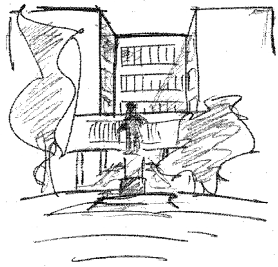
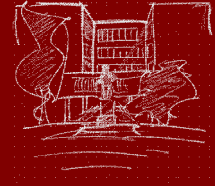


[P273] **Пројектовање база података** 4.3



Саша Малков
Универзитет у Београду
Математички факултет
2023/2024

[P273] **Пројектовање база података**
Саша Малков



Тема 7.2

**Релациони модел података
(манипулативни део)**

Манипулативни део модела



- Манипулативни део релационог модела дефинише формалне начине руковања моделираним подацима
- Кључно место у манипулативном делу модела има појам упита:
 - *Упити* је дефиниција (или израчунавање) нове релације на основу већ познатих релација базе података

Формални упитни језици



- Најважнији формални упитни језици у релационом моделу су:
 - Релациона алгебра и
 - Релациони рачун



Релациона алгебра

- Релациона алгебра је проширење скуповне алгебре
 - Задржавамо основне скуповне операције
 - Унија, Пресек, Разлика, Производ
 - Код производа подразумевамо да је $(A \times B) \times (C \times D) = A \times B \times C \times D$
 - То је у реду зато што претпостављамо атомичност атрибута
- Основне додатне операције су:
 - Пројекција
 - Рестрикција
 - Природно сјајање
- Резултати свих ових операција су нове релације



Имена релација и атрибута

- У случају производа и неких других бинарних операција може да се догоди да резултат има два атрибута истог имена
 - у том случају се назив атрибута допуњава слева називом релације из које потиче

$$R(A,B) \times Q(B,C) = W(A, R_B, Q_B, C)$$
- Ако је потребно да се у операцији употреби више пута иста релација, да не би долазило до неспоразума уводе се *алиаси* (надимци) релација

$$\text{DEFINE ALIAS R1 FOR R}$$

$$R \dots R1$$



Пројекција

- **Пројекција** се дефинише као избор колона X једне релације:

$$R[X] = \{ x \mid x \in \text{Dom}(X) \wedge (\exists y \in \text{Dom}(Y)) (x, y) \in R \}$$
 - где су X и Y комплементарни скупови атрибута релације $R: Y=A(R)\setminus X$
- Резултат пројекције може да има мање елемената него полазна релација
 - Због тога што се ради о скуповима n -торки, пројекција не може да садржи поновљене елементе
 - Резултат пројекције сигурно има исти број елемената као релација R ако је X наткључ релације R
- У терминима табела, пројекција је вертикално “сечење” табеле
- Пример:
 - $\text{RADNIK}[\text{ime}, \text{prezime}] = \{ (x.\text{ime}, x.\text{prezime}) \mid x \in \text{RADNIK} \}$



Рестрикција

- **Рестрикција** се дефинише као избор торки релације R у којима дати атрибут $X \in A(R)$ има дату вредност $x \in \text{Dom}(X)$:

$$R[X=x] = \{ e \mid e \in R \wedge X(e) = x \}$$
 - Слободније посматрано, X може да буде произвољан подскуп атрибута
- Општије, рестрикција по произвољном предикату $p: R \rightarrow \{T, \perp\}$ је:

$$R[p] = \{ e \mid e \in R \wedge p(e) \}$$
- У терминима табела, рестрикција је хоризонтално “сечење” табеле
- Пример:
 - $\text{RADNIK}[\text{org_jed_id} = 40] = \{ x \mid x \in \text{RADNIK} \wedge x.\text{org_jed_id} = 40 \}$



Природно спајање

- **Природно спајање** релација R и Q са заједничким скупом атрибута X дефинише се као:

