא NULL.

לא אפשרי כאשר S הוא חלק ממפתח

תכנון סכמות רלציוניות

Customer

Cust_Id	Faculty	Track
12345	CS	Software
45678	EE	Hardware
11111	IE	IS
22222	IE	Accounting

Ordered

Cust_Id	Book_Name
12345	<i>Database Systems</i>
45678	Anatomy
12345	Database And Knowledge
11111	Anatomy
22222	Intro. To Economy

תכנון א'

CustomerOrders

Cust_Id	Faculty	Track	Book_Name
12345	CS	Software	<i>Database Systems</i>
45678	EE	Hardware	Anatomy
12345	CS	Software	Database And Knowledge
11111	IE	IS	Anatomy
22222	IE	Accounting	Intro. To Economy

תכנון ב'

איזה תכנון טוב יותר? למה?

תכנון סכמות רלציוניות

חסרונות התכנון עם טבלה אחת:

בטבלה תופעות לא רצויות הנובעות משילוב נתונים משתי ישויות שונות

■ כפילות הגורמת:

✓ בזבוז מקום בזיכרון (פרטי לקוח נשמרים כמספר הספרים שהוא הזמין).

✓ עדכון יותר קשה – עדכון פרטי סטודנט בכל הספרים שהזמין).

✓ אי דיוק בנתונים - במקרה ולקוח ביצע מספר הזמנות - שינוי כלשהו בפרטי לקוח

מצריך עדכון בכל השורות הרלוונטיות.

✓ אם עודכנו רק חלק מהרשומות, המסד לא עקבי.

■ קשה לייצג לקוח שלא הזמין אף ספר.

■ ביטול הזמנה שהינה ההזמנה היחידה של סטודנט מסוים, יוצר איבוד נתונים

נוספים (פרטי סטודנט)

נרמול נתונים

במטרה להתגבר על בעיות אלו פותחה שיטת הנרמול.

היעד של שיטת הנרמול הוא:

יצירת סכמת בסיס נתונים המורכבת מרלציות
פשוטות, שכל אחת מבטאת ישות מסוימת ותוך
מניעת כפילויות בנתונים

החיסרון: סרבול בתהליכי בחירה ושליפה.



תלויות פונקציונליות

Functional Dependency

תלות פונקציונלית



$B \leftarrow A$

ערך שדה A קובע חד-ערכית את ערכו של שדה B
או ערך שדה B תלוי בערך שדה A

לדוגמא:

מ"ז \leftarrow שם_סטודנט, ש. לידה

(מ"ז קובע שם של תלמיד אחד, אך יתכנו מספר תלמידים עם אותו שם)

מ"ז \leftarrow מחלקה (תלמיד מסוים שייך למחלקה אחת, אך למחלקה מסוימת יתכנו מספר תלמידים)



ת.ז.	שם פרטי	ש. לידה
111	אליהו	1970
222	יניב	1971
333	אליהו	1965

מקרה פרטי של תלות פונקציונלית

$B <----> A$

ערך שבו שדה A קובע את ערכו של B וגם להפך.
זהו קשר חד-חד-ערכי, אחד-לאחד, 1:1

לדוגמא:

קוד מחלקה <----> מ"ז מנהל

למחלקה יש מנהל אחד ומנהל יכול לנהל מחלקה
אחת בלבד

תלות מורכבת (מרובה)

B <<--->> **A**

ערך של שדה A יכול להתקשר למספר ערכים של שדה B ולהיפך
קיים קשר רב-רב-ערכי, רבים לרבים

דוגמא:

ת.ז. סטודנט <<--->> קוד קורס



ת.ז. סטוד'	קוד קורס	ציון
111	A	70
111	B	90
222	A	80

טבלת ציונים

מייצג קשר
N:M

תלות עוברת (טרנזיטיבית)

תלות פונקציונלית עקיפה

אם $B \leftarrow A$ ו $C \leftarrow B$ אזי $C \leftarrow A$

(נכון אבל בעייתי)

לדוגמא:

