

Tudásalapú információ integráció

(A Szemantikus Web megközelítés és a
másik irány)

1

Miért van szükségünk ilyesmire? (Alkalmazások)

- **WWW:**
 - Keresés a weben (pl. összehasonlítás alapú vásárlás)
 - Portál építések több adatforrás felhasználásával
 - B2B, elektronikus piacterek
- **Tudomány és kultúra:**
 - Asztrofizika: égi jelenségek gyűjtése.
 - Kultúra: kulturális információs adatbázisok egységes elérése országhatárokon túl
 - Genetika: gén információk integrálása
- **Vállalati adatintegráció**
 - Egy átlagos KKV 49 adatbázist alkalmaz és IT költségvetésének 30%-át az adatintegrációra költi (US, 2002)

2

Információ integráció

Az információ integráció célja strukturált és félig-strukturált információforrások együttes kezelése, információigények egységes felületen történő kiszolgálása

Területek:

- Intézmények (közép- és nagyvállalatok, államigazgatás)
- Web

Megközelítési módok:

- Források kiegészítése (szemantikus web koncepció)
- Lekérdező eszközök fejlesztése (mediátor és információintegrációs rendszerek)

3

Szkeptikusoknak

Miért nem elég:

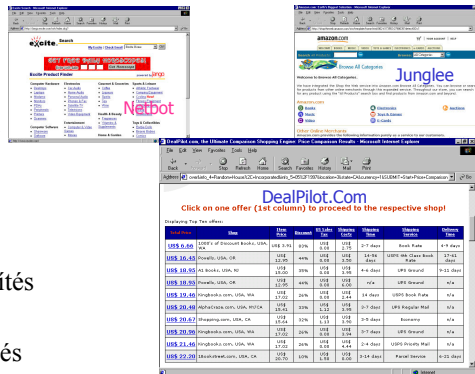


- Keresőgépek szövegalapú keresést végeznek
 - Jól működik egyedi dokumentumokon
 - Nem tudnak integrálni több dokumentumból származó információkat
 - Nem képesek hatékony általánosításra
 - Nem tudnak dokumentumokat és adatbázisokat összekapcsolni

4

Összehasonlítás alapú vásárlás?

- Hasonló források
- De:
 - Eltérő fókusz
 - Eltérő a spektrumuk az adatbázisoknak
 - Szolgáltatások
 - Új kihