

Adatbázisok 1.

Relációs algebra lekérdezések optimalizációja

Algebrai lekérdezés

- Lekérdezés példa: Sziszi SZ01 azonosítójú számláján mennyi pénz van?

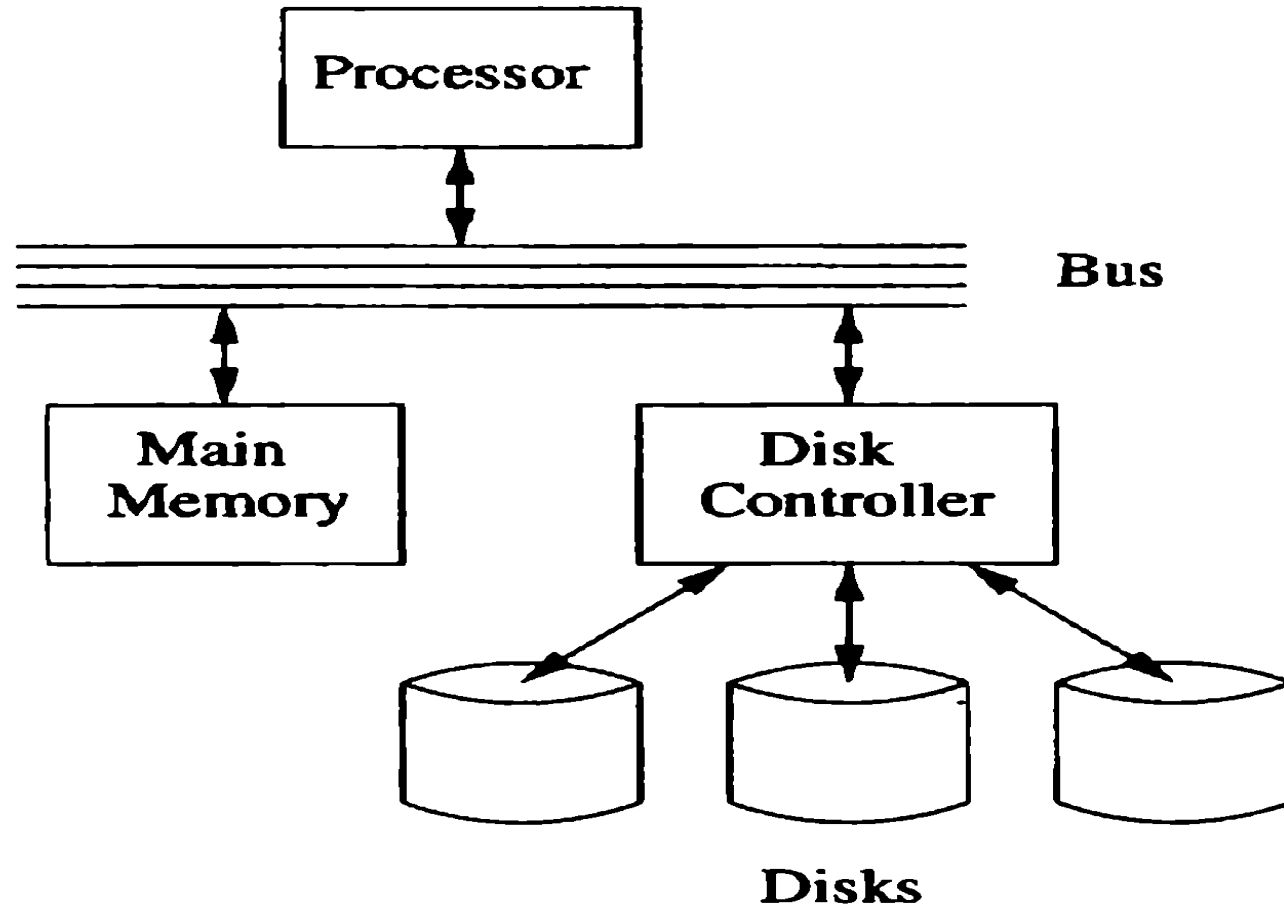
Számla

név	számla_azon	összeg
Sziszi	SZ01	45000
Peti	SZ02	543000
Sziszi	SZ03	120000

- $\Pi_{\text{összeg}} (\sigma_{\text{név}='Sziszi' \wedge \text{számla_azon}='SZ01'}(\text{Számla}))$

összeg
45000

Számítógép rendszer sematikus ábrája



Mi a cél?

- **Moore-törvénye:** (Gordon Moore) szerint az integrált áramkörök sok jellemzőjének fejlődése exponenciális, ezek az értékek 18 havonta duplázódnak. Ilyenek például:
 - (i) processzorok sebességének és árának aránya,
 - (ii) lemez egy bitre eső ára és a lemezen tárolható bájtok száma.
- Más paraméterek azonban sokkal lassabban fejlődnek. Ilyenek például:
 - (i) központi memóriában milyen gyorsan lehet az adatokat elérni,
 - (ii) az a sebesség, amellyel a lemez mozog a merevlemezben.
- Emiatt egy-egy nagy adathalmazmal dolgozó algoritmus optimalizációjánál az a lényeges szempont, hogy a feladatot minél kevesebb adatmozgatással tudjunk megoldani a háttértároló és a központi memória között.

Egy lehetséges megközelítés

- Minél kevesebb **lemez olvasási és írási** (I/O) műveletet végrehajtani egy-egy lekérdezés végrehajtása során
- → **igyekszünk minél kisebb méretű relációkkal dolgozni**
- Az optimalizáció során **relációs algebrai azonosság** alkalmazása
- Eredetivel ekvivalens lekérdezés készítése, kiszámításához az esetek többségében kevesebb I/O művelet
- A q , q' relációs algebrai lekérdezések **ekvivalensek**, ha tetszőleges I előfordulás esetén $q(I) = q'(I)$ fennáll.
Jelben: **$q \equiv q'$** .

Egy példa...