■ מ"ז תלמיד \leftarrow קוד מחלקה \leftarrow תקציב

אזי: מ"ז תלמיד \leftarrow תקציב

■ אם קוד פריט \leftarrow קוד ספק וגם קוד ספק \leftarrow כתובת

אזי: קוד פריט \leftarrow כתובת

כללי הכרמל

כללי נרמול

- בתורת הנרמול יש כללים אחדים המתייחסים לסוגי תלויות הקיימים בין שדות.
- באמצעות הכללים ניתן לבחון אם רלציה קיימת תקינה, ואם לא – לתקן אותה.
- התיקון בד"כ מתבטא בפרוק של רלציות
- **התוצאה:** רלציות העונות על הכללים = נמצאות בדרגת נרמול

אנו נלמד את דרגות הנרמול:

✓ דרגת נרמול ראשונה 1NF

✓ דרגת נרמול שנייה 2NF

✓ דרגת נרמול שלישית 3NF

דרגת נרמול ראשונה - 1 NF

כל רלציה במודל הרלציוני היא לפחות בדרגת נרמול ראשונה (First Normal Form), כלומר יש לה צורת טבלה, שדותיה מכילים ערכים אטומיים (לא ניתנים לפירוק. שימו לב, תכונה מורכבת ומרובת ערכים אינה אטומית) והמפתח שלה מכיל ערכים (אינו ריק).

דרגת נרמול ראשונה - 1 NF

דרגת נרמול ראשונה 1NF היא נקודת המוצא לנרמול נתונים

כדי להשיג נרמול בדרגה 1NF:

✓ עמודה מורכבת, תפוצל לעמודות המכילות שדות אטומיים.

✓ עבור שדה חוזר ניצור טבלה חדשה שתכלול את השדה החוזר +

מפתח הטבלה המקורית.

דוגמא:

document: title, author-list, date (year, month, day), keywords


document: title, {author}, date (year, month, day), {keyword}

דרגת נרמול שנייה - 2 NF

רלציה היא בדרגת נרמול שנייה (Second Normal Form), אם כל שדה שאינו שדה מפתח תלוי פונקציונלית במפתח כולו.

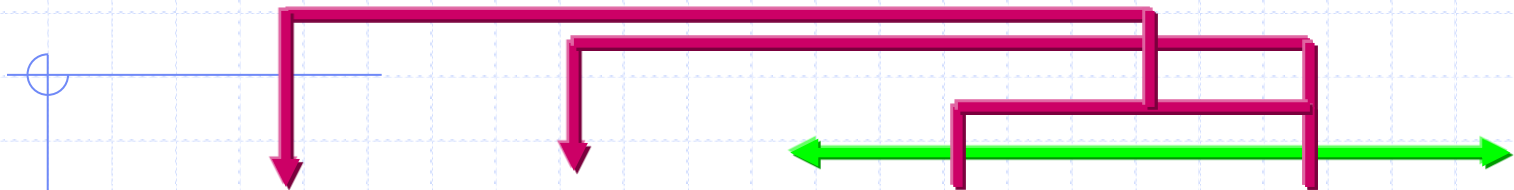
הערה: כלל זה חל על רלציות שהמפתח שלהן מורכב מיותר משדה אחד.

דוגמא - נתונה הרלציה הבאה:



