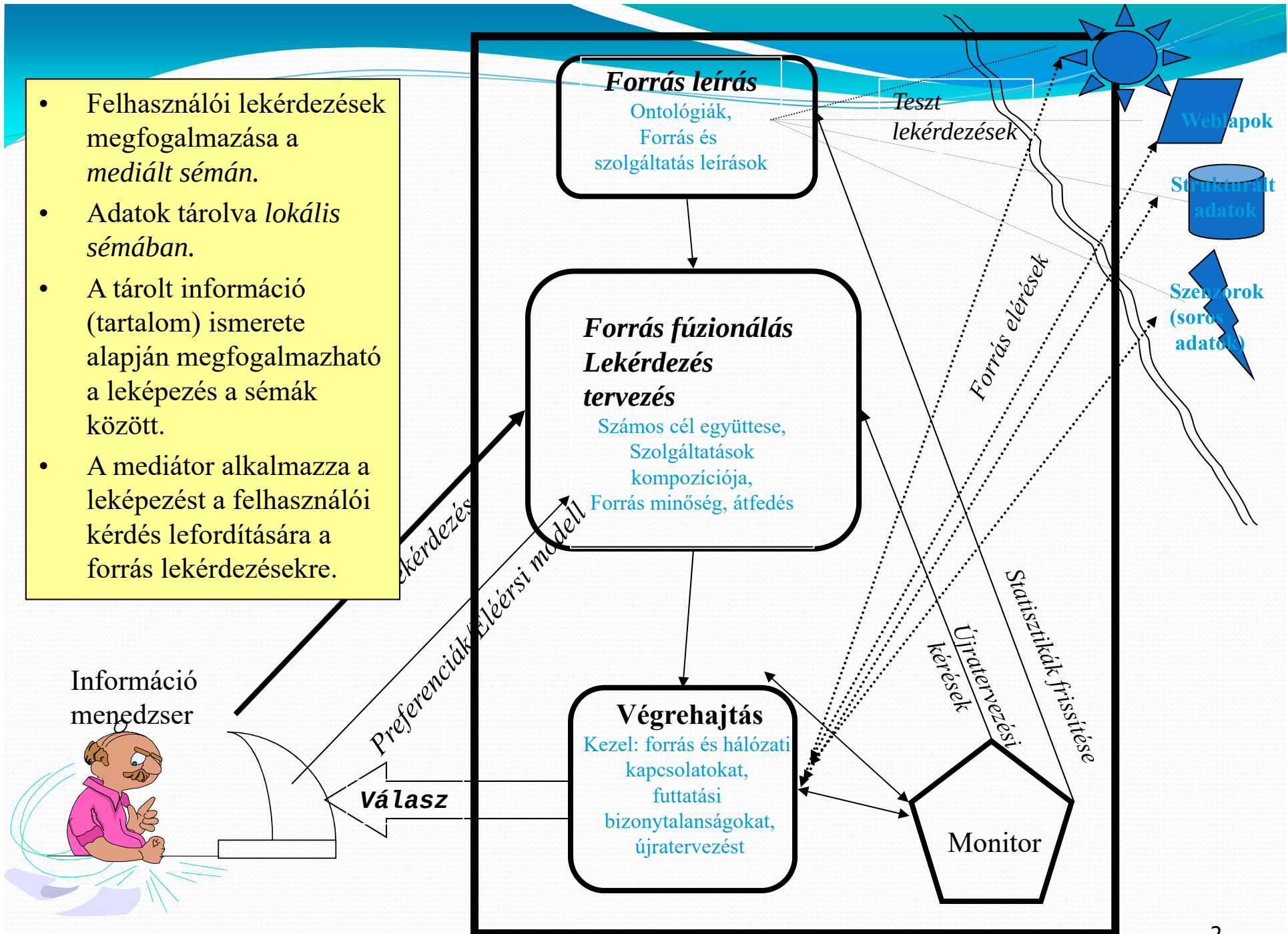


Integrációs és ellenőrzési technikák
VIMIAC04, 2019. tavasz

Információ integráció (GAV példa) 6. Előadás

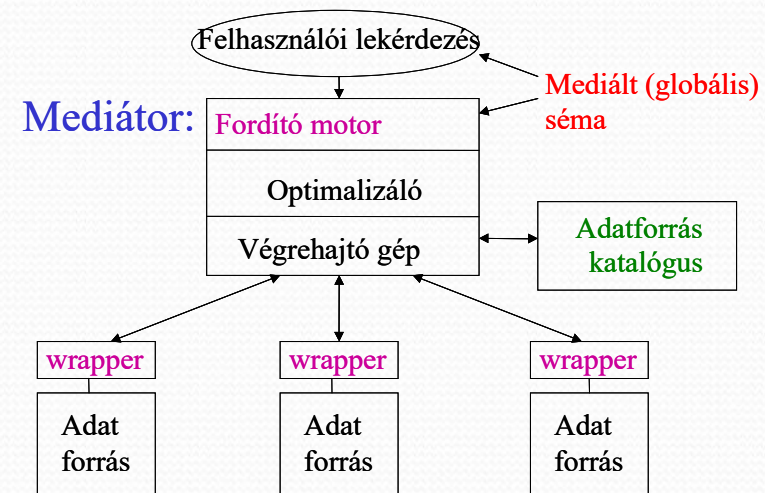
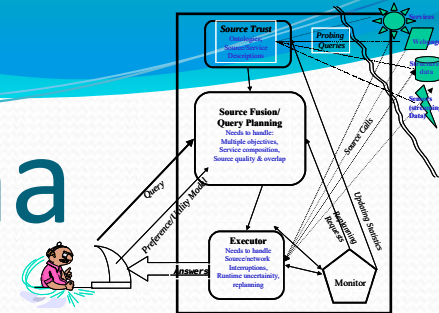
Méréstechnika és Információs Rendszerek Tanszék
<https://www.mit.bme.hu/oktatas/targyak/vimiaco4>

- Felhasználói lekérdezések megfogalmazása a *mediált sémán*.
- Adatok tárolva *lokális sémában*.
- A tárolt információ (tartalom) ismerete alapján megfogalmazható a leképezés a sémák között.
- A mediátor alkalmazza a leképezést a felhasználói kérdés lefordítására a forrás lekérdezésekre.



Virtuális integrációs séma

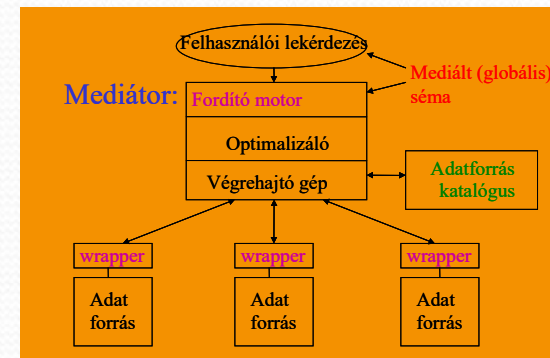
- Adatok a forrásokban maradnak
- Lekérdezés végrehajtásakor:
 - Releváns források meghatározása
 - Lekérdezés szétválasztása forrásokra vonatkozó lekérdezésekre.
 - Válaszok begyűjtése a forrásokból, és megfelelő kombinálása a válasz előállításához.
- Friss adatok
- A megoldás skálázható



Garlic [IBM], Hermes[UMD];Tsimmis, InfoMaster[Stanford]; DISCO[INRIA]; Information Manifold [AT&T]; SIMS/Ariadne[USC];Emerac/Havasu[ASU]

Forrás-mediátor relációs sémával szembeni elvárások

- **Kifejező erő:** hasonló adattartalommal rendelkező források megkülönböztetése, irreleváns források felismerése.
- **Egyszerű bővíthetőség:** tegyük könnyűvé források hozzáadását.
- **Fordítás/átalakítás:** felhasználói lekérdezés lefordítása forrásokon értelmezett lekérdezésekre hatékonyan és eredményesen.
- **Vesztességmentesség:** minden lehetséges adatelérés biztosítása



Lekérdezés átalakítás

- **Adott:**
 - Egy Q lekérdezés a mediátor sémára vonatkozóan
 - Adat források leírása
- **Létrehozandó:**
 - Egy Q' lekérdezés az adat forrásokra vonatkozóan, amely:
 - Q' csak *helyes válaszokat* ad a Q lekérdezéshez és
 - Q' minden lehetséges választ megtalál Q-hoz az elérhető forrásokból.

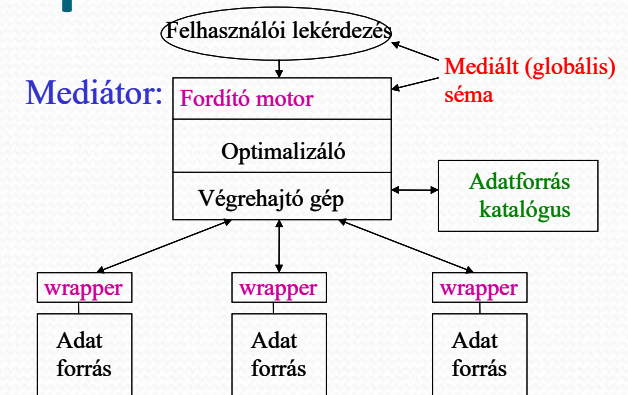
Fordítási/átfogalmazási probléma

- **Adott:**

- Egy Q lekérdezés a mediátor sémára vonatkozóan
- Adat források leírása

- **Létrehozandó:**

- Egy Q' lekérdezés az adat forrásokra vonatkozóan, amely:
 - Q' csak *helyes válaszokat* ad a Q lekérdezéshez és
 - Q' minden lehetséges választ megtalál Q-hoz az elérhető forrásokból.