- A táblák legyenek:

Film (cím, év, hossz)

Szerepel (filmcím, év, színésznév)

- Ekkor a következő lekérdezés:

Film (F)

Cím	Év	Hossz
Alma	2010	120
Körte	2012	65
Barack	2016	85

Szerepel (Sz)

Filmcím	Év	Színésznév
Alma	2010	Edus
Körte	2012	Gianni
Barack	2016	Edus

$\Pi_{\text{hossz}}(\sigma_{\text{cím=filmcím} \wedge \text{F.év}=\text{Sz.év} \wedge \text{színésznév}=\text{'Edus'}}(\text{F} \times \text{Sz}))$ ekvivalens a

$\Pi_{\text{hossz}}(\sigma_{\text{cím=filmcím} \wedge \text{F.év}=\text{Sz.év}}(\text{F} \times (\sigma_{\text{színésznév}=\text{'Edus'}}(\text{Sz}))))$ lekérdezéssel.

- Emellett az utóbbi valószínűleg gyorsabban végrehajtható.

Illusztráció

$$\sigma_{\text{cím=filmcím} \wedge \text{F.év}=\text{Sz.év} \wedge \text{színésznév}=\text{'Edus'}} (\mathbf{F} \times \mathbf{Sz})$$

Cím	F.Év	Hossz	Filmcím	Sz.év	Színésznév
Alma	2010	120	Alma	2010	Edus ✓
Alma	2010	120	Körte	2012	Gianni ✗
Alma	2010	120	Barack	2016	Edus ✓
Körte	2012	65	Alma	2010	Edus ✓
Körte	2012	65	Körte	2012	Gianni ✗
Körte	2012	65	Barack	2016	Edus ✓
Barack	2016	85	Alma	2010	Edus ✓
Barack	2016	85	Körte	2012	Gianni ✗
Barack	2016	85	Barack	2016	Edus ✓

$$\sigma_{\text{cím=filmcím} \wedge \text{F.év}=\text{Sz.év}} (\mathbf{F} \times (\sigma_{\text{színésznév}=\text{'Edus'}} (\mathbf{Sz})))$$

Cím	F.Év	Hossz	Filmcím	Sz.év	Színésznév
Alma	2010	120	Alma	2010	Edus ✓
Alma	2010	120	Barack	2016	Edus ✓
Körte	2012	65	Alma	2010	Edus ✓
Körte	2012	65	Barack	2016	Edus ✓
Barack	2016	85	Alma	2010	Edus ✓
Barack	2016	85	Barack	2016	Edus ✓

Descartes-szorzat és összekapcsolások

- Asszociativitás:

$(E_1 \Delta E_2) \Delta E_3 \equiv E_1 \Delta (E_2 \Delta E_3)$, ahol $\Delta \in \{\times, |X|\}$ és

[természetes összekapcsolás]

$(E_1 |X|_{F1} E_2) |X|_{F2} E_3 \equiv E_1 |X|_{F1} (E_2 |X|_{F2} E_3)$, ha

$\text{attr}(F1) \subseteq \text{attr}(E1) \cup \text{attr}(E2)$ és $\text{attr}(F2) \subseteq \text{attr}(E2) \cup \text{attr}(E3)$

- [θ összekapcsolás]

- Kommutativitás:

$E_1 \Delta E_2 \equiv E_2 \Delta E_1$, ahol $\Delta \in \{\times, |X|, |X|_F\}$.

Projekció és szelekció

- Projekció sorozat:

$$\Pi_X(\Pi_Y(E)) \equiv \Pi_X(E), \text{ ha } X \subseteq Y.$$

- Kiválasztás és a feltételek konjunkciója:

$$\sigma_{F_1 \wedge F_2}(E) \equiv \sigma_{F_1}(\sigma_{F_2}(E)).$$

- Kiválasztás és a feltételek diszjunkciója:

$$\sigma_{F_1 \vee F_2}(E) \equiv \sigma_{F_1}(E) \cup \sigma_{F_2}(E).$$

- Kiválasztás elé projekció beillesztése:

$$\Pi_X(\sigma_F(E)) \equiv \Pi_X(\sigma_F(\Pi_Y(E))), \text{ ahol } Y = \text{attr}(F) \cup X.$$

Kiválasztás és Descartes-szorzat/összekapcsolás

- Kiválasztás és Descartes-szorzat, összekapcsolás felcserélése:

$\sigma_F (E_1 \Delta E_2) \equiv \sigma_F (E_1) \Delta E_2$, ahol $\text{attr}(F) \subseteq \text{attr}(E_1)$ és $\Delta \in \{\times, |X|\}$.

- Általánosabban:

$\sigma_F (E_1 \Delta E_2) \equiv \sigma_{F_1} (E_1) \Delta \sigma_{F_2} (E_2)$, ahol $\text{attr}(F_i) \subseteq \text{attr}(E_i)$ ($i = (1, 2)$)
 $F = F_1 \wedge F_2$ és $\Delta \in \{\times, |X|\}$.

- Ezekből levezethető:

$\sigma_F (E_1 \Delta E_2) \equiv \sigma_{F_2} (\sigma_{F_1} (E_1) \Delta E_2)$, ahol $\text{attr}(F_1) \subseteq \text{attr}(E_1)$, $F = F_1 \wedge F_2$,
de $\text{attr}(F_2) \subseteq \text{attr}(E_i)$ nem teljesül ($i = (1, 2)$), $\Delta \in \{\times, |X|\}$.

Projekció és Descartes-szorzat/összekapcsolás

- Projekció és Descartes-szorzat, összekapcsolás felcserélése:

$$\Pi_X(E_1 \Delta E_2) \equiv \Pi_Y(E_1) \Delta \Pi_Z(E_2),$$

ahol $X = Y \cup Z$, $Y \subseteq \text{attr}(E_1)$, $Z \subseteq \text{attr}(E_2)$ és $\Delta \in \{\times, \bowtie\}$.