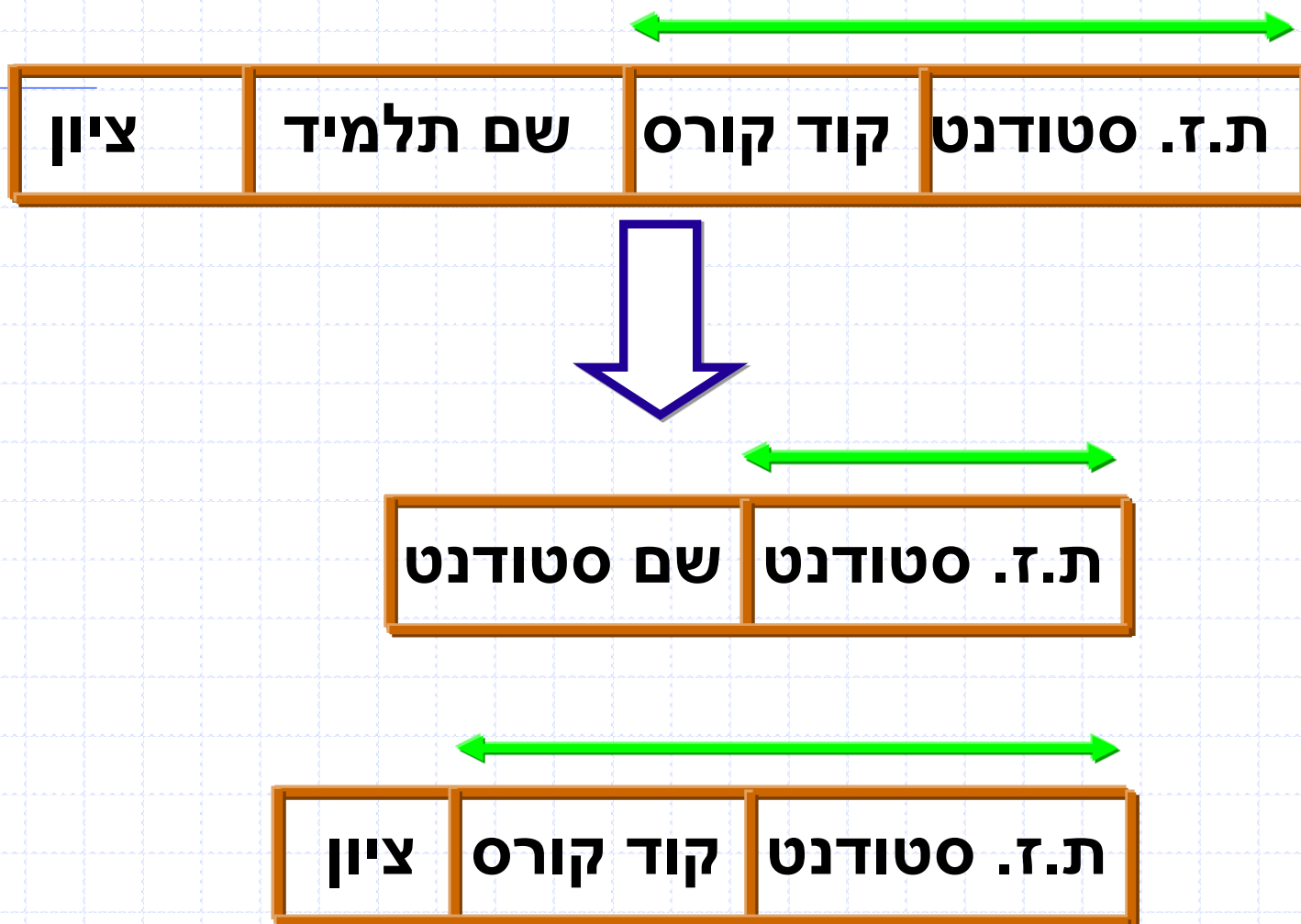
ת.ז. סטודנט	קוד קורס	שם תלמיד	ציון
-------------	----------	----------	------

הבעיה



ציון	שם תלמיד	קוד קורס	ת.ז. סטודנט
80	אלי כהן	A	111
100	אלי כהן	B	111
70	אלי כהן	C	111
60	תומר שדה	A	222
85	תומר שדה	C	222

הפתרון



דרגת נרמול שלישית - 3 NF

רלציה היא בדרגת נרמול שלישית (Third Normal Form),
אם כל שדה שאינו שדה מפתח אינו תלוי פונקציונלית בשום
שדה אחר שאינו במפתח.