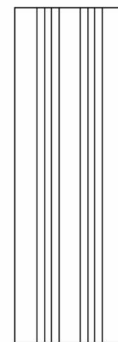
$$R * Q = \{ (e, f) \mid e \in R \wedge f \in Q \wedge X(e) = X(f) \} [X \cup Y \cup Z]$$

$$= \{ (x, y, z) \mid (x, y) \in R \wedge (x, z) \in Q \}$$
- где су:
 - $Y = A(R) \setminus X$
 - $Z = A(Q) \setminus X$
- Спајају се торке једне релације са торкама друге релације које имају једнаке вредности заједничких атрибута
- Резултат садржи само једну копију заједничких атрибута
- Резултат садржи најмање нула торки
 - ако нема торки са истим вредностима заједничких атрибута
- Резултат садржи највише $M \times N$ торки, где су M и N бројеви торки релација R и Q
 - ако све торке обе релације имају исте вредности заједничких атрибута
- Пример:
 - $RADNIK * ORG_JEDINICA =$
 $\{ (org_jed_id, x, y) \mid (org_jed_id, x) \in RADNIK \wedge (org_jed_id, y) \in ORG_JEDINICA \}$

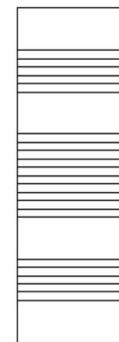


Илустрација операција

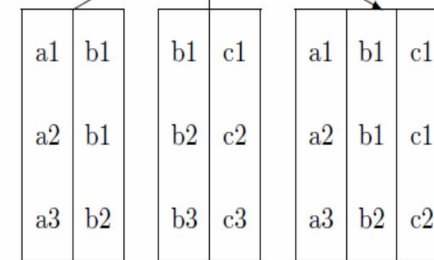
пројекција



рестрикција



природно спајање



Природно спајање (2)

- Општији случај природног спајања је када релације R и Q имају подскупове атрибута X и Y који нису исти, али су исте величине и истих домена.
- Тада се природно спајање дефинише и записује као:

$$R[X*Y]Q = \{ (z, x, w) \mid (z, x) \in R \wedge (y, w) \in Q \wedge x=y \}$$



Додатне операције

- Дефинишу се и друге операције, као:
 - *слободно спајање*
 - *дељење*



Слободно спајање

- **Слободно спајање** релација R и Q по датом бинарном предикату $p: R \times Q \rightarrow \{T, \perp\}$ се дефинише као:

$$R[p]Q = \{ (e, f) \mid e \in R \wedge f \in Q \wedge p(e, f) \}$$
- Спајају се торке једне релације са торкама друге релације са којима испуњавају дати услов
- Слободно спајање је општије од природног



Дељење

- **Дељење** релације R релацијом Q по подскупу атрибута $Y \subseteq A(R)$, где је $Z = A(Q)$ скуп атрибута који по броју и домену одговара скупу Y , дефинише се као:

$$R[Y \% Z]Q = \{ x \mid x \in R[X] \wedge \{x\} \times Q[Z] \subseteq R \}$$
- где је $X = A(R) \setminus Y$
- где је $Z = A(Q)$ скуп атрибута који по броју и домену одговара скупу Y
- Резултат је максимални подскуп релације $R[X]$ такав да је његов производ са Q подскуп релације R
- издвајају се оне торке атрибута Y које су у R упарене са свим торкама из Q
- Пример - издвојити раднике који су радили на свим пројектима:
`RESENIJE[RAD_ID, PROJ_ID] [PROJ_ID % PROJEKAT_ID] PROJEKAT[PROJEKAT_ID]`



Проширена релациона алгебра

- **Проширена релациона алгебра** обухвата рад са **недефинисаним вредностима**
 - Без обзира на разлог недефинисане вредности, користи се једна иста ознака недефинисане или недостајуће вредности
 - Сваки домен се проширује недефинисаном вредношћу
- Тиме се мења и структурни део модела, али су важније промене у манипулативном делу, па зато то обрађујемо овде
- Скуп истинитосних вредности се проширује **непознатом логичком вредношћу**
 - Све релације поређења неке вредности са недефинисаним вредностима имају за резултат **непознато**
 - Логичке операције се проширују да раде са непознатом вредношћу