Projekció/kiválasztás és halmazműveletek

- Kiválasztás és unió (metszet, különbség) felcserélése:

$$\sigma_F (E_1 \Delta E_2) \equiv \sigma_F (E_1) \Delta \sigma_F (E_2), \text{ ahol } \Delta \in \{\cap, \cup, -\}.$$

- Projekció unióval való felcserélése:

$$\pi_X(E_1 \cup E_2) \equiv \pi_X(E_1) \cup \pi_X(E_2).$$

- **Megjegyzés:** nincs általános szabály a projekció különbséggel való felcserélésére.
- **Kérdés:** a metszettel mi a helyzet? [[reláció séma](#)]

Projekció/kiválasztás és halmazműveletek

- Nincs általános szabály a projekció metszettel való felcserélésére sem.
Pl. amikor nem működik:

$$\begin{aligned}\pi_A(\{\langle A = a, B = b \rangle\} \cap \{\langle A = a, B = b' \rangle\}) &= \emptyset \\ \pi_A(\{\langle A = a, B = b \rangle\}) \cap \pi_A(\{\langle A = a, B = b' \rangle\}) &= \{\langle A = a \rangle\}\end{aligned}$$

Példa optimalizálásra

- A következő két feladathoz használt táblák:

Személy (név, kor, város, ISBN)

Könyv (cím, író, ISBN, ár)

Kiad (k_cím, k_író, város, ország)

- Kik azok, akik 20 évesek, és moszkvai kiadású könyvet kölcsönöztek ki?

Példa optimalizálásra

- A következő két feladathoz használt táblák:

Személy (név, kor, város, ISBN)

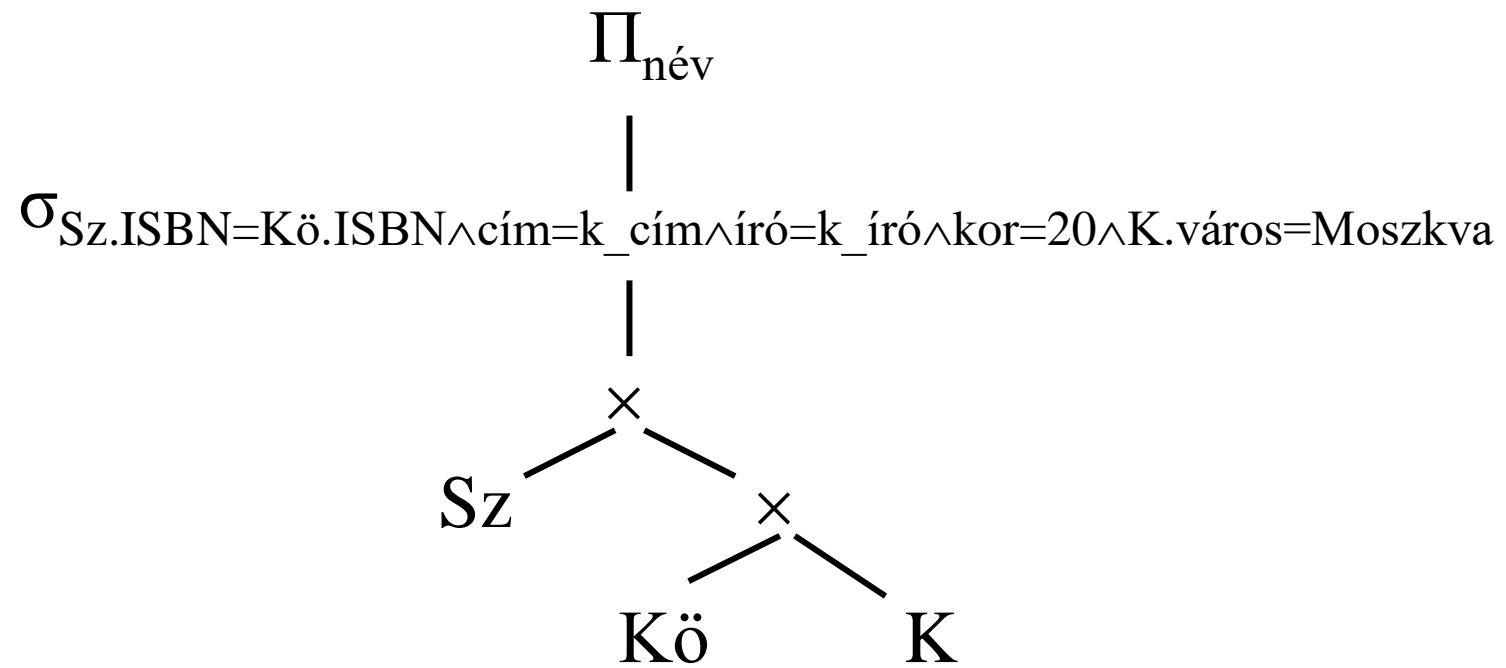
Könyv (cím, író, ISBN, ár)

Kiad (k_cím, k_író, város, ország)

- Kik azok, akik 20 évesek, és moszkvai kiadású könyvet kölcsönöztek ki?

$$\Pi_N(\sigma_{S_z.ISBN=Kö.ISBN \wedge cím=k_cím \wedge író=k_író \wedge kor=20 \wedge K.város=Moszkva} (S_z \times K_ö \times K))$$