במילים אחרות - אסורה תלות טרנזיטיבית

דוגמא - נתונה הרלציה הבאה:



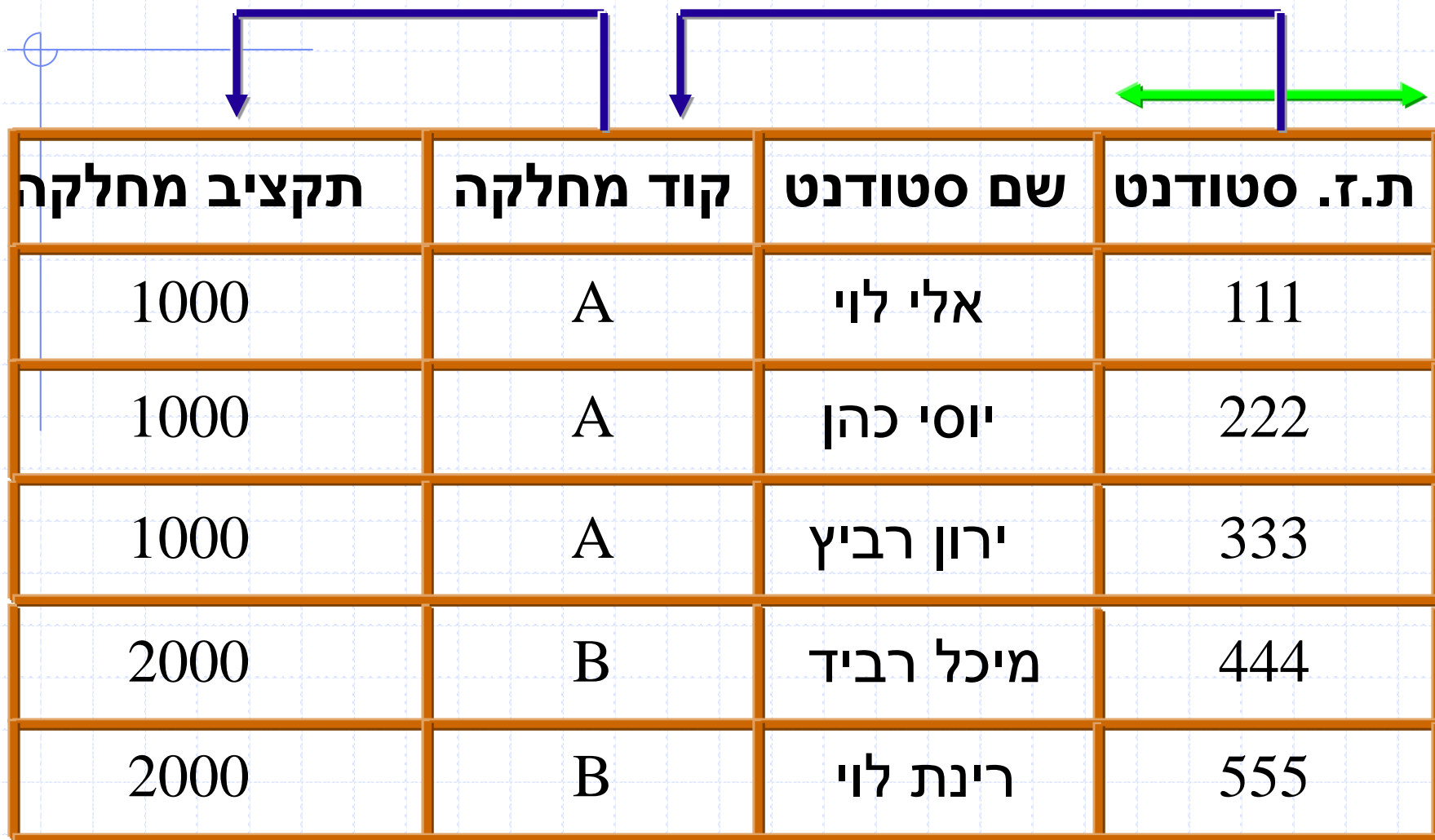
ת.ז. סטודנט	שם סטודנט	קוד מחלקה	תקציב מחלקה
-------------	-----------	-----------	-------------