Проширене логичке операције

\wedge	T	F	n
T	T	F	n
F	F	F	F
n	n	F	n

\vee	T	F	n
T	T	T	T
F	T	F	n
n	T	n	n

\neg	
T	F
F	T
n	n



Проширена рестрикција

- Уводе се нове врсте рестрикције
 - **тачна рестрикција**, одговара оригиналној рестрикцији
 - издвајају се торке за које је вредност услова **тачно**
 - означава се навођењем индекса услова "Т"
 - $R[p_T] = R[p]_T = \{ x \mid x \in R \wedge p(x) = T \}$
 - **можда рестрикција** представља нову операцију
 - издвајају се торке за које је вредност услова **непознат**
 - означава се навођењем индекса услова "n" (или неки други знак, зависно од извора)
 - $R[p_n] = R[p]_n = \{ x \mid x \in R \wedge p(x) = n \}$
 - **не-лажна рестрикција** представља унију претходне две
 - издвајају се торке за које вредност услова **није лажан**
 - може да се означава комбинацијом претходна два "Tn"
 - $R[p_{Tn}] = R[p]_{Tn} = \{ x \mid x \in R \wedge p(x) \neq \perp \}$
 - строго формално $p(x) \neq \perp$ има недефинисану вредност за недефинисано $p(x)$ па је исправније $R[p_{Tn}] = R \setminus R[\neg p_T]$



Проширено спајање

- Операције спајања почивају на поређењима и проверама вредности предиката, па и оне могу да се прошире:
 - Природно спајање
 - **тачно** природно спајање
 - **можда** природно спајање
 - **не-лажно** природно спајање
 - Слободно спајање
 - **тачно** слободно спајање
 - **можда** слободно спајање
 - **не-лажно** слободно спајање



Проширене скуповне операције

- **Спољашња унија** релација R и Q са заједничким пресеком атрибута X дефинише се као:

$$R \dot{\cup} Q = (R \times (Z:n)) \cup ((Y:n) \times Q)$$
 где су:
 - $Y = A(R) \setminus X$
 - $Z = A(Q) \setminus X$
 - $(A:n)$ = релација са атрибутима A која има само једну торку са свим недефинисаним вредностима атрибута
- Спољашња унија представља унију релација које немају све исте атрибуте
- Алтернативна дефиниција:

$$R \dot{\cup} Q = \{ (x,y,z) \mid ((x,y) \in R \wedge z=n) \vee (x,z) \in Q \wedge y=n \}$$
 - при чему се провера припадности одвија по **не-лажно** правилима

A	B	C	D
a1	b1	c1	n
a1	b2	c2	n
a2	b3	c3	n
a2	b2	c3	n

A	B	C	D
n	b1	c1	d1
n	b2	c1	d2
n	b2	c2	d1



A	B	C	D
a1	b1	c1	n
a1	b2	c2	n
a2	b3	c3	n
a2	b2	c3	n
n	b1	c1	d1
n	b2	c1	d2
n	b2	c2	d1



Проширене скуповне операције (2)

- **Спољашњи пресек и спољашња разлика** релација R и Q са заједничким пресеком атрибута X се дефинишу на сличан начин:

$$R \cap_o Q = (R \times (Z:n)) \cap ((Y:n) \times Q)$$

$$R \setminus_o Q = (R \times (Z:n)) \setminus ((Y:n) \times Q)$$
- Ове операције немају много смисла:
 - спољашњи пресек је увек празан, зато што се пореде недефинисане вредности по *шачним* правилима
 - спољашња разлика је увек једнака првој релацији, зато што је пресек празан



Спољашње слободно спајање

- **Спољашње слободно спајање** релација R и Q са атрибутима $A(R) = X$ и $A(Q) = Y$, дефинише се као:

$$R[p]_o Q = T \cup (R \times (Y:n)) \cup (Q \times (X:n))$$
 где су:
 - $T = R[p]Q$ = резултат обичног слободног спајања по истом услову
 - $R_1 = R \setminus T[X]$ = подскуп торки из R које се не спајају са торкама из Q
 - $Q_1 = Q \setminus T[Y]$ = подскуп торки из Q које се не спајају са торкама из R
- Представља проширење слободног спајања тако да резултат увек садржи *све* торке из ових релација, чак и када не постоје одговарајуће торке са којима би се спајали



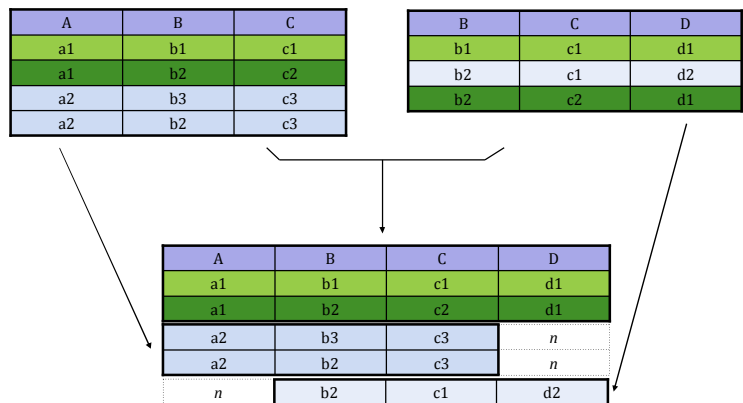
Спољашње слободно спајање (2)