Lekérdezésfa

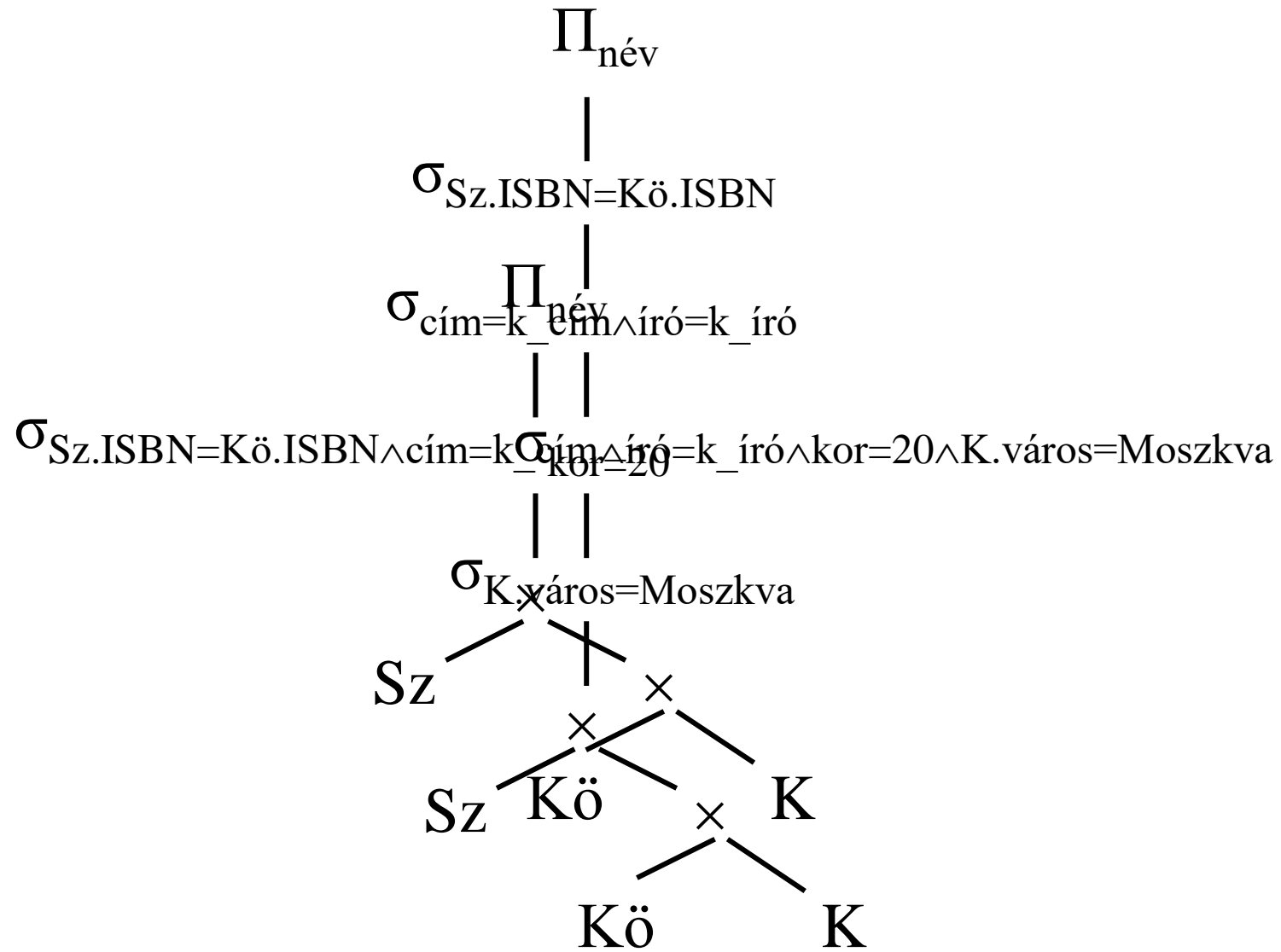


Kiválasztások "lejjebb csúsztatása"

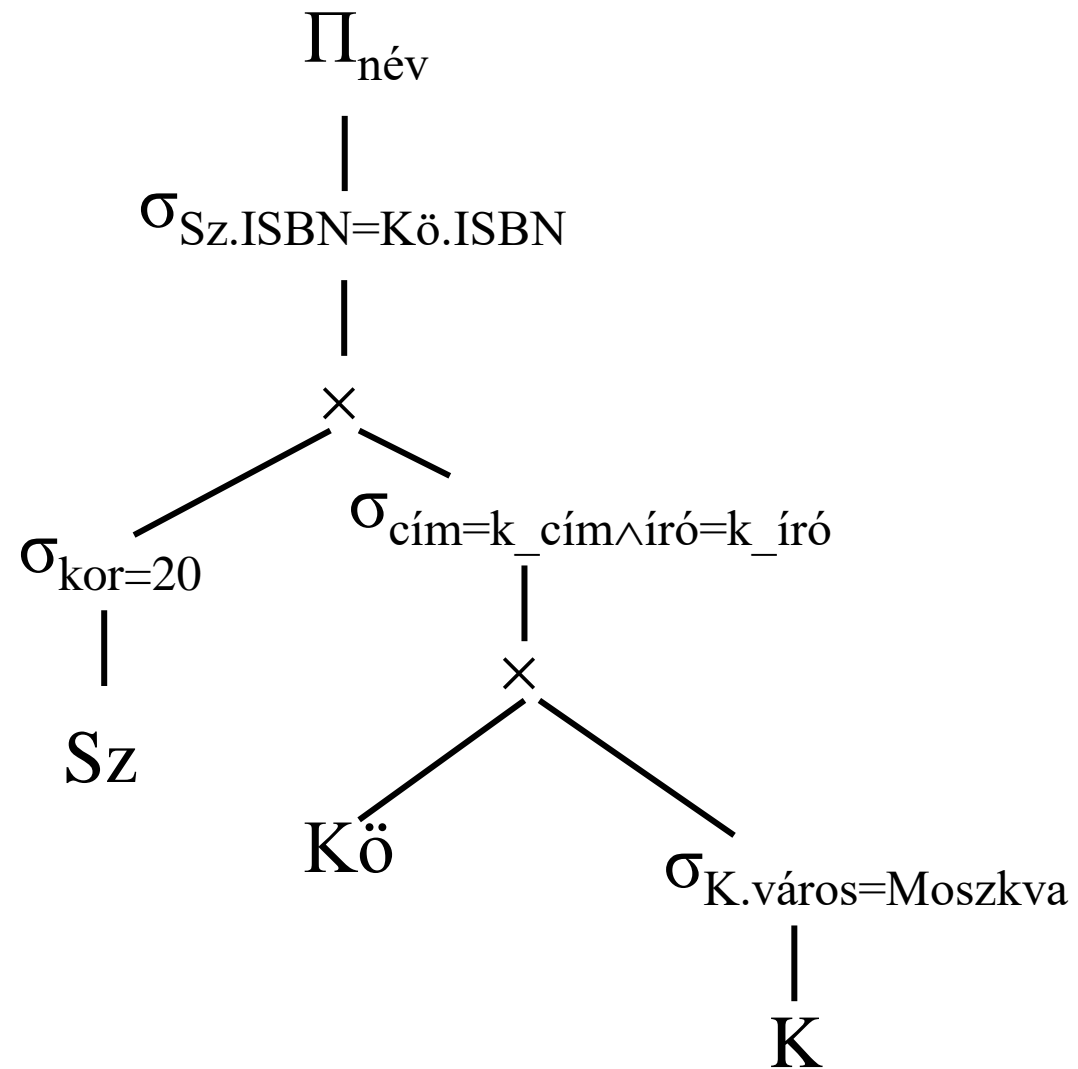
1. A kiválasztások **konjunkciós feltételeinek szétदारabolása elemi feltételekké** a $\sigma_{F_1 \wedge F_2}(E) \equiv \sigma_{F_1}(\sigma_{F_2}(E))$ szabály segítségével.
2. A **kiválasztás halmazműveletekkel illetve Descartes-szorozattal és a természetes összekapcsolással** való felcserélésének szabályainak alkalmazása.
 - Cél: a kiválasztásokat minél hamarabb végrehajtani
 - A Théta-összekapcsolást itt jobb, ha egy Descartes-szorozatra és egy azt követő kiválasztásra bontjuk.