Forrás és felhasználói sémák reláció leírásának megközelítései

- **Globális mediált sémák(Global-as-view, GAV):**
 - a mediált séma kifejezése a forrásokra vonatkozó nézetek relációjaként
- **Lokális mediált sémák (Local-as-view, LAV):**
 - forrás relációk kifejezése a mediált sémákon értelmezett relációkkal.
- Módszerek kombinációja...?

GAV

VS.

LAV

- Nem moduláris
 - Források hozzáadása módosítja a meglévő mediált séma definícióját
- Nehézkes lehet veszteségmentes mediátort készíteni.
- Lekérdezés átalakítás egyszerű
 - *Nézetek kibontását jelenti (polinomiális)*
 - Hierarchikus mediátor sémák létrehozása lehetséges
- Hatékony, ha
 - Kis számú, ritkán változó adatforrás van
 - Feladat teljesen ismert a mediátor tervezésekor (pl. vállalati adatintegráció)
 - **Garlic, TSIMMIS, HERMES**

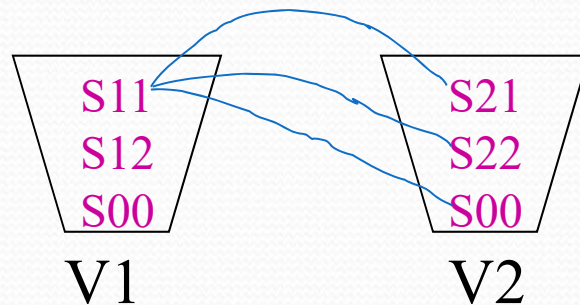
- Moduláris—új forrás hozzáadása egyszerű
- Igen rugalmas – a lekérdező nyelv közvetlenül alkalmazható a források leírására
- Lekérdezés átalakítás bonyolult
 - Válaszokat a nézeteken keresztül kell előállítani (nem mindig megoldható)
- Hatékony, ha
 - Sok, kevésbé korrelált forrás
 - Források dinamikus hozzáadása és törlése
 - **Information Manifold, InfoMaster, Emerac, Havasu**

Átalakítási algoritmusok

$Q(.) :- V1() \& V2()$

$S11() :- V1()$ $S12 :- V1()$
 $S21() :- V2()$ $S22 :- V2()$
 $S00() :- V1(), V2()$

Veder algoritmus



- Veder algoritmus
 - Vedrek kombinációjából előállított terv
 - Utána tartalmazási ellenőrzés

Inverz szabályok

$Q(.) :- V1() \& V2()$

$V1() :- S11()$
 $V1() :- S12()$
 $V1() :- S00()$
 $V2() :- S21()$
 $V2() :- S22()$
 $V2() :- S00()$

- Inverz szabályok
 - Rész lekérdezések tervezés

[Levy]

P_1 contains P_2 if
 $P_2 \models P_1$

[Duschka]

Példa integrátor rendszerekre

- Tsimmis
 - Stanford-IBM Manager of Multiple Information Sources
 - Global-as-View (Gav)
 - Mediátor nézetek a forrás nézetek feletti relációkként definiálva
- Megközelítés:
 - Rugalmas adatmodell
 - Általános lekérdező nyelv
 - Egyedi eszközök

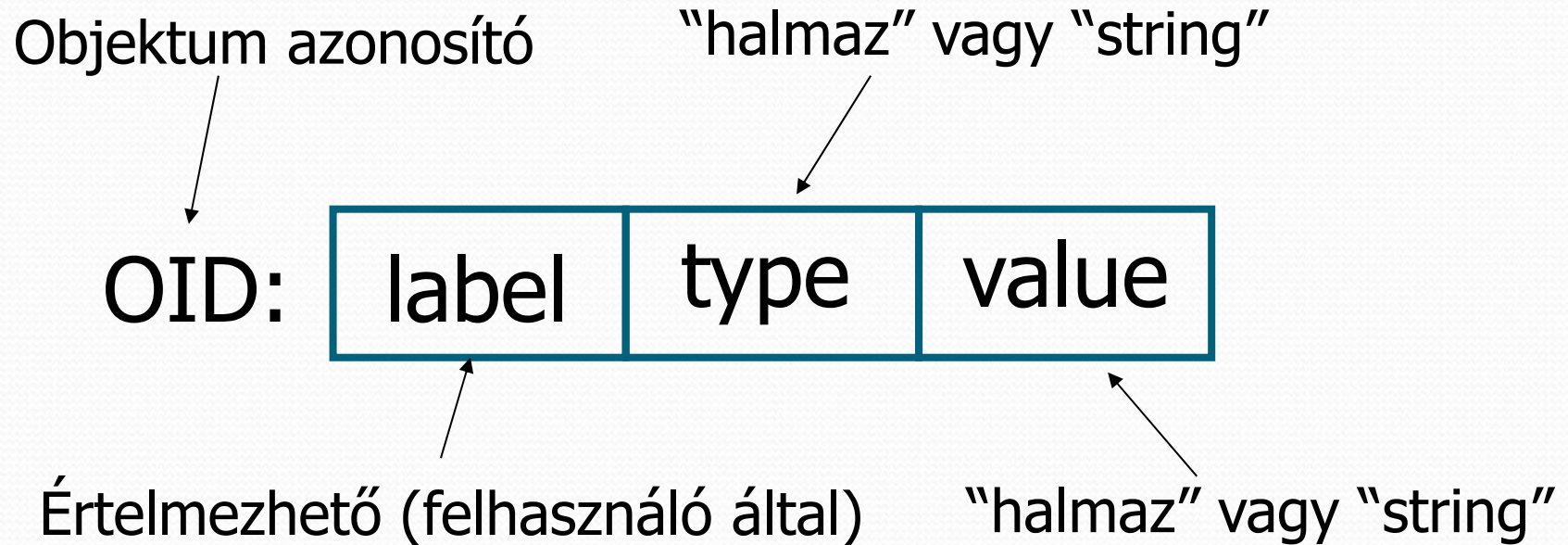
TSIMMIS – Komponensek

- *Adatmodell:*
OEM (Object-Exchange Model)
- *Lekérdező nyelv:*
LOREL (Lightweight Object REpository Language)
- *Mediátor leírás:*
MSL (Mediator Specification Language)
- *Wrapper leírás:*
WSL (Wrapper Specification Language)

