

Пројектовање база података - Пример испита

16. avgust 2021.

Нормализација

Захтев

Нека је дата база података за потребе евиденције књига, њихових писаца и њихових издавача.

1. За сваког писца се бележи шифра писца (п_шиф), име, презиме и држава из које потиче. Додатно се за сваког писца бележи и број наслова тј. број написаних књига.
2. За сваку књигу (коју је написао неки писац) се евидентира шифра књиге (к_шиф), наслов и област којој припада.
3. Књига може бити издавана од стране неког издавача за којег се евидентира шифра издавача (и_шиф), назив, статус и држава.
4. Писац може да напише више верзија једне књиге, где сваког верзији одговара неки редни број. За сваки редни број књиге се прати година када је написана, шифра издавача и издање (број издања).
5. За свако издање (може их бити више) неке књиге од стране неког издавача се бележи година издања и тираж издања.

Формулисати и именовати све зависности међу уведеним атрибутима. Пројектовати шему релација које су у БКНФ и које приказују претходно описане податке.

Решење

Поступак: Извући све атрибуте из текста у једну релацију и на основу текста записати све релације. Даље, довести базу података у БКНФ.

Функционалне зависности из текста:

1. {п_шиф} \longrightarrow {име, презиме, држава_писца, број_наслова}
2. {к_шиф} \longrightarrow {наслов_књиге, област}
3. {и_шиф} \longrightarrow {назив_издавача, статус, држава_издавача}
4. {п_шиф, к_шиф, редни_број} \longrightarrow {година, и_шиф, издање}
5. {и_шиф, к_шиф, издање} \longrightarrow {година_издања, тираж}

Напомена: За неке атрибуте су додати суфикси да би се разликовали при спајању у једну релацију (пример: писац и издавач имају атрибут држава).

Почетна релација: $R = \{п_шиф, име, презиме, држава_писца, број_наслова, к_шиф, наслов_књиге, област, и_шиф, назив_издавача, статус, држава_издавача, редни_број, година, издање, година_издања, тираж\}$. Кандидат за кључ је {п_шиф, к_шиф, редни_број}, јер се ти атрибути појављују само са леве стране, па морају да буду део кључа, и функционално одређују све остале атрибуте. Релација није у 2НФ и то нарушава свака од горе наведених ФЗ сем Ф34 (са леве стране је цео кључ, а не подскуп) и Ф35 (није подскуп кључа, јер има неке неключне атрибуте).

Декомпозиција релације **R** по **Ф31**:

- $R = \{п_шиф, к_шиф, редни_број, издање, наслов_књиге, област, и_шиф, назив_издавача, статус, држава_издавача, година, година_издања, тираж\}$ (важи: Ф32, Ф33, Ф34, Ф35)
- **Писац** = {п_шиф, име, презиме, држава_писца, број_наслова} (важи: Ф31)

Декомпозиција релације **R** по **Ф32**:

- $R = \{п_шиф, к_шиф, редни_број, издање, и_шиф, назив_издавача, статус, држава_издавача, година, година_издања, тираж\}$ (важи: Ф33, Ф34, Ф35)
- **Писац** = {п_шиф, име, презиме, држава_писца, број_наслова} (важи: Ф31)
- **Књига** = {к_шиф, наслов_књиге, област} (важи: Ф32)

Декомпозиција релације **R** по **Ф33**:

- **Р** = {п_шиф, к_шиф, редни_број, издање, и_шиф, година, година_издања, тираж} (важи: Ф34, Ф35)
- **Писац** = {п_шиф, име, презиме, држава_писца, број_наслова} (важи: Ф31)
- **Књига** = {к_шиф, наслов_књиге, област} (важи: Ф32)
- **Издавач** = {и_шиф, назив_издавача, статус, држава_издавача} (важи: Ф33)

Све релације су сада у 2НФ. Релација **Р** није у 3НФ због **Ф35**. У то можемо да се уверимо тако што проверимо да ли важи неки од услова за испуњавање 3НФ за Ф35 тј.

{и_шиф, к_шиф, издање} \rightarrow {година_издања, тираж}:

- Да ли је Ф3 тривијална? Не.
- Да ли је са леве стране наткључ? Не.
- Да ли је десна страна део неког кључа? Не, Ф35 нарушава 3НФ!

Остале релације чак су чак и у БКНФ. Декомпозиција релације **Р** по **Ф35**:

- **Р** = {п_шиф, к_шиф, редни_број, и_шиф, издање, година} (важи: Ф34)
- **Писац** = {п_шиф, име, презиме, држава_писца, број_наслова} (важи: Ф31)
- **Књига** = {к_шиф, наслов_књиге, област} (важи: Ф32)
- **Издавач** = {и_шиф, назив_издавача, статус, држава_издавача} (важи: Ф33)
- **Издање_књиге** = {и_шиф, к_шиф, издање, година_издања, тираж} (важи: Ф35)

Како нема функционалних зависности неключног атрибута у неключни атрибут, релација је у 3НФ. Како нема додатних функционалних зависности неключних атрибута у кључни атрибут, релација је и у БКНФ. Релација **Р** се може преименовати у **Књига_писца**.