$$R \bowtie_X S \equiv \sigma_F(R \times S).$$

Darabolás

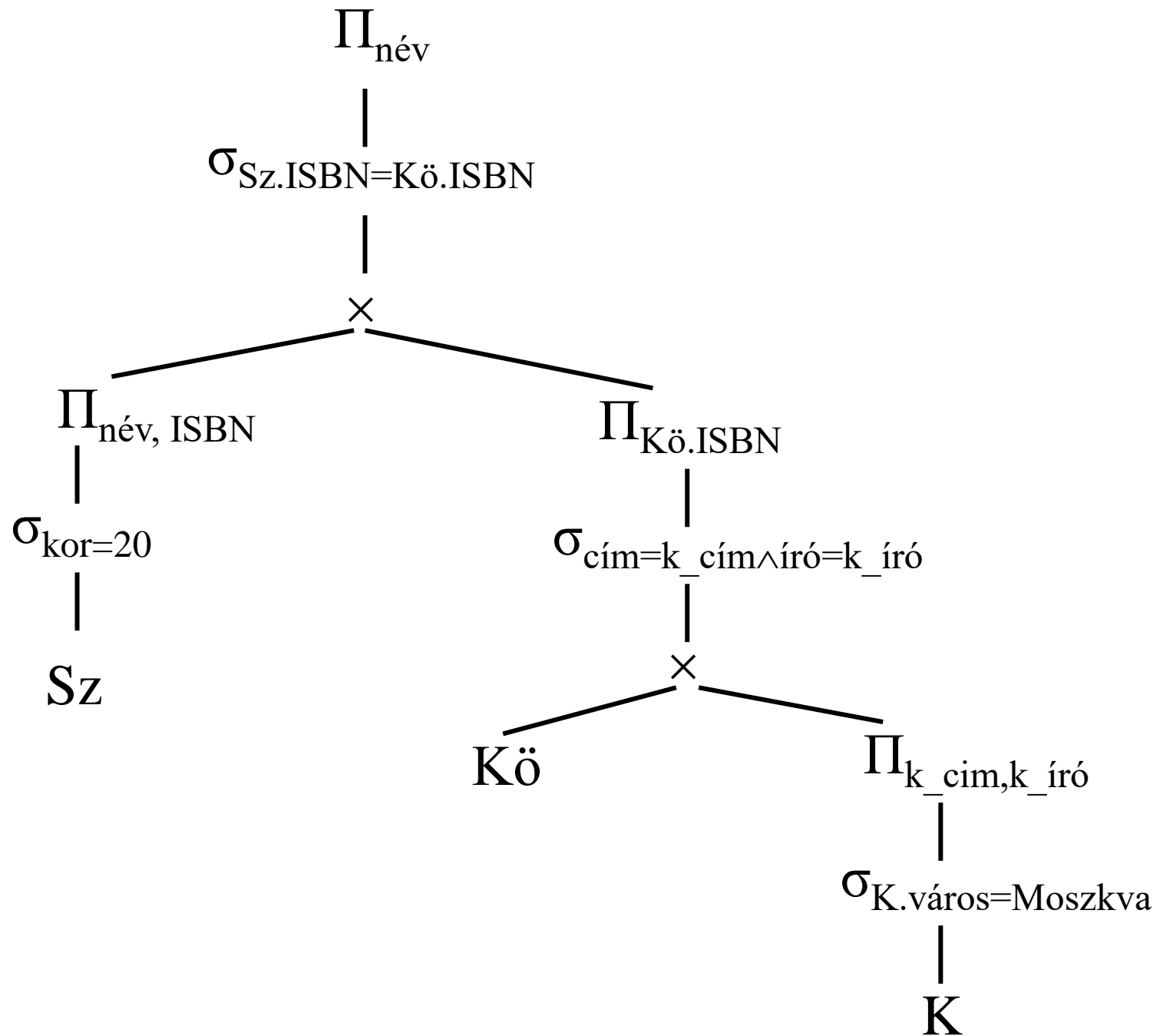


Letolás



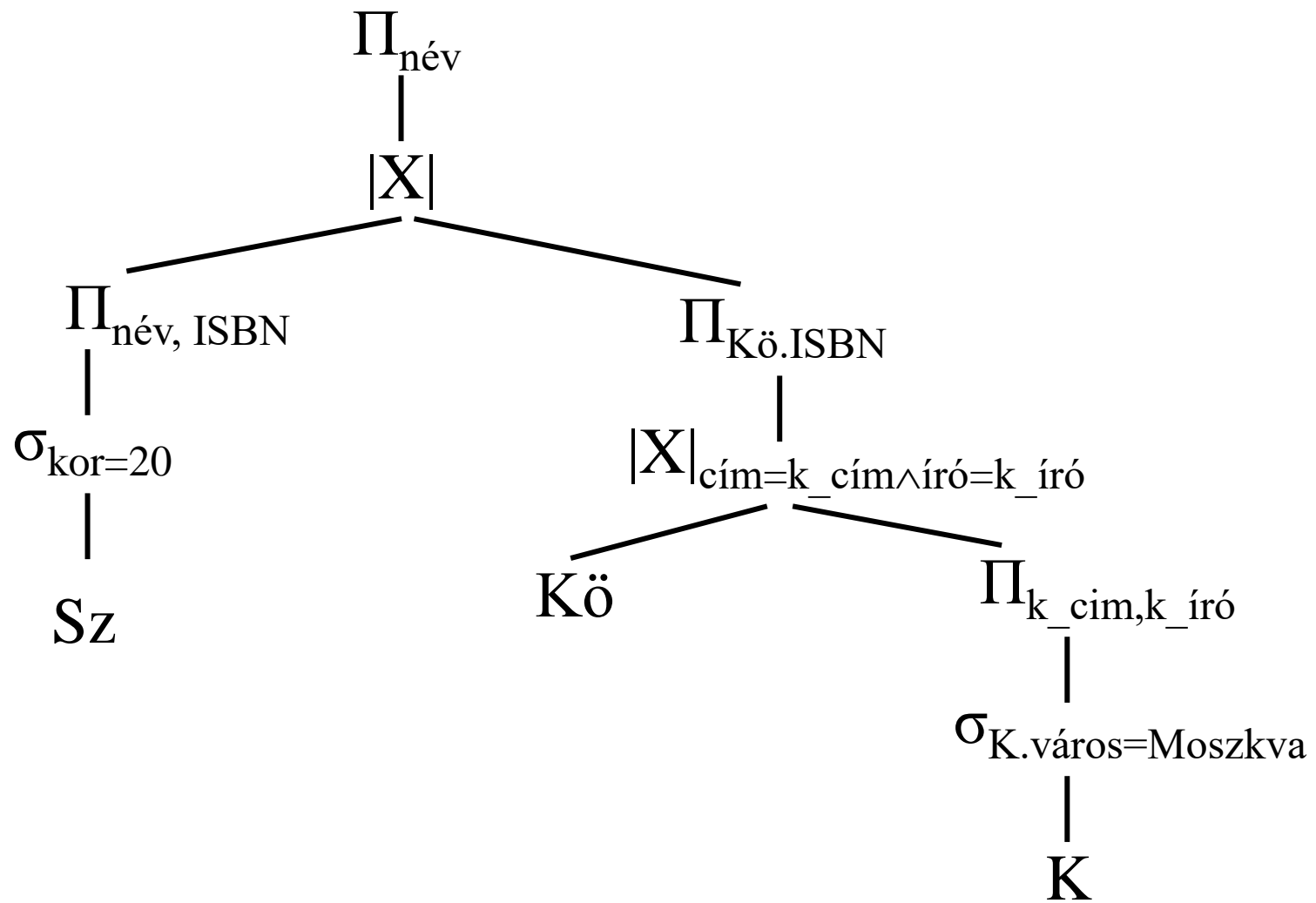
Projekciók "beírása"

- Cél: csak azokat az oszlopokat megtartani a (köztes) relációkban, amelyekre később szükség lesz
- Általában itt nem olyan nagy a nyereség, a projekciók végrehajtása időigényes stb.
- Az átalakításokhoz a projekciókra vonatkozó szabályok használata



Összekapcsolások

- Utolsó lépés: $\Pi_L(\sigma_C(R \times S))$, $\sigma_C(R \times S)$ kifejezések helyettesítése természetes összekapcsolással, Théta-összekapcsolással



Személy (név, kor, város, ISBN)

Könyv (cím, író, ISBN, ár)

Kiad (k_cím, k_író, város, ország)

Mi történik, ha a diszjunkció is megjelenik?

- Kik azok, akik 1000 forintos könyvet vásároltak, és még nincsenek 40 évesek, vagy moszkvaiak, és orosz kiadású könyvet vettek?

Személy (név, kor, város, ISBN)
Könyv (cím, író, ISBN, ár)
Kiad (k_cím, k_író, város, ország)

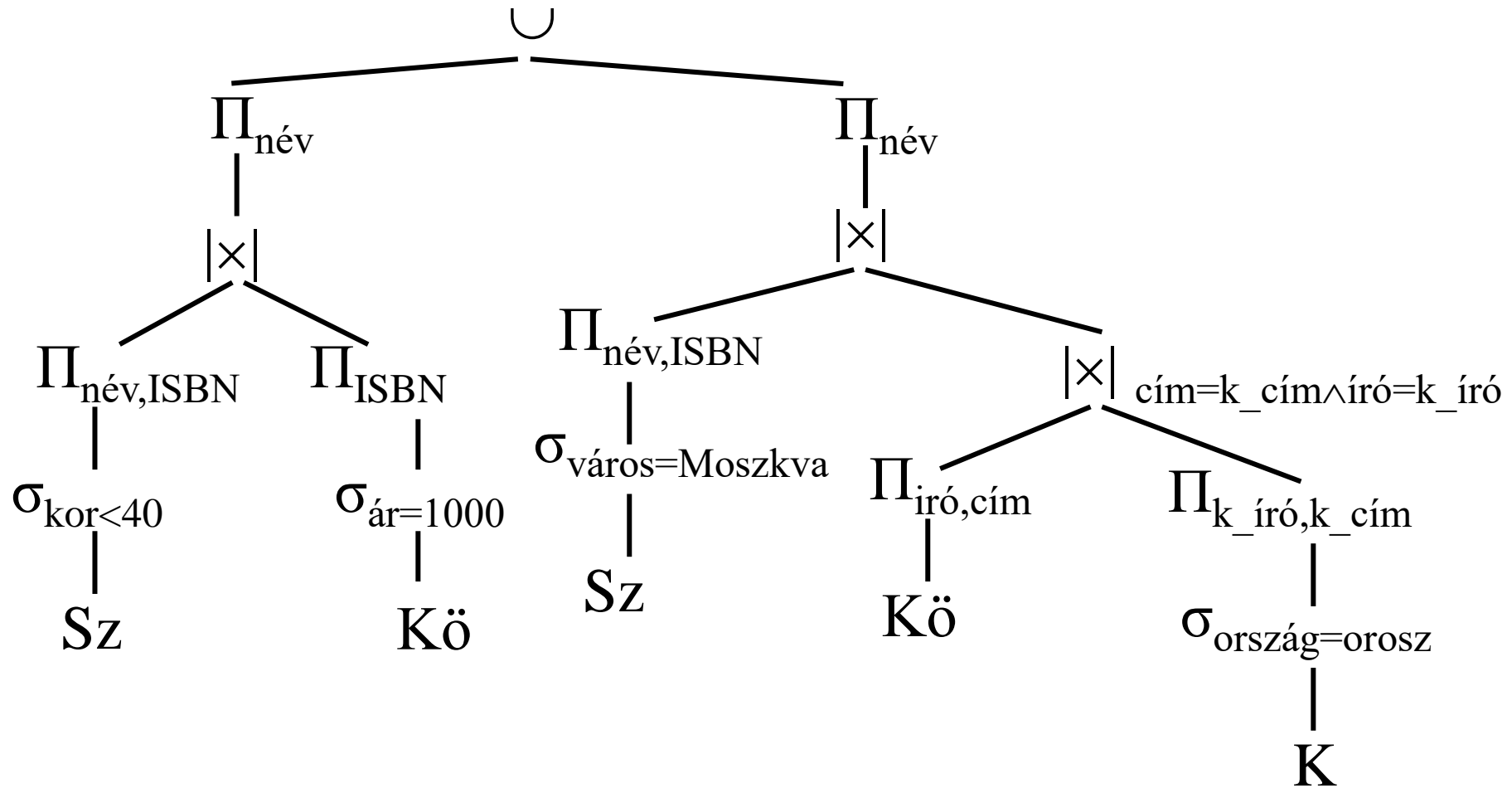
Mi történik, ha a diszjunkció is megjelenik?