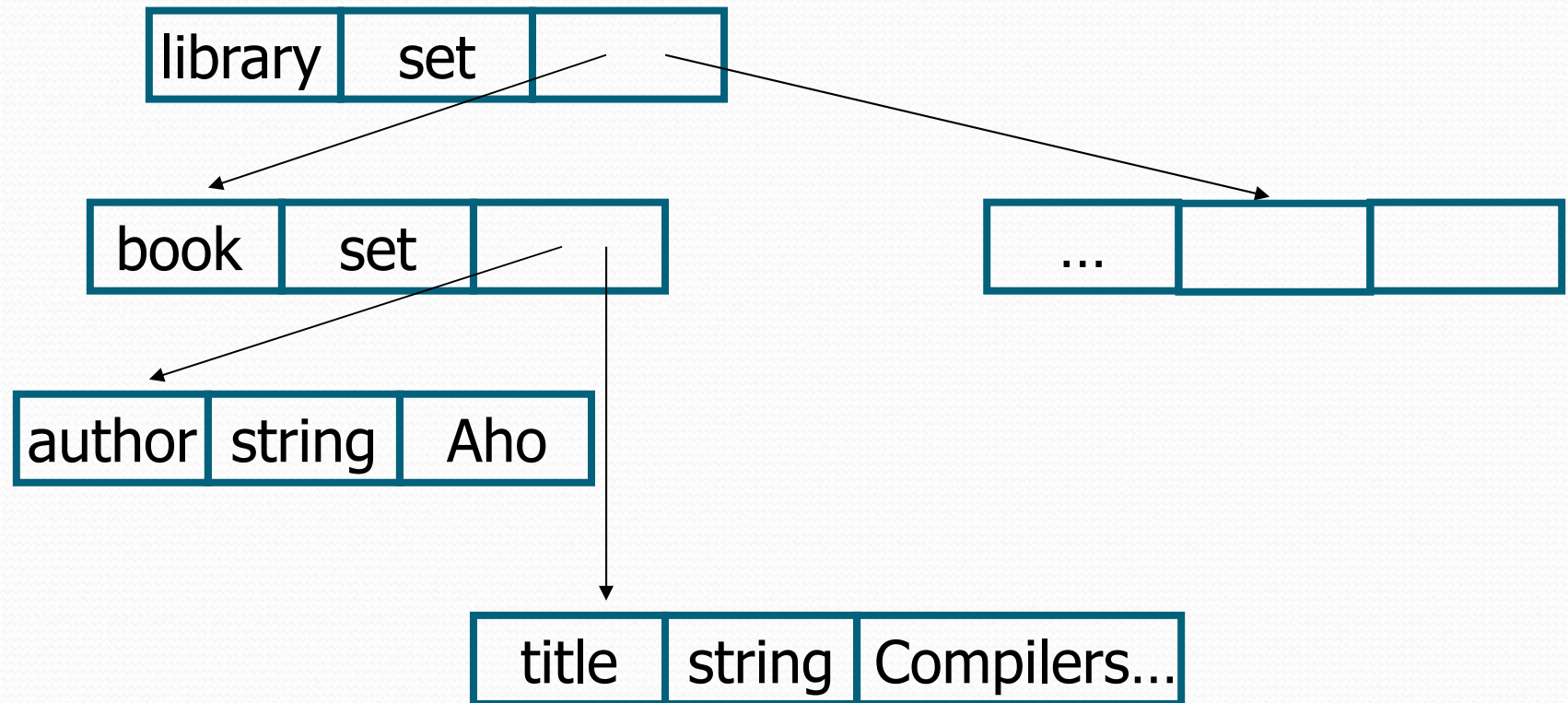
TSIMMIS – OEM

- Objektumok alkalmazása:
Object Exchange Model
- A TSIMMIS adatmodellje
- “önleíró” (az objektum leíró információkat
címkékben (label) helyezi el)
- Rugalmas
- Első rendű logikai megközelítés

TSIMMIS – Objektum modell



TSIMMIS – OEM



TSIMMIS – OEM

Elsőrendű predikátum logikai átírás lehetséges:

123

author	string	Aho
--------	--------	-----

author(T, "Aho")

Objektum azonosító alapján elérhetőek
(lekérdezhetőek) az elemek, akár az
"author" vagy az "Aho" érték.

TSIMMIS – LOREL

- Lekérdező nyelv:
Lightweight Object REpository Language
- Egy OQL lekérdező nyelv OEM adatmodellhez
- A végfelhasználó elérése a TSIMMIS modellhez és az adatokhoz.

Példa:

```
select library.book.title  
from library  
where library.book.author = "Aho"
```

TSIMMIS – LOREL

- Részleges illeszkedést megengedő szemantika:

```
select R.A  
from R, S, T  
where R.A = S.A or R.A = T.A
```

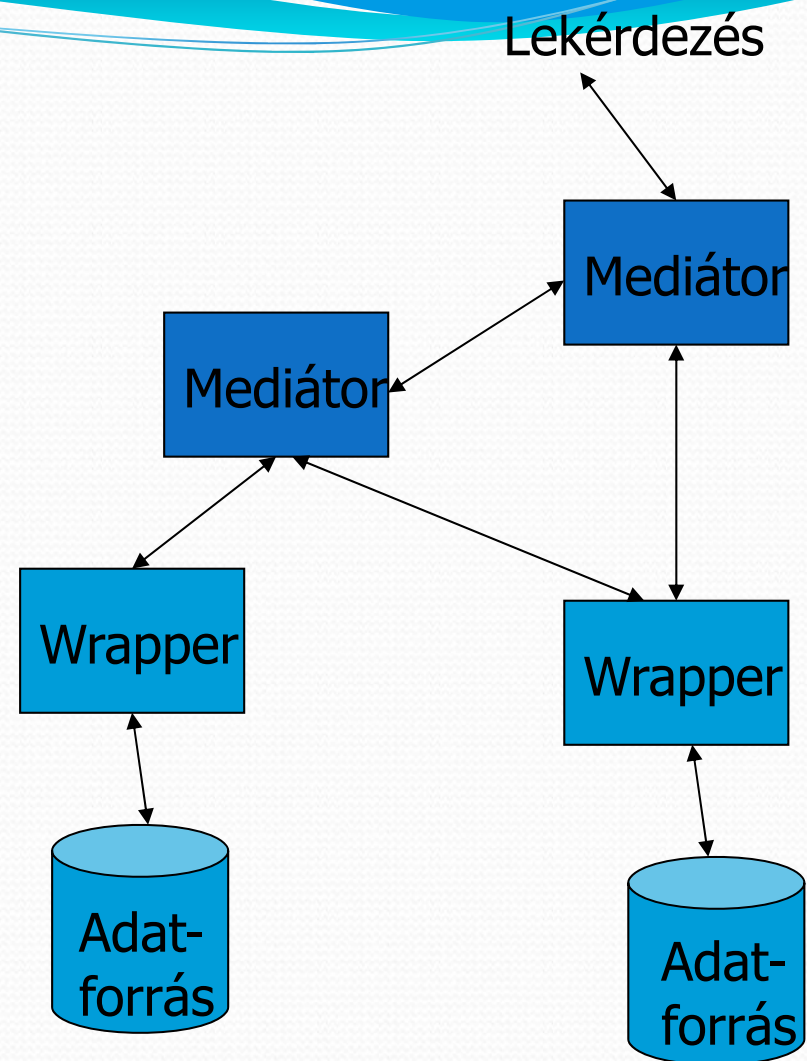
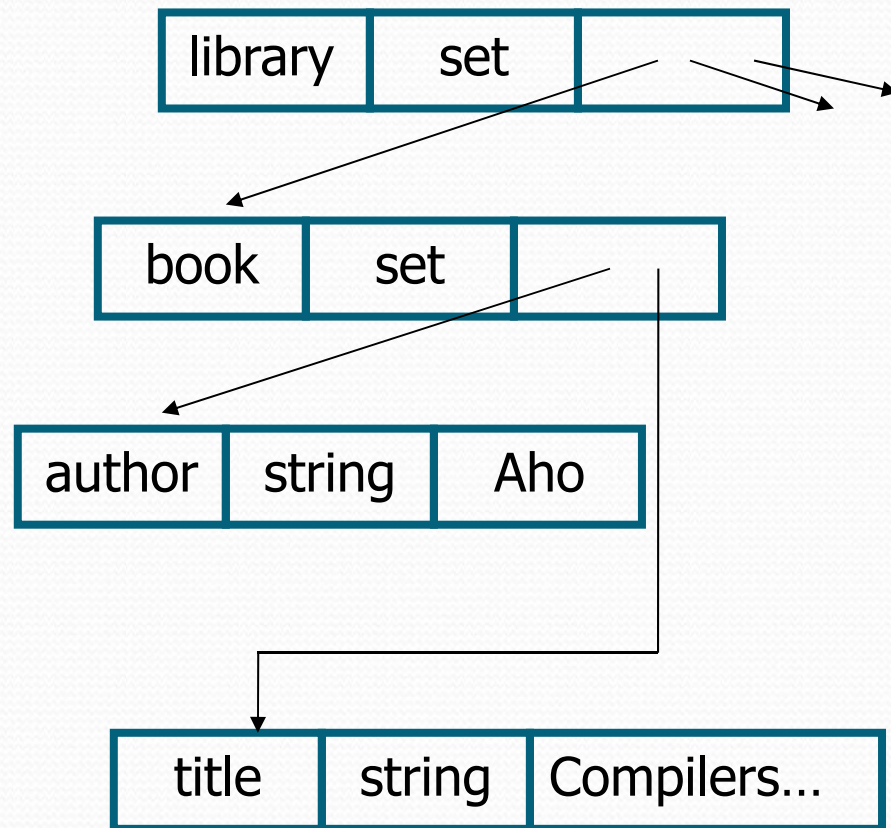
- Csak akkor sikertelen (és nem tér vissza értékkel), ha T és S is üres.

TSIMMIS – MSL

(Mediator Specification Language)

- Mediátorok deklaratív leírása
- Objektum orientált, logikai leírás
- OEM adatmodellre épül

TSIMMIS – MSL



<booktitle X> :-

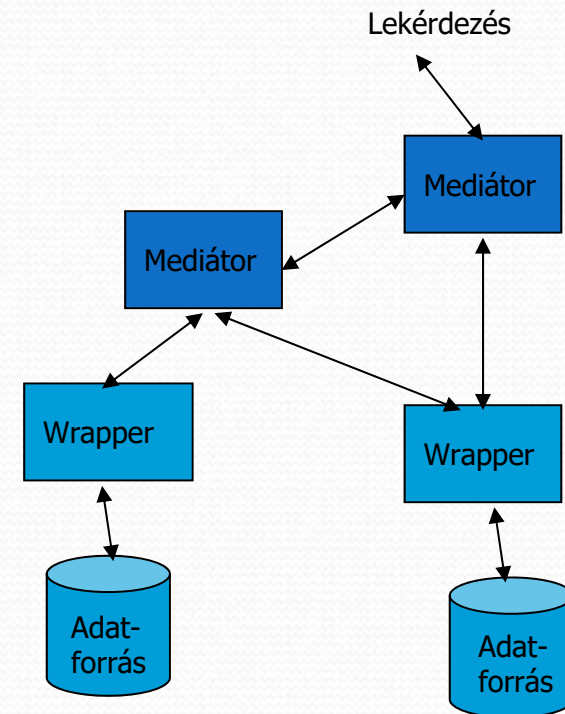
<library { <book { <title X> <author "Aho"> } > } > @s1