- Дефинишу се додатне две специјализације спољашњег слободног спајања
 - **Лево спољашње слободно спајање** се дефинише слично, али тако да се резултату обичног слободног спајања унијом додаје само проширен подскуп торки *прве* релације који се не спајају са торкама друге релације
 - **Десно спољашње слободно спајање** се дефинише слично, али тако да се резултату обичног слободног спајања унијом додаје само проширен подскуп торки *друге* релације који се не спајају са торкама прве релације
- Већ описано **спољашње слободно спајање** се назива и **пуно спољашње спајање**



Спољашње природно спајање

- Аналогно слободном спајању се дефинишу и
 - **(Пуно) спољашње природно спајање**
 - **Лево спољашње природно спајање**
 - **Десно спољашње природно спајање**



Релациона алгебра – резиме



- Релациона алгебра као проширење скуповне алгебре:
 - скуповне операције: унија, пресек, разлика, производ
 - додатне операције: пројекција, рестрикција, природно спајање, слободно спајање, дељење
- Проширена релациона алгебра:
 - недефинисане вредности
 - проширене логичке операције
 - проширена рестрикција, проширено спајање, проширене скуповне операције
 - спољашње спајање

Релациони рачун



- Примена предикатског рачуна на релације
- Код је дефинисао две варијанте рачуна:
 - релациони рачун *n*-торки
 - релациони рачун домена
- Еквивалентан релационој алгебри

Релациони рачун *n*-торки



- **Израз** релационог рачуна *n*-торки је:

$$\{(t_1, t_2, \dots, t_k) : f\}$$
- за који важи:
 - променљиве t_1, t_2, \dots, t_k су *n*-торне **променљиве** или **индексиране *n*-торне променљиве** облика $t[i]$ где је i редни број атрибута торке t
 - скуп *n*-торних променљивих чини тзв. **циљну листу**
 - формула f је тзв. **квалификациони израз**
 - слободне променљиве у квалификационом изразу су оне и само оне које чине циљну листу



Релациони рачун н-торки (2)

- Квалификациони израз се гради од атома и операција:
 - (a1) Ако је R релација а s н-торна променљива везана за релацију R онда ознака $R(s)$ означава то везивање и зове се **атом припадности**
 - (a2) Ако су s и p н-торне променљиве, a нека константа и Θ операција поређења, онда ознаке $s[i] \Theta p[j]$ и $s[i] \Theta a$ представљају поређење i -те компоненте променљиве s са j -том компонентом променљиве p или константом a и називају се **атоми поређења**
 - (ф1) ...
 - (ф2) ...
 - (ф3) ...
 - (ф4) ...
 - (ф5) ...



Релациони рачун н-торки (3)

- Квалификациони израз се гради од атома и операција:
 - (a1) ...
 - (a2) ...
 - (ф1) сваки атом је **формула**; све н-торне променљиве атома су слободне у формули
 - (ф2) ако су f и g формуле, онда су формуле и: f AND g , f OR g , NOT f ; појављивање променљиве у формули је тачно онакво какво је у f и g
 - (ф3) ако је f формула, онда су формуле и: $(\exists s)(f)$ и $(\forall s)(f)$; променљива s је везана у тој формули а слободна у f
 - (ф4) ако је f формула, онда је формула и $(\neg f)$
 - (ф5) ништа друго није формула релационог рачуна н-торки



Примери упита

- Наредни упит издваја мушкарце који имају основу плате мању од 32000:

$$\{ x : \text{RADNIK}(x) \text{ AND } x[5]='M' \text{ AND } x[7] < 32000 \}$$
 - $x[5]$ је пол, $x[7]$ је основа плате
- Имена, презимена и датуми запослења радника који зарађују бар 40000:

$$\{(x[3], x[4], x[6]) : \text{RADNIK}(x) \text{ AND } x[7] \geq 40000 \}$$
 - $x[3]$ је име, $x[4]$ је презиме
 - $x[6]$ је датум запослења, $x[7]$ је основа плате



Релациони рачун н-торки - синтакса

- Уводи се синтакса прилагођена раду на рачунару
- Укратко:
 - уместо индексирања колона користе се имена колона иза тачке
 - уместо $x[i]$ користимо $x.kolona_i$
 - уместо записа у загради $\{ (t_1, t_2, \dots, t_k) : f \}$ користимо запис:
 - t_1, t_2, \dots, t_k WHERE f
 - у формулама се користе логички оператори AND, OR, NOT и предикати
 - EXISTS променљива (формула)
 - FORALL променљива (формула)
 - пре записа формуле везујемо имена променљивих за домене, тј. релације
 - RANGE OF променљива IS релација
 - обично подразумевамо да су променљиве облика $k, k1, k2, \dots$ везане за релацију k



Примери упита

- Наредни упит издваја мушкарце који имају основу плате мању од 32000:

$$\text{radnik WHERE radnik.pol='M'}$$

$$\text{AND radnik.osnova_plate < 32000}$$
- Имена, презимена и датуми запослења радника који зарађују бар 40000:

$$\text{radnik.ime, radnik.prezime, radnik.dat_zaposlenja}$$

$$\text{WHERE radnik.osnova_plate } \geq 40000$$



Релациони рачун домена

- Личи на рачун н-торки, али се променљиве везују за домене атрибута а не за домене торки
- **Израз** релационог рачуна домена је:

$$\{x_1, x_2, \dots, x_k \mid f\}$$

- за који важи:
 - променљиве x_1, x_2, \dots, x_k су **доменске променљиве**
 - f је **формула** релационог рачуна домена чије су слободне променљиве x_1, x_2, \dots, x_k
 - формуле се дефинишу слично као у рел.рачуну н-торки
 - измењен је облик атома



Релациони рачун домена (2)

- Атоми израза релационог рачуна домена су:
 - (a1') Ако је R релација степена n (има n атрибута) и y_1, y_2, \dots, y_n константе или доменске променљиве над доменима одговарајућих атрибута релације R , онда је $R(y_1, y_2, \dots, y_n)$ атом чије је значење да вредности y_1, y_2, \dots, y_n морају бити такве да постоји одговарајућа н-торка релације R
 - (a2') Ако су x и y н-торне променљиве, a нека константа и Θ операција поређења, онда су $x \Theta y$ и $x \Theta a$ **атомима поређења** и значење им је да вредности x и y морају да буде такве да је поређење тачно



Примери упита

- Наредни упит издваја мушкарце који имају основу плате мању од 32000:

$$\{xyzq'M'st \mid \text{RADNIK}(xyzq'M'st) \text{ AND } t < 32000 \}$$
 - t је основа плате
- Имена, презимена и датуми запослења радника који зарађују бар 40000:

$$\{zqs \mid (\exists x)(\exists y)(\exists w)(\exists t) \text{ RADNIK}(xyzqwst) \text{ AND } t \geq 40000 \}$$
 - z је име, q је презиме, s је датум запослења
 - t је основа плате



Релациони рачун – резиме

- Оба представљена релациона рачуна почивају на предикатском рачуну првог реда
 - Изрази
 - Атоми
 - Формуле
- Релациони рачун n-торки у облику који је прилагођен употреби на рачунару има сличности са SQL-ом



Примери упита 1.1

- Имена и презимена запослених у “Планирању”
- Релациона алгебра:


```
( ORG_JEDINICA [naziv = 'Planiranje']
  * RADNIK
  )[ime, prezime]
```



Примери упита 1.2

- Имена и презимена запослених у “Планирању”, чији стаж улази у категорију “Средња”
- Релациони рачун n-торки:


```
radnik1.ime, radnik1.prezime
WHERE EXISTS org_jedinica1 (
  radnik1.org_jed_id = org_jedinica1.org_jed_id
  AND org_jedinica1.naziv = 'Planiranje'
)
```



Примери упита 1.3

- Имена и презимена запослених у “Планирању”, чији стаж улази у категорију “Средња”
- Релациони рачун домена:

$$\{ ip \mid (\exists a)(\exists b)(\exists c)(\exists d)(\exists e) \text{ RADNIK}(abipcde) \wedge (\exists f) \text{ ORG_JEDINICA}(bf\text{'Planiranje'}) \}$$