- Kik azok, akik 1000 forintos könyvet vásároltak, és még nincsenek 40 évesek, vagy moszkvaiak, és orosz kiadású könyvet vettek?

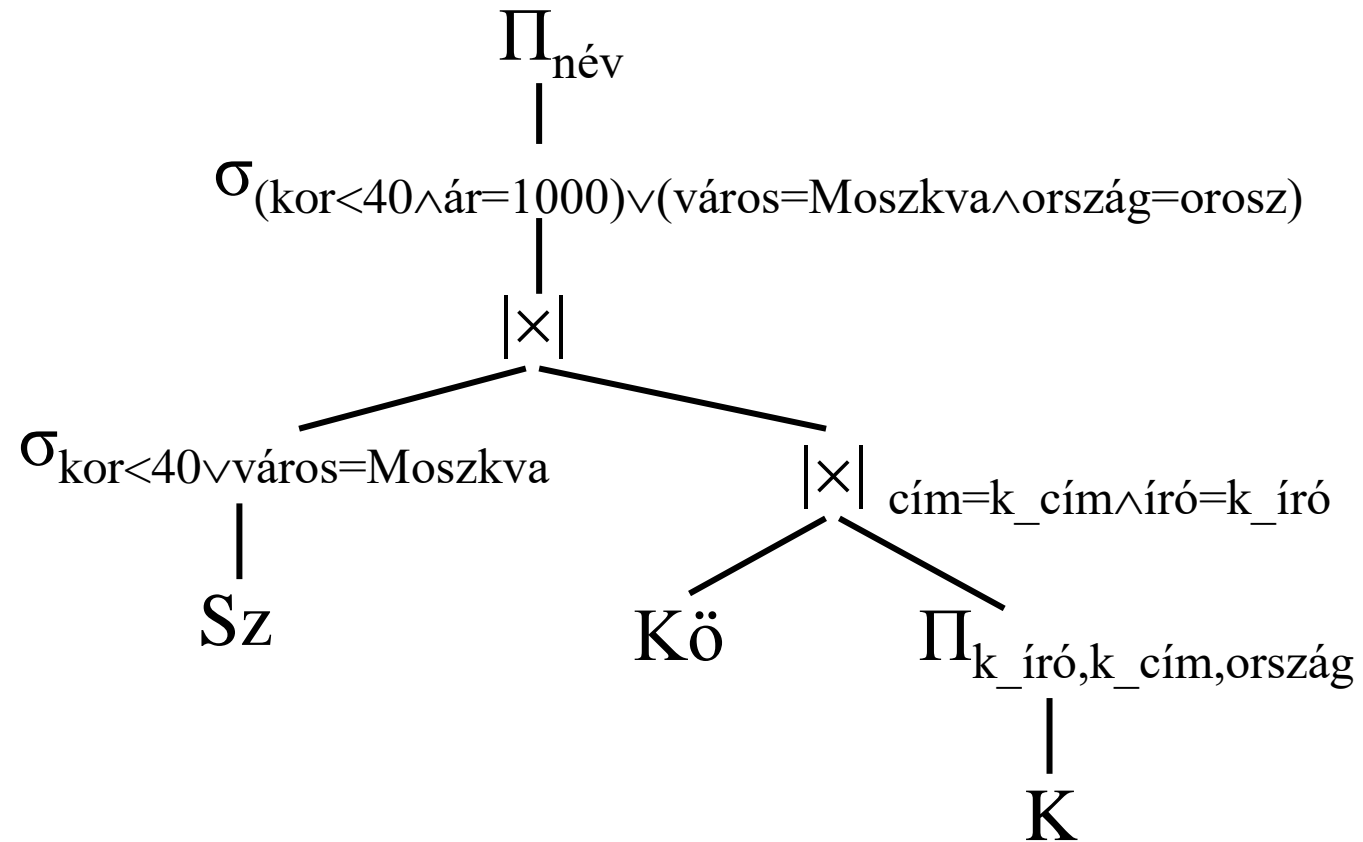
$$\Pi_N(\sigma_{C \wedge ((\text{ár}=1000 \wedge \text{kor} < 40) \vee (\text{Sz.város}=\text{Moszkva} \wedge \text{ország}=\text{orosz}))} (\text{Sz} \times \text{Kö} \times \text{K})).$$

- Itt **C** az $\text{Sz.ISBN} = \text{Kö.ISBN} \wedge \text{Kö.cím} = \text{K.k_cím} \wedge \text{Kö.író} = \text{K.k_író}$ feltételt jelöli.

Megoldás I.



Megoldás II.



Összegzés

- Ha tehát a kiválasztások feltételei diszjunkciót is tartalmaznak, **a helyzet bonyolultabbá válik**, és nem adható olyan egyértelmű optimalizációs algoritmus, mint konjunkciók esetén.