הבעיה



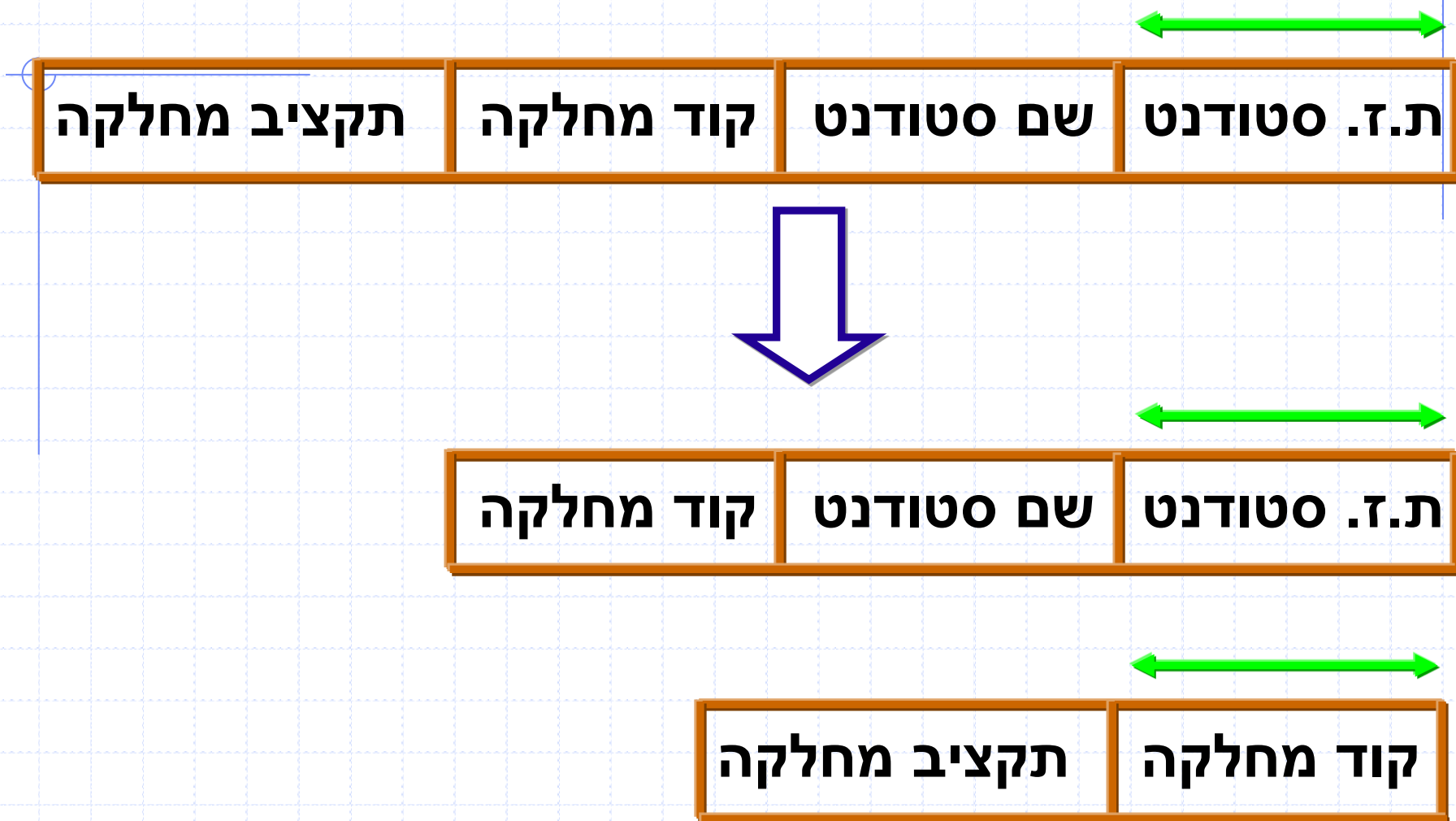
ת.ז. סטודנט	שם סטודנט	קוד מחלקה	תקציב מחלקה
111	אלי לוי	A	1000
222	יוסי כהן	A	1000
333	ירון רביץ	A	1000
444	מיכל רביד	B	2000
555	רינת לוי	B	2000

הבעיה



ת.ז. סטודנט	שם סטודנט	קוד מחלקה	תקציב מחלקה
111	אלי לוי	A	1000
222	יוסי כהן	A	1000
333	ירון רביץ	A	1000
444	מיכל רביד	B	2000
555	רינת לוי	B	2000