Примери упита 2.1

- Имена и презимена запослених у “Планирању” који су били ангажовани на бар два различита пројекта
- Релациона алгебра:

```
(
  RADNIK * ORG_JEDINICA [naziv = 'Planiranje']
  [radnik_id, ime, prezime]
) [radnik_id * rad_id] (
  (RESENJE1 [rad_id] RESENJE2 )
  [resenje1_proj_id <> resenje2_proj_id]
  [rad_id]
) [ime, prezime]
```



Примери упита 2.2

- Имена и презимена запослених у “Планирању”, чији стаж улази у категорију “Средња”
- Релациони рачун н-торки:


```
radnik1.ime, radnik1.prezime
WHERE EXISTS org_jedinica1 (
  radnik1.org_jed_id = org_jedinica1.org_jed_id
  AND org_jedinica1.naziv = 'Planiranje'
) AND EXISTS resenje1 (
  radnik1.radnik_id = resenje1.rad_id
  AND EXISTS resenje2(
    radnik1.radnik_id = resenje2.rad_id
    AND resenje1.proj_id <> resenje2.proj_id
  )
)
```



Примери упита 2.3

- Имена и презимена запослених у “Планирању”, чији стаж улази у категорију “Средња”
- Релациони рачун домена:

$$\{ ip \mid (\exists a)(\exists b)(\exists c)(\exists d)(\exists e) \text{ RADNIK}(abipcde) \wedge (\exists f) \text{ ORG_JEDINICA}(bf\text{Planiranje}') \wedge (\exists g)(\exists h)(\exists i)(\exists j) \text{ RESENJE}(aghij) \wedge (\exists k)(\exists l)(\exists m)(\exists n) \text{ RESENJE}(aklmn) \wedge g \neq k \}$$


Значај формалних упитних језика

- Из угла теорије, кључно је да се утврди да је упитни језик довољно добар да се њиме може исказати било који упит
- У пракси је потребна и ефикасност
- Оптимизација упита обухвата послове анализе упита и проналажења најефикаснијег пута за његово израчунавање
- Релациони упитни језици омогућавају релативно једноставно аутоматско модификовање, па тиме и оптимизацију, имајући у виду формалну заснованост

* Оптимизација базе података обухвата послове анализе стварне или процењене употребе базе података и пројектовања физичког модела базе података у циљу омогућавања свеукупно ефикаснијег рада система



Ажурирање базе података

- Поред упита, важно место заузима ажурирање базе података
- Има више приступа, али суштина је да захтева проширења и нове појмове или промене дефиниција постојећих појмова:
 - *релациона променљива*
= променљива релације
 - *релација*
= вредност релације
= садржај релационе променљиве
 - *променљива базе података*
= низ променљивих релација
 - *база података*
= вредност базе података
= садржај променљиве базе података



Ажурирање базе података

- *Ажурирање базе података* се дефинише као замењивање вредности променљиве базе података новом вредношћу базе података
- Променљивост података базе података се на тај начин апстражује секвенцом вредности променљиве базе података које она добија у различитим временским тачкама (тренуцима)
- Таква дефиниција омогућава и разматрање сагласних конкурентних промена
- Штавише, омогућава и дефинисање чисто функционалних објектно оријентисаних упитних језика
- Нећемо детаљније разматрати
- Видети, на пример: *Darwen, Date, The Third Manifesto, 1995.*



Манипулативни део модела – резиме

- Формални упитни језици
 - релациона алгебра
 - релациони рачун n-торки
 - релациони рачун домена
- Релациона алгебра има операциону семантику
 - описује се *како* се рачуна резултат
- Рачун има декларативну семантику
 - описује се *шта* је потребно израчунати а не *како*

Литература за тему

- Гордана Павловић-Лажетић, *Увод у релационе базе података, 2. изд. Математички факултет, 1999.*
 - доступно онлајн: <http://poincare.matf.bg.ac.rs/~gordana/urbp-2016.htm>
- Ramakrishnan, Gehrke, *Database Management Systems, 2.ed, 2000.*
- Codd, *A relational model of data for large shared data banks, Comm.ACM, 13(6), 1970.*
- Codd, *Extending the database relational model to capture more meaning, ACM ToDS, 4(4), 1979.*
- Codd *The Relational Model for Database Management – Version 2, AddisonWesley Publ. Inc., 1990.*
- Darwen, Date, *The Third Manifesto, 1995.*
- IBM, *Database Administration Concepts and Configuration Reference, 2012.*