Индекси

Захтев

Нека су дате релације:

Биоскоп (id_bioskopa, naziv, lokacija)

Пројекција (id_bioskopa, id_filma, datum_projekcije, vreme_projekcije, broj_posetilaca, cena_karte, tip_filma)

Филм (id_filma, naziv_filma, zanzr)

Претпоставимо да су следећи упити најважнији у радном оптерећењу, као и да су од једнаком значаја и приближне учесталости:

1. Излистати назив биоскопа, датум и време пројекције за све пројекције филмова који су жанра који задаје корисник и који се приказују у датом опсегу датума.
2. Издвојити ид-еве филмова који су на репертоару текућег датума, и који су типа који задаје корисник. Тип филма може бити 2Д или 3Д.
3. Издвојити назив филма са највећим укупним бројем посетилаца на свим његовим пројекцијама.
4. Издвојити називе биоскопа у којима је постојала пројекција таква да је број посетилаца био већи од вредности коју задаје корисник.

Претпоставимо да су наведени упити чешћи од ажурирања, стога треба направити све потребне индексе, али треба водити рачуна на се не додају непотребни индекси. Одредити које атрибуте треба индексирати, за сваки индекс навести да ли треба да буде кластерован или не и да ли индекс треба бити хеш или Б+ стабло. Претпоставити да су подржани и индекси са више атрибута. Обавезно је сваку одлуку детаљно образложити.

Решење

Прво желимо да преведемо упите у *sql* облик:

Упит 1 ('X', 'Y' i 'Z' су променљиве коју уноси корисник)

```
select b.naziv, p.datum_projekcije, p.vreme_projekcije
from bioskop as b join projekcije as p
  on b.id_bioskopa = p.id_bioskopa
where p.datum_projekcije in between(X, Y)
and p.id_filma in (
  select f.id_filma
  from film as f
  where f.zanr = Z
)
```

Закључак (независно од осталих упита):

- Имамо интервални упит за *p.datum_projekcije in between(X, Y)*. Можемо да поставимо кластеровани Б+ стабло индекс над *p.datum_projekcije*.
- Идеално је да биоскоп буде у унутрашњој петљи при спајању, јер се не користи у селекцији. То је мотивација да се дода некластерован хеш индекс над *p.id_bioskopa*.
- Због угњевженог упита, можемо да тестирамо *index-only plan* над релацијом за филм преко композитног некласерованог хеш индекса *<f.zanr, f.id_filma>*

Упит 2 ('X' је променљива коју уноси корисник)

```
select f.id_filma
from film as f join projekcije as p
  on f.id_filma = p.id_filma
where p.datum_projekcije = now()
and p.tip_filma = X
```

Закључак (независно од осталих упита):

- Идеално је да филм буде у унутрашњој петљи при спајању, јер се не користи у селекцији. То је мотивација да се дода некластерован хеш индекс над *p.id_filma*.
- Кандидати за индексе за релацију пројекције су:
 - *<p.datum_projekcije>*,
 - *<p.tip_filma>*,
 - *<p.tip_filma, p.datum_projekcije>* и
 - *<p.datum_projekcije, p.tip_filma>*.

Композитни индекс *<f.datum_projekcije, f.tip_filma>* највероватније има бољу селективност од *<f.datum_projekcije, f.tip_filma>*, јер очекујемо да мали проценат досадашњих пројекција одговара данашњем дану. Из истог разлога индекс *<f.datum_projekcije>* делује као добар избор у односу на остале.

Упит 3

```
with (
  select p.id_filma, sum(p.broj_posetilaca) as suma
  from projekcije as p
  group by p.id_filma
  order by suma desc
  limit 1
) as najpopularniji_film
select f.naziv
from film as f
where f.id_filma = (
  select nf.id_filma
  from najpopularniji_film as nf
)
```

Закључак (независно од осталих упита):

- Ако посматрамо помоћни упит, некластерован композитни хеш индекс $\langle p.id_filma, p.broj_posetilaca \rangle$ може да нам омогући боље перформансе.

Упит 4 ('X' је променљива коју уноси корисник)

```
select b.naziv
from bioskop as b
where id_bioskopa in (
  select p.id_bioskopa
  from projekcije as p
  where p.broj_posetilaca > X
)
```

Закључак (независно од осталих упита):

- Због интервалног упита, кластеровани Б+ стабло индекс је добар кандидат на атрибуту $p.broj_posetilaca$.
- Алтернатива је да са постави композитни некластеровани Б+ стабло индекс над $\langle p.broj_posetilaca \rangle, p.id_bioskopa \rangle$.

Коначно:

- Над релацијом за пројекције за кластеровани индекс имамо два кандидата $p.broj_posetilaca$ и $p.datum_projekcije$. Треба да се тестирају обе опције. Индекс над $p.datum_projekcije$ нам значи за први и други упит, а $p.broj_posetilaca$ само за четврти упит.
- Што се тиче некластерованих индекса, најбоље да се тестирају сви предлози и направе индекси ако значајно побољшавају перформансе. Ако су перформансе сличне, онда је боље да се не праве индекси.