TSIMMIS – Wrapper-ek

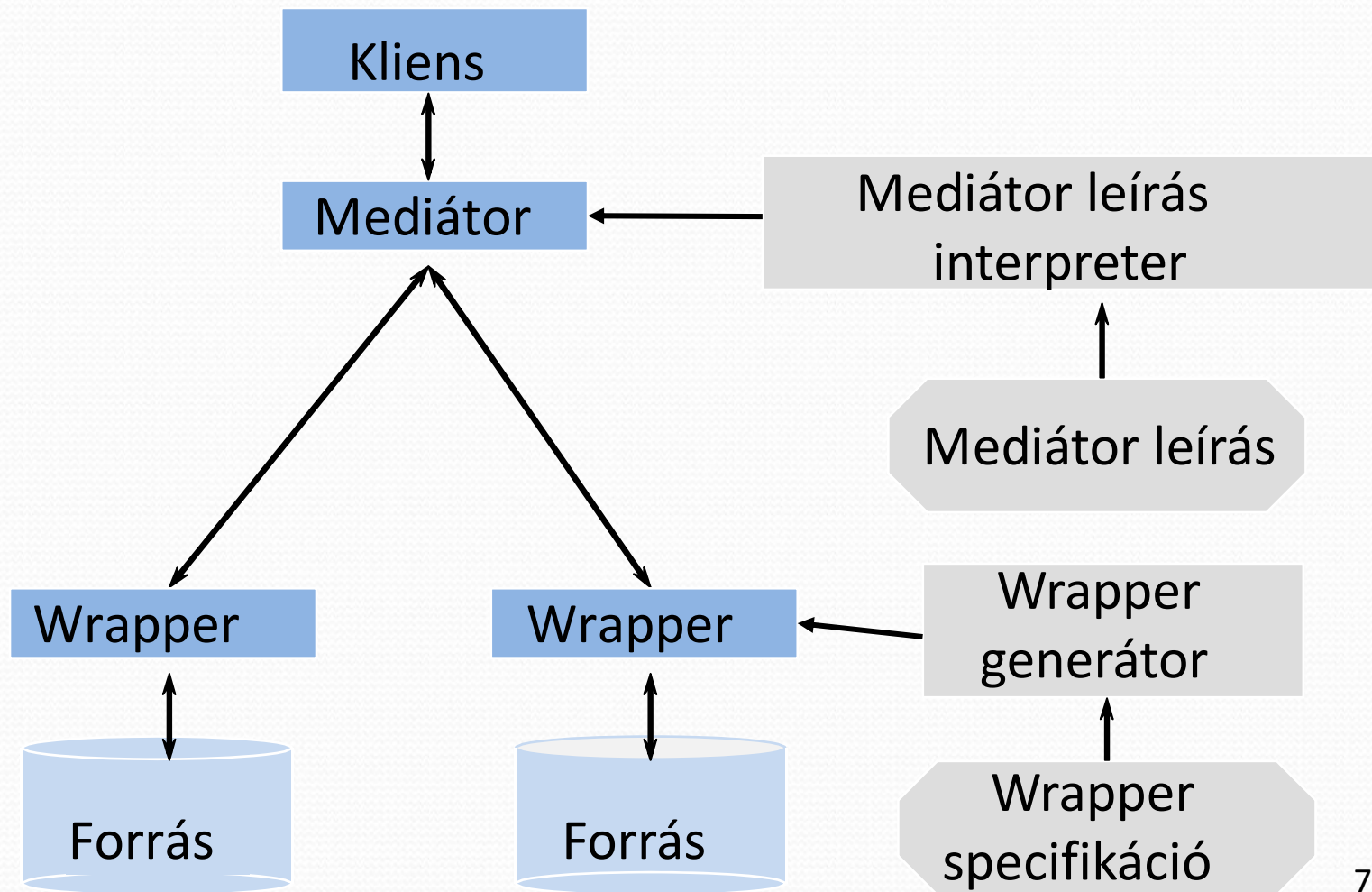
- Wrapper-ek szerepe hasonló, mint az adatbázis meghajtóké
- A Wrapper-ek logikai leírása MSL nyelven történik
- Wrapper leírások formája:
MSL template
// action //

Példa

```
<books X> :- <library { X:<book {<title X>  
                                <author $AU>}}> }>@s1  
// sprintf(lookup-query, "find author %s", $AU) //
```



Mediátor rendszer specifikálása



Kihívás: Strukturálatlan források elemzése

- Félig-struktúrált
 - irreguláris
 - mélyen beágyazott
 - kereszthivatkozások
- Hiányos séma információ
 - autonóm
 - dinamikus
- Példák
 - HTML lapok
 - SGML dokumentumok
 - Gén adatbázisok
 - kémiai elemek
 - könyvtári információk

Adatreprezentáció -TSIMMIS

- **Félig-strukturált adat reprezentáció**
- **Mediátor generálás**
- **Wrapper generálás**
- **Képességeket figyelembe vevő átírás**

Félig-strukturált információk reprezentálása OEM modellel

Szemantikus
objektum azonosító

label

Értékadás

```
<http://www/~doe, faculty, {&f1,&l1,&r1}>
```

```
<&f1, first_name, "John">
```

```
<&l1, last_name, "Doe">
```

```
<&r1, rank, "professor">
```