Kiválasztások feljebb csúsztatása

- Lehet, hogy egy kiválasztást először felfelé kell csúsztatni, hogy aztán le lehessen tolni.
- A táblák:

Film (cím, év, hossz)

Szerepel (filmcím, év, színésznev)

```
CREATE VIEW film04 AS
```

```
(SELECT *  
FROM film  
WHERE év = 2004);
```

$\sigma_{\text{év}=2004}(\text{Film})$

- Kik a 2004-ben megjelent filmeknek a szereplői?

Kiválasztások feljebb csúsztatása

- Lehet, hogy egy kiválasztást először felfelé kell csúsztatni, hogy aztán le lehessen tolni.
- A táblák:

Film (cím, év, hossz)

Szerepel (filmcím, év, színésznev)

```
CREATE VIEW film04 AS
```

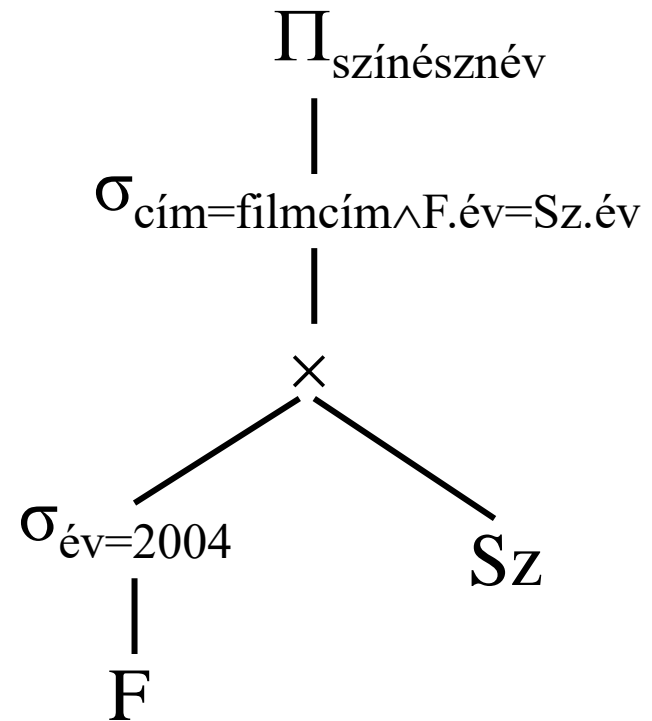
```
(SELECT *  
FROM film  
WHERE év = 2004);
```

$\sigma_{\text{év}=2004}(\text{Film})$

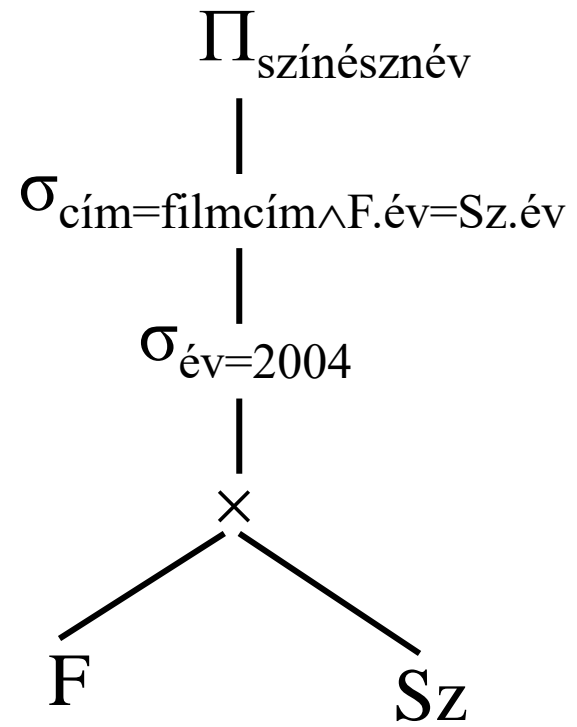
- Kik a 2004-ben megjelent filmeknek a szereplői?

$\Pi_{\text{színésznev}}(\sigma_{\text{cím=filmcím} \wedge \text{film04.év=Szerepel.év}}(\text{film04} \times \text{Szerepel}))$

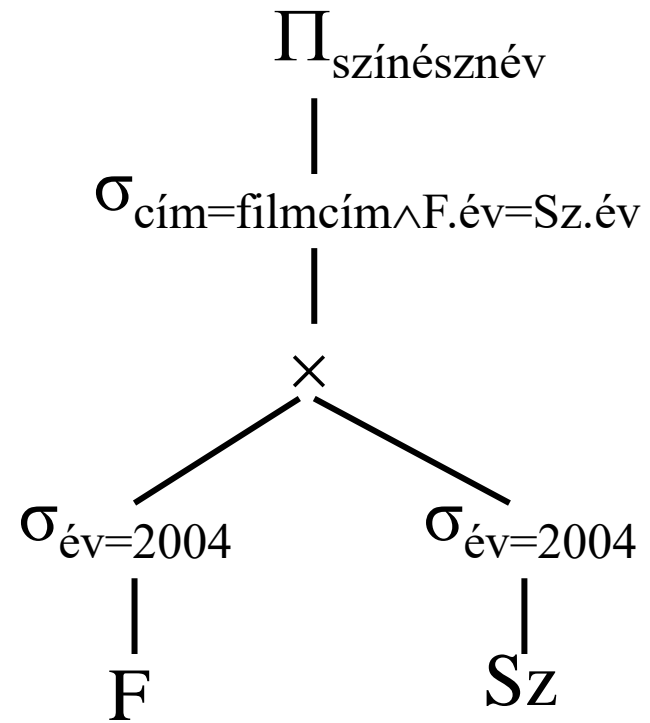
Kezdeti lekérdezésfa



Második lépés



És az eredmény...



Feladat

- A táblák legyenek:
Film (cím, év, hossz)
Szerepel (filmcím, év, színésznév)
Színész (név, kor, város)
- Adjuk meg, hogy a nem budapesti, negyven évesnél idősebb színészek milyen filmekben játszottak 1998-ban. A lekérdezést optimalizáljuk.