Превођење ЕЕР у релациони модел базе података

Захтев

Превести дати ЕЕР дијаграм у релациони модел података. Креирати скупове схема и међурелационих ограничења.

Решење

Први корак: Превођење ентитета (плавим текстом су обележени страни кључеви додати у трећем кораку)

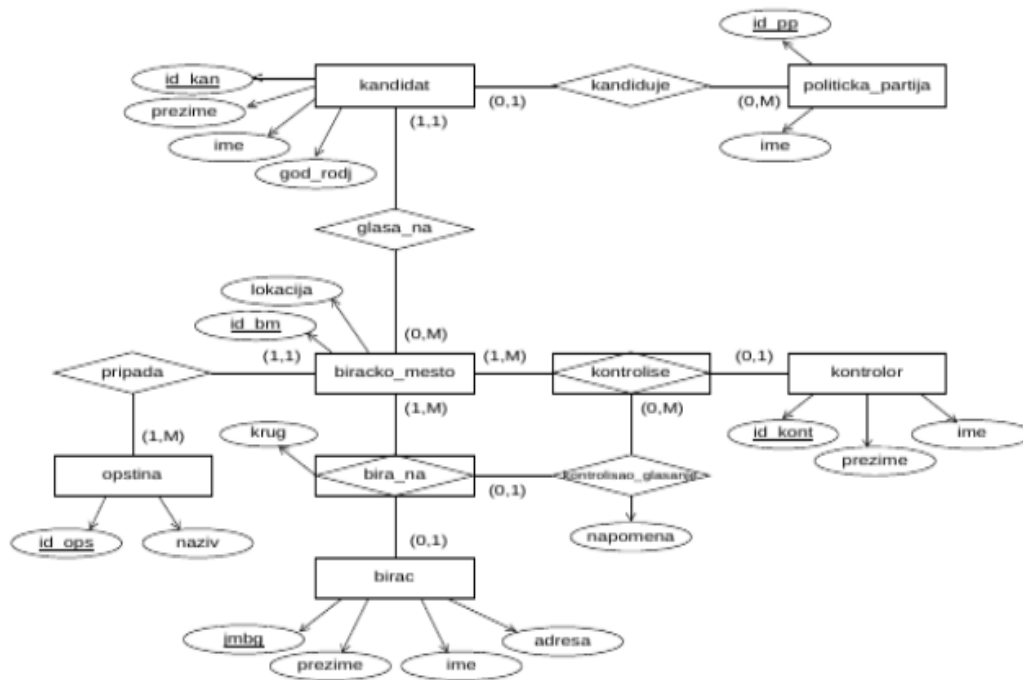
- кандидат(ид_кан, презиме, име, год_рођ, ид_бм)
- политичка_партија(ид_пп, име)
- бирачко_место(ид_бм, локација, ид_опш)
- општина(ид_опш, назив)
- бирач(ЈМБГ, презиме, име, адреса)
- контролор(ид_конт, презиме, име)

Други Корак: Превођење зависних ентитета и специјализација

- Нема зависних ентитета и специјализација.

Трећи корак: Превођење односа

- кандидује:
 - Однос се преводи у табелу.
 - кандидује(ид_кан, ид_пп)



- $\text{kandiduje}[\text{id_kan}] \subseteq \text{kandidat}[\text{id_kan}]$
- $\text{kandiduje}[\text{id_pp}] \subseteq \text{politicka_partija}[\text{id_kan}]$

• **гласа_на:**

- Додаје се примарни кључ **ид_бм** табеле из **бирачко_место** као страни кључ у табелу **кандидат**.
- $\text{kandidat}[\text{id_бм}] \subseteq \text{бирачко_место}[\text{id_бм}]$

• **припада:**

- Додаје се примарни кључ **ид_опш** табеле из **општина** као страни кључ у табелу **бирачко_место**.
- $\text{бирачко_место}[\text{id_опш}] \subseteq \text{општина}[\text{id_опш}]$

• **бира_на:**

- Однос се преводи у табелу.
- **бира_на**(ЈМБГ, ид_бм, круг)
- $\text{бира_на}[\text{ЈМБГ}] \subseteq \text{бирач}[\text{ЈМБГ}]$
- $\text{бира_на}[\text{id_бм}] \subseteq \text{бирачко_место}[\text{id_бм}]$

• **контролише:**

- Однос се преводи у табелу.
- **контролише**(ид_конт, ид_бм)
- $\text{контролише}[\text{id_конт}] \subseteq \text{контролор}[\text{id_конт}]$
- $\text{контролише}[\text{id_бм}] \subseteq \text{бирачко_место}[\text{id_бм}]$

• **контролисао_гласање:**

- Однос се преводи у табелу.
- **контролисао_гласање**(ЈМБГ, ид_конт, напомена)
- $\text{контролисао_гласање}[\text{ЈМБГ}] \subseteq \text{бира_на}[\text{ЈМБГ}]$
- $\text{контролисао_гласање}[\text{id_конт}] \subseteq \text{контролише}[\text{id_конт}]$