Atomi érték

Strukturális
objektum azonosító

OEM – Gráf reprezentáció

<http://www/~doe, **faculty**, {&f1,&l1,&r1}>

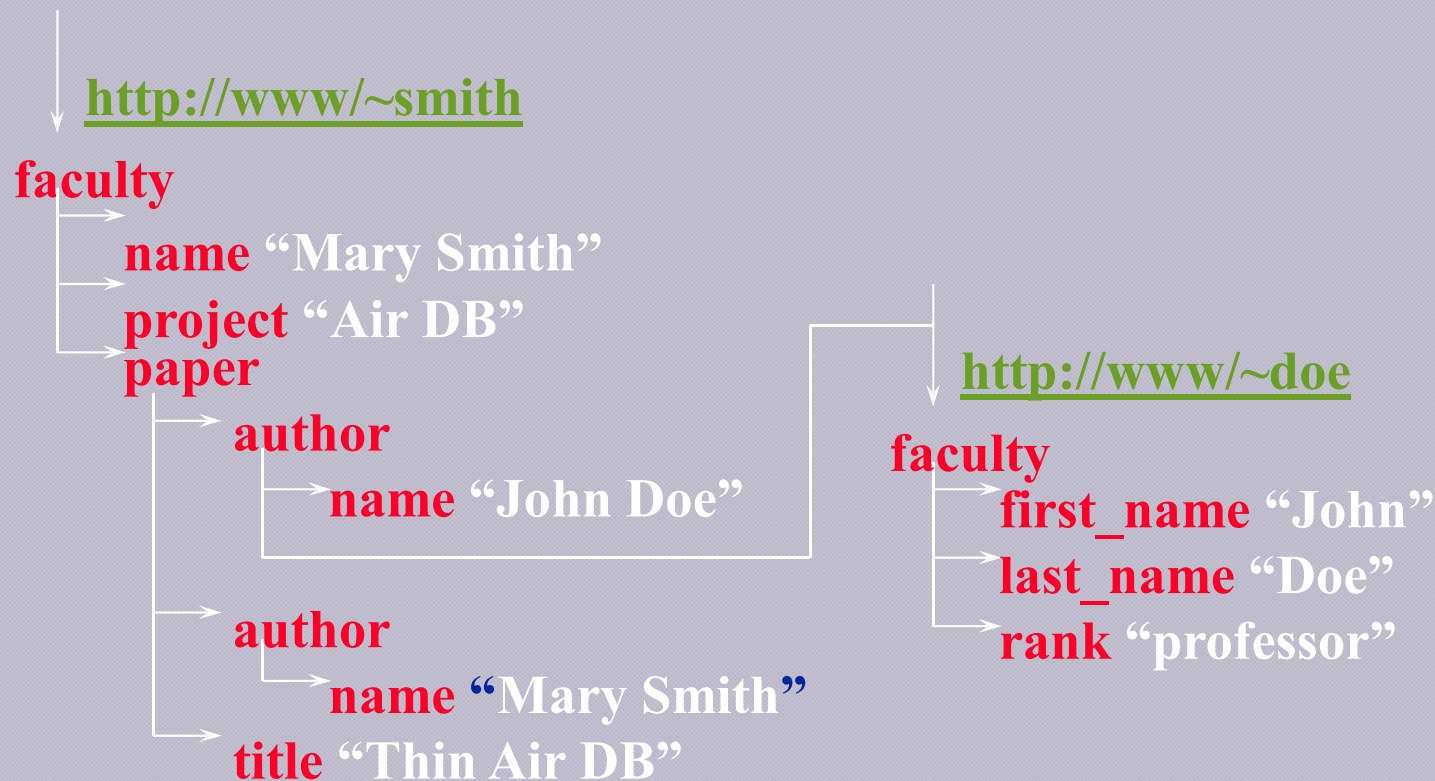
<&f1, **first_name**, “John”>

<&l1 **last_name**, “Doe”>

<&r1 **rank**, “professor”>



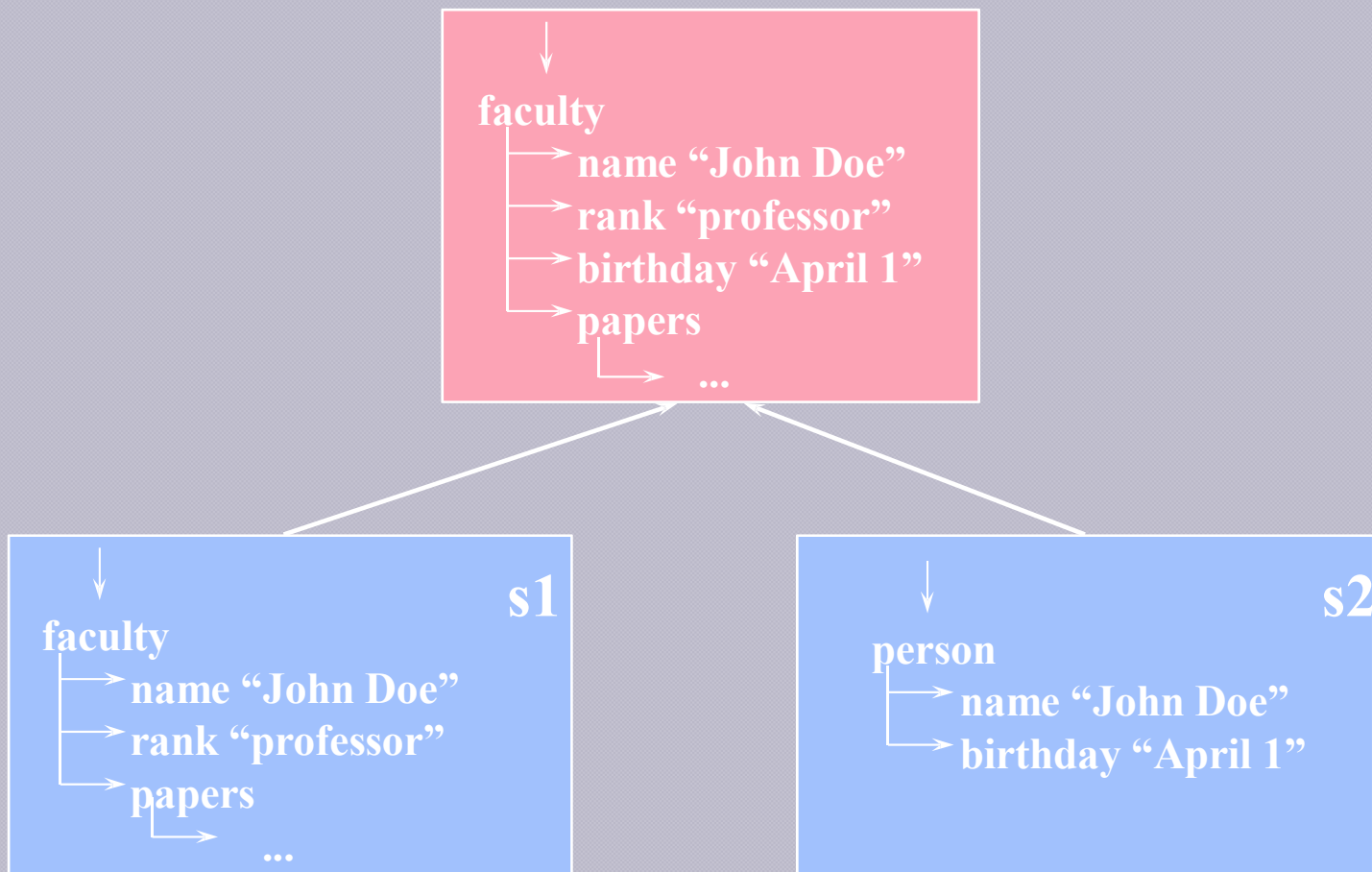
OEM struktúra tetszőleges gráf reprezentációt megenged



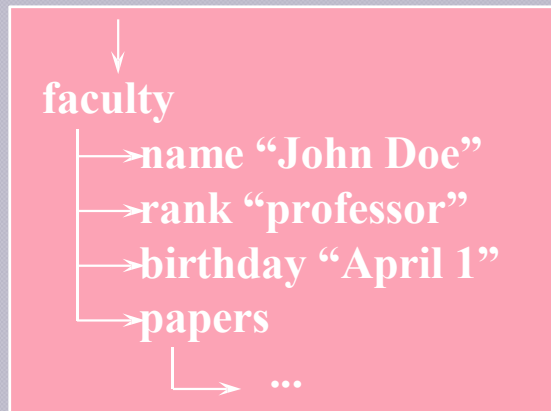
Áttekintés

- Félig-strukturált adat reprezentáció
- ➔ • Mediátor generálás
 - Mediátor specifikációs példák
 - Nyelv kifejezőerő
 - Megvalósítás és teljesítmény
- Wrapper generálás
- Képességeket figyelembe vevő átírás

Információ egyesítés

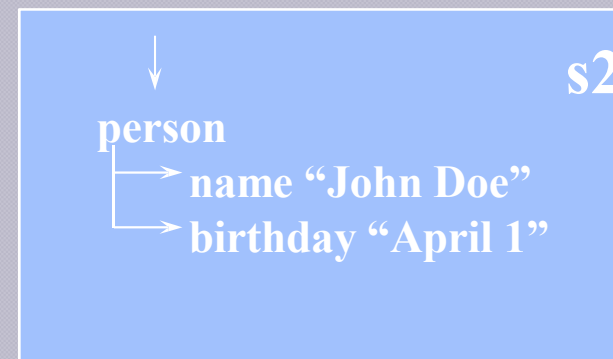
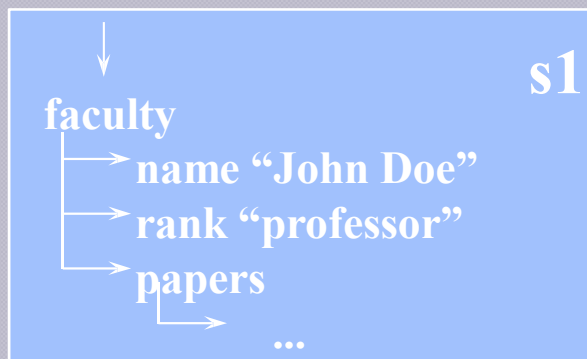


Mediátor specifikáció példa

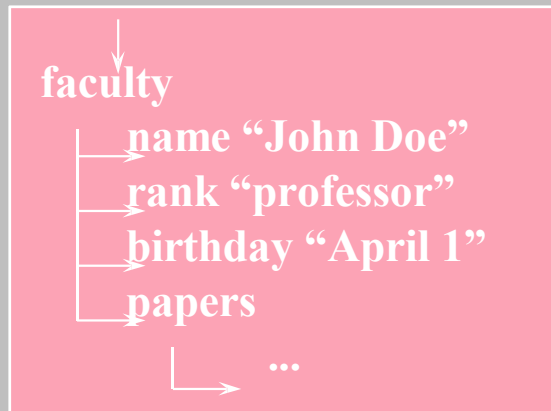


$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } N \rangle \ \langle L \ V \rangle \} \rangle @s1$

$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{person } \{ \langle \text{name } N \rangle \ \langle L \ V \rangle \} \rangle @s2$

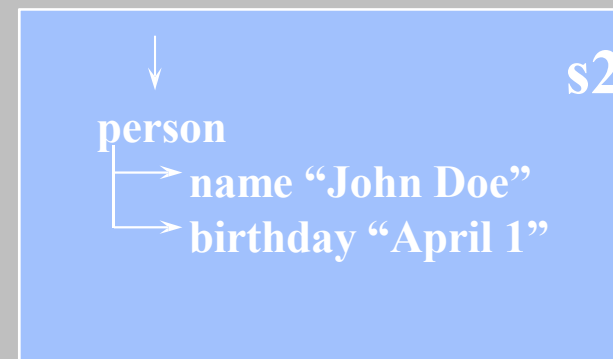
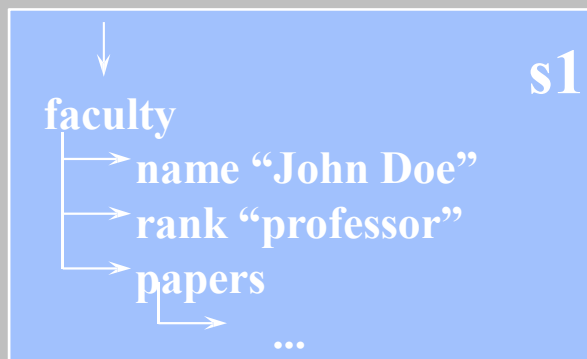


Mediátor specifikáció példa: Szabályok törzsének elemei

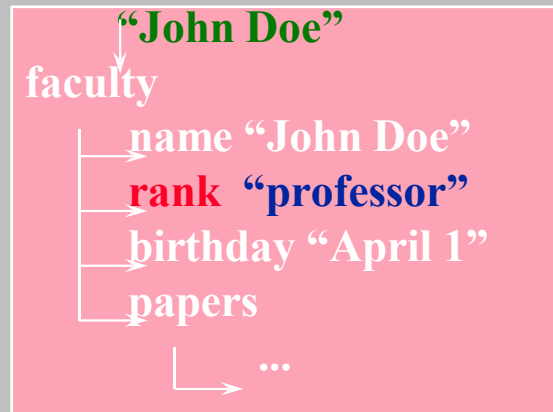


$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } N \rangle \langle L \ V \rangle \} \rangle @s1$