הפתרון



סיכום כללי הנרמול

רלציה מנורמלת עומדת ב - 3 דרגות
הנרמול דהיינו:

כל שדה ברלציה שאינו במפתח צריך
להיות תלוי פונקציונלית אך ורק
במפתח כולו (ולא בשום שדה אחר)

תרגיל

בחברת שיווק גדולה הוחלט לתעד בבסיס הנתונים את בני/ות הזוג של הלקוחות כדי לבצע שיווק אגרסיבי וממוקד יותר הפונה למשפחות הלקוחות, באמצעות הטבלה הבאה:

מס' ילדים	תאריך נישואין	שנת לידת בן/בת הזוג	שם בן/בת הזוג	ת.ז. בן/ת הזוג	שם לקוח/ה	ת.ז. לקוח/ה
1	12/9/02	1978	רונית בן-בסט	204835	אלי סוויסה	215657
2	10/8/01	1979	מירי גולדפרב	338015	בני וקס	318742
1	8/8/02	1976	עידו הראל	345602	רונית לוי	324187
0	14/10/01	1978	מירב כרמל	348106	יוסף כהן	322456

✓ הסברו האם טבלה זו מנורמלת ?

✓ אם עניתם לא, הגדירו תלויות פונקציונאליות

✓ קבעו רלציות ומפתחות מתאימים לתלויות שמצאתם

תרגיל

נתון המאגר "עבודה" הבא המתייחס לעובדים המשמשים כנציגי מכירות וללקוחות שקנו מהם פריט כלשהו:

עבודה (מספר עובד, שם (פרטי + משפחה), קוד תפקיד, תיאור תפקיד, מספר לקוח, שם לקוח, מספר פריט, כמות, תיאור פריט)

נתונות התלויות הפונקציונליות הבאות:

✓ מספר עובד ← שם פרטי, שם משפחה, קוד תפקיד

✓ קוד תפקיד ← תיאור תפקיד

✓ מספר פריט ← תיאור פריט

✓ מספר לקוח ← שם לקוח

✓ ((מספר עובד << -- >> מספר לקוח) << -- >> מספר פריט) ← כמות

הצג את סכמת הרלציות למאגר "עבודה" לאחר נרמול (בדרגת 3NF) – הצג את השינויים לכל דרגה

תרגיל

יש להקים בסיס נתונים למערכת מידע שתנהל נתונים על הצעות מחקר שמגישים מרצים באוניברסיטה לקרנות המחקר.

מרצה שייך למחלקה אחת ומגיש מעת לעת הצעות לביצוע מחקרים שונים. לכל מחלקה באוניברסיטה יש קוד זיהוי, שם ומנהל (שהוא מרצה חוקר). את הצעת המחקר יכול להגיש מרצה אחד או מרצים אחדים. להצעה יש שם שנותן לה החוקר. ביחידה לקרנות מחקר שבאוניברסיטה ניתן לכל הצעה קוד מזהה ייחודי. ההצעה מוגשת לקרן מחקר מסוימת המזוהה בשם. כשמוגשת הצעת מחקר לראשונה מציינים בה כמה שנים ימשך המחקר ולכל שנה מפרטים את הסכום המבוקש ואת חוקרי המשנה שישתתפו במחקר, תוך פירוט מספרי הזיהוי של חוקרי המשנה, פירוט שמותיהם ופירוט מספר החודשים שכל אחד מהם יעבוד במחקר. בתום כל שנת מחקר מוגשת ההצעה שוב לקרן שדנה באישורה. במערכת המידע מעדכנים מעת לעת את סטטוס הבקשה בכל שנה (למשל "הוגש", "בדיון", "אושר"). אם בקשה מאושרת מציינים את תאריך האישור ואת הסכום שאושר.

בצע:

- א. מצא את רשימת התלויות בין השדות לפי המתואר בתרגיל.
- ב. קבע את הרלציות ואת מפתחותיהן לפי התלויות שמצאת בסעיף א'.