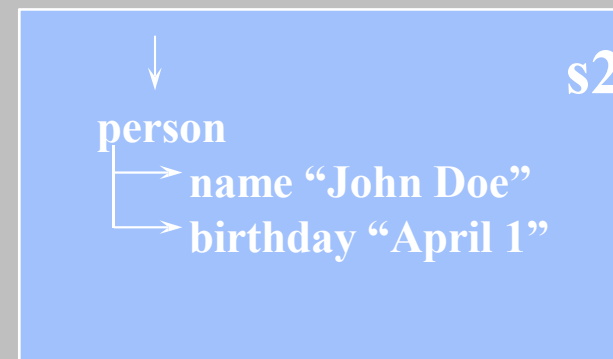
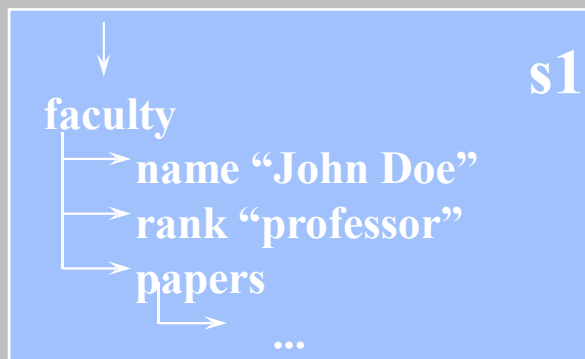
$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{person } \{ \langle \text{name } N \rangle \langle L \ V \rangle \} \rangle @s2$



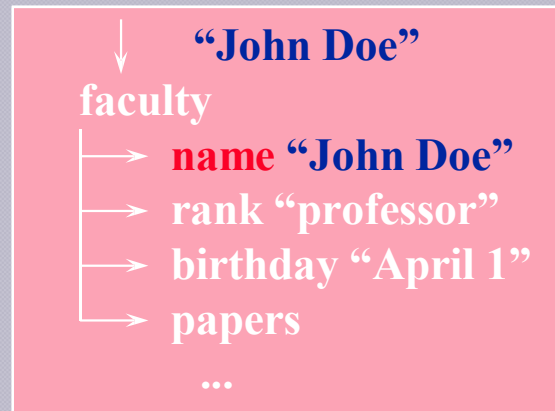
Mediátor specifikáció példa: Szabályok fejének értelmezése



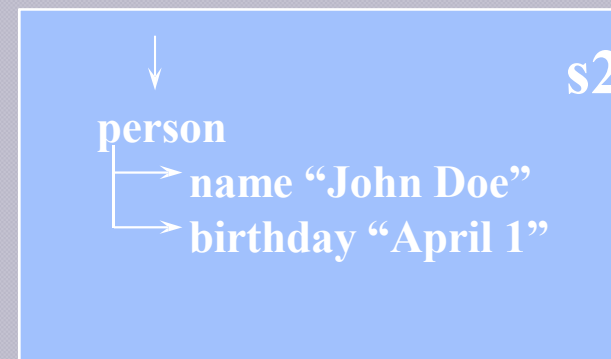
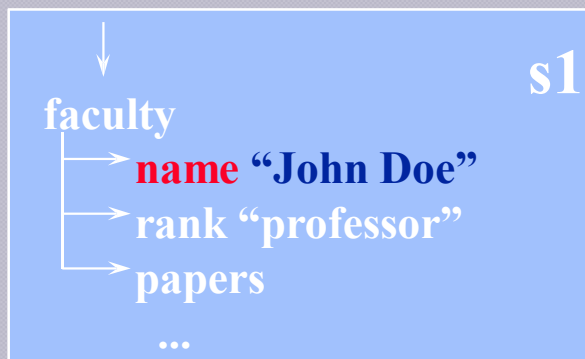
$\langle \mathbf{N} \text{ faculty } \{ \langle \mathbf{L} \mathbf{V} \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } \mathbf{N} \rangle \langle \mathbf{L} \mathbf{V} \rangle \} \rangle @s1$
 $\langle \mathbf{N} \text{ faculty } \{ \langle \mathbf{L} \mathbf{V} \rangle \} \rangle \text{ :- } \langle \text{person } \{ \langle \text{name } \mathbf{N} \rangle \langle \mathbf{L} \mathbf{V} \rangle \} \rangle @s2$



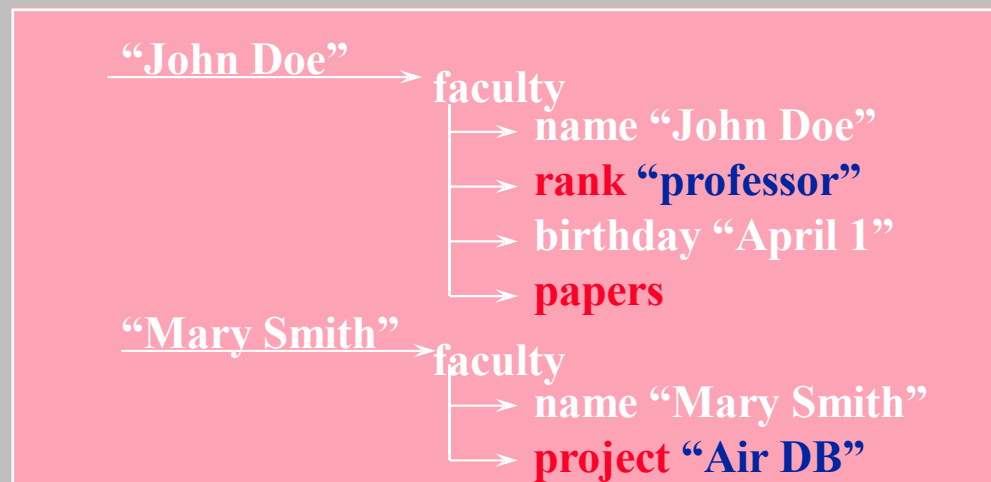
Szemantikusan azonosított objektum hozzáadása



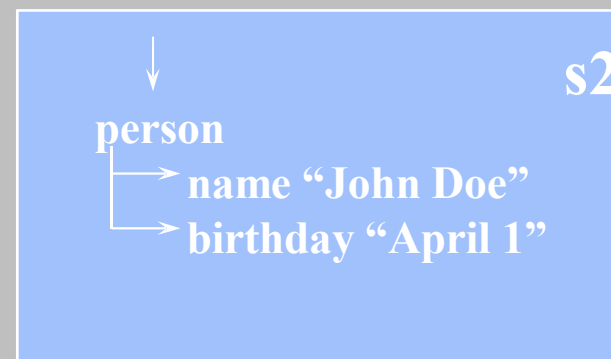
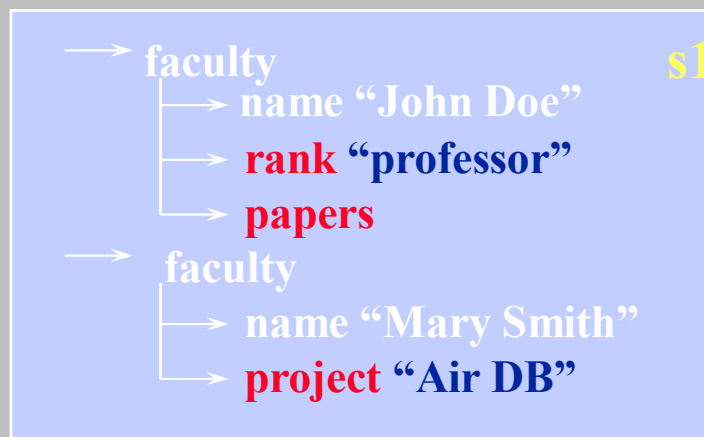
$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } N \rangle \langle L \ V \rangle \} \rangle @s1$
 $\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{person } \{ \langle \text{name } N \rangle \langle L \ V \rangle \} \rangle @s2$



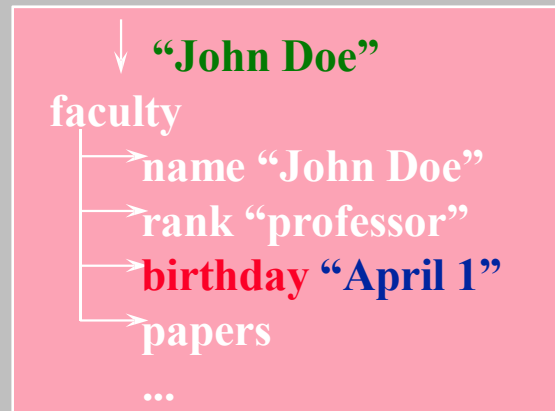
Irregularis és hiányzó információ



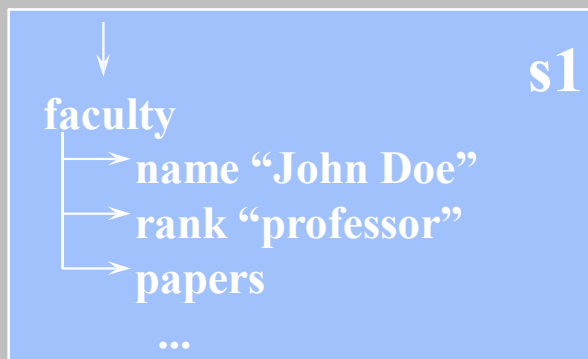
$\langle N \text{ faculty } \{ \langle \mathbf{L} \mathbf{V} \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } N \rangle \langle \mathbf{L} \mathbf{V} \rangle \} \rangle @s1$



A második szabály további elemet rendel a nézet objektumaihoz



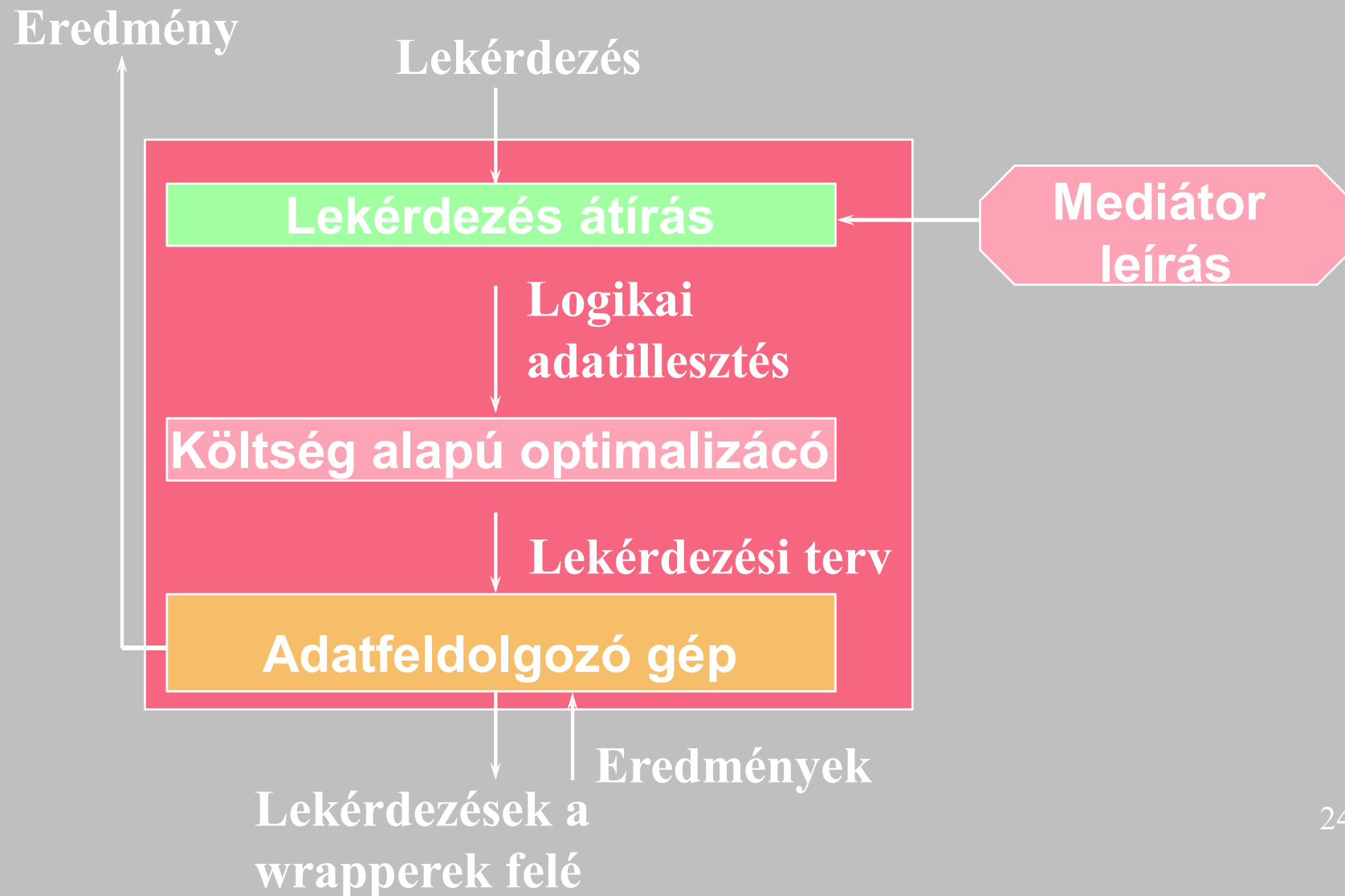
$\langle N \text{ faculty } \{ \langle L V \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } N \rangle \langle L V \rangle \} \rangle @s1$
 $\langle N \text{ faculty } \{ \langle L V \rangle \} \rangle \text{ :- } \langle \text{person } \{ \langle \text{name } N \rangle \langle L V \rangle \} \rangle @s2$



A nyelv kifejezőereje

- MSL nyelvvel megoldott problémák
 - Irregularitások
 - *Hiányos struktúra információk*
 - Kereszthivatkozások kezelése
 - Inkonzisztens és redundáns adat
 - Tetszőleges illeszkedési kritérium

Mediátor leírás interpreter architektúra



Lekérdezés átírás

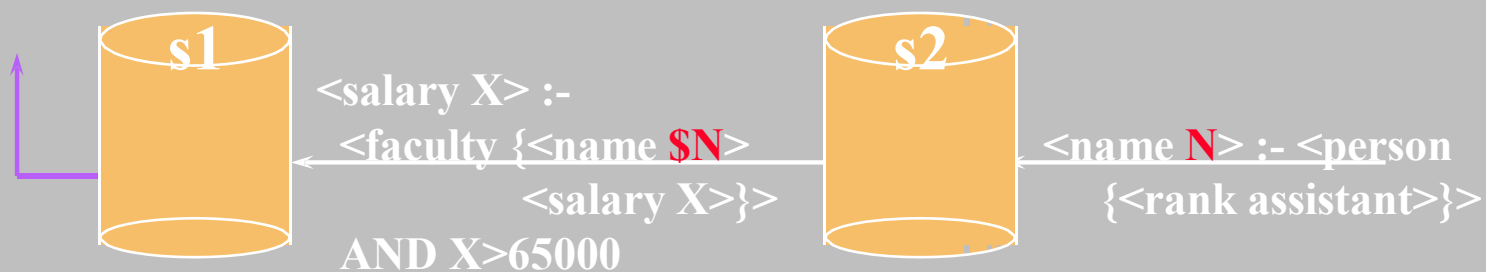
- $\langle N \text{ faculty } \{ \langle \text{salary } S \rangle \} \rangle :-$
 $:- \langle \text{faculty } \{ \langle \text{name } N \rangle \langle \text{salary } S \rangle \} \rangle @s1$
 $\langle N \text{ faculty } \{ \langle \text{rank } R \rangle \} \rangle$
 $:- \langle \text{person } \{ \langle \text{name } N \rangle \langle \text{rank } R \rangle \} \rangle @s2$
- $\langle \text{well-paid } \{ \langle \text{name } N \rangle \langle \text{salary } X \rangle \} \rangle$
 $:- \langle N \text{ faculty } \{ \langle \text{salary } X \rangle \langle \text{rank assistant} \rangle \} \rangle$
 $\text{AND } X > 65000$

Lekérdezés átírás: feltételek átadása a források fele

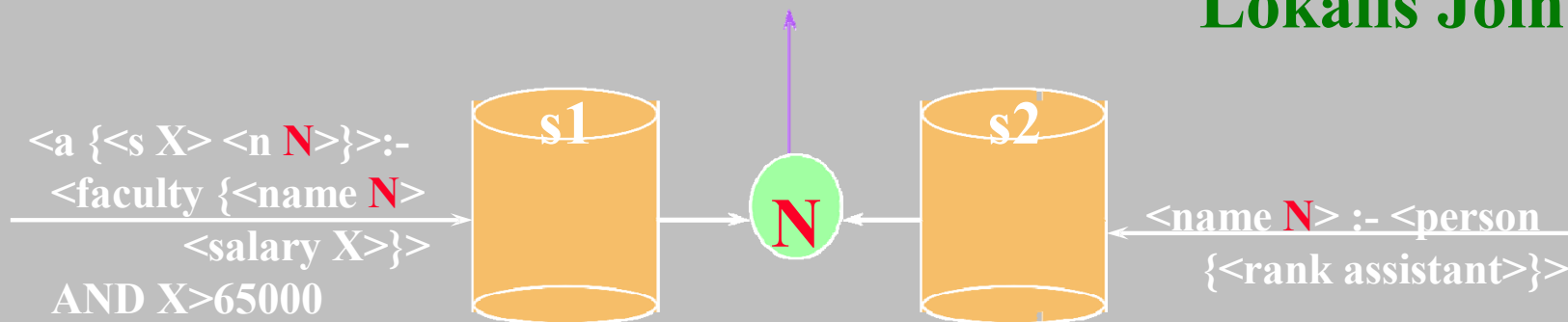
- $\langle N \text{ faculty } \{ \langle \text{salary } S \rangle \} \rangle :-$
 $:- \langle \text{faculty } \{ \langle \text{name } N \rangle \langle \text{salary } S \rangle \} \rangle @s1$
 $\langle N \text{ faculty } \{ \langle \text{rank } R \rangle \} \rangle$
 $:- \langle \text{person } \{ \langle \text{name } N \rangle \langle \text{rank } R \rangle \} \rangle @s2$
- $\langle \text{well-paid } \{ \langle \text{name } N \rangle \langle \text{salary } X \rangle \} \rangle$
 $:- \langle N \text{ faculty } \{ \langle \text{salary } X \rangle \langle \text{rank assistant} \rangle \} \rangle \text{ AND }$
 $X > 65000$
- **logical datamerge program**
 $\langle \text{well-paid } \{ \langle \text{name } N \rangle \langle \text{salary } X \rangle \} \rangle$
 $:- (\langle \text{faculty } \{ \langle \text{name } N \rangle \langle \text{salary } X \rangle \} \rangle \text{ AND }$
 $X > 65000) @s1$
 $\text{AND } \langle \text{person } \{ \langle \text{name } N \rangle \langle \text{rank assistant} \rangle \} \rangle @s2$

Lekötések átadása & Lokális join tervek

Lekötések átadása



Lokális Join



Lekérdezés dekomponálása ismeretlen eredet figyelembevételével

<N faculty {<L V>}> :- <faculty {<name N> <L V>}>@s1

<N faculty {<L V>}> :- <person {<name N> <L V>}>@s2

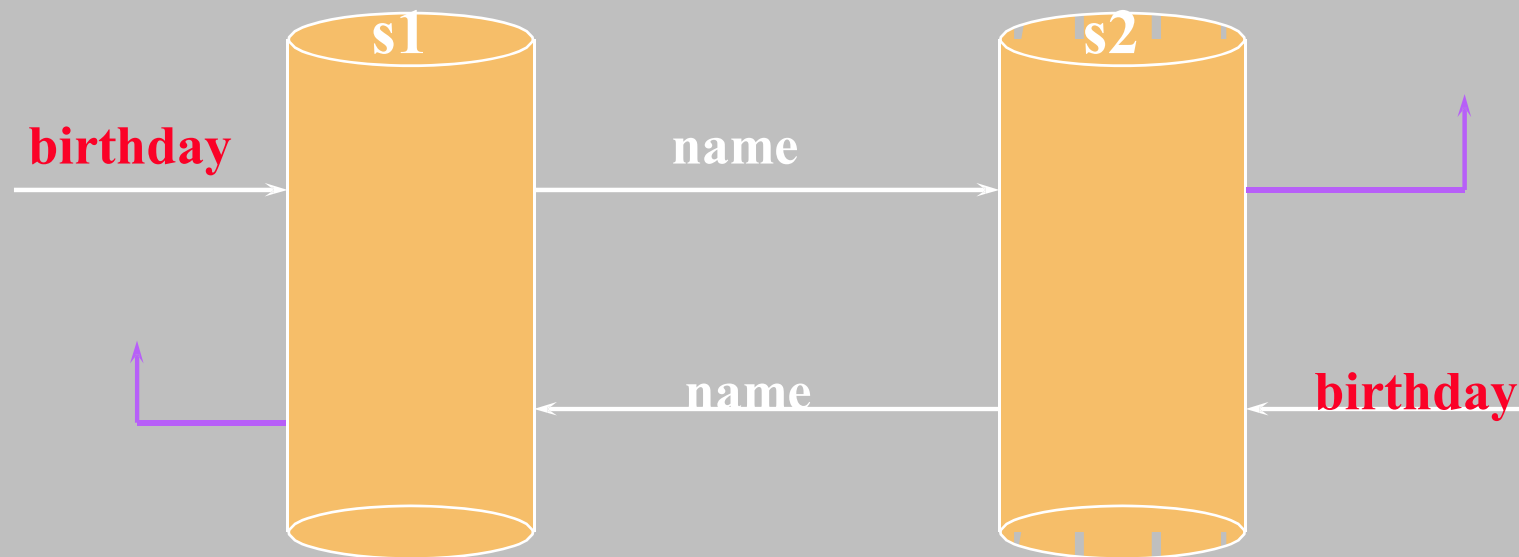
<X faculty {<S Y>}> :- <X faculty {<birthday “1/20”> <S Y>}>

A terv figyelembe vesz minden lehetséges forrást a „birthday„ értékére

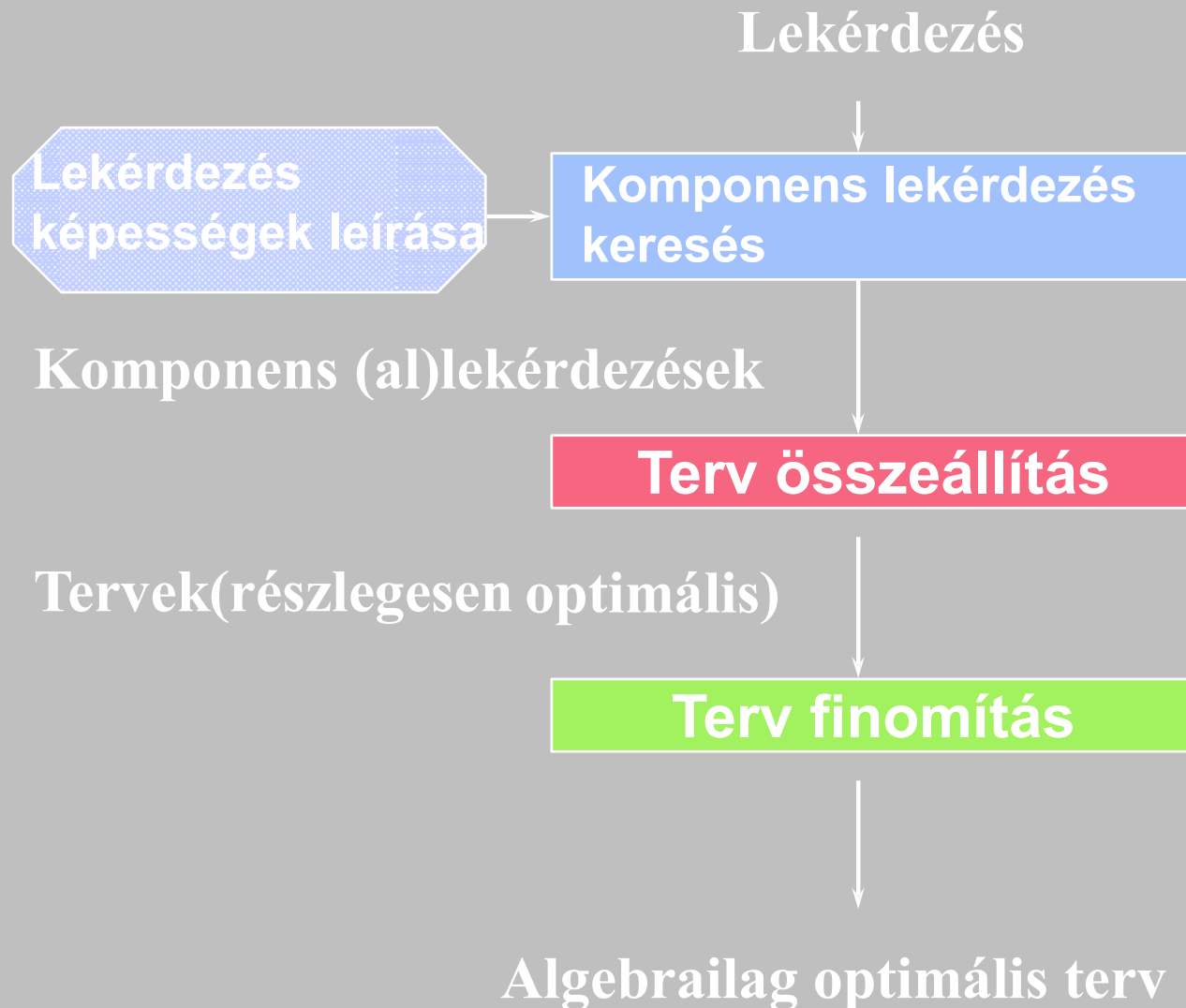
$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{faculty } \{ \langle \text{name } N \rangle \ \langle L \ V \rangle \} \rangle @s1$

$\langle N \text{ faculty } \{ \langle L \ V \rangle \} \rangle \text{ :- } \langle \text{person } \{ \langle \text{name } N \rangle \ \langle L \ V \rangle \} \rangle @s2$

$\langle X \text{ faculty } \{ \langle S \ Y \rangle \} \rangle \text{ :- } \langle X \text{ faculty } \{ \langle \text{birthday "1/20"} \rangle \ \langle S \ Y \rangle \} \rangle$



Képesség alapú átírás architektura



TSIMMIS összefoglalás

- **Információ integráció heterogén forrásokra**
- **Kihívások és megoldások**
 - **Félig-struktúrák adatok & hiányzó séma információk**
 - **Megfelelő lekérdezés definiáló nyelv és lekérdezés lefordítás tervezése**
- **Eltérő képességű források kezelése**
 - **Lekérdezések speciális átalakítása**
 - **Képesség alapú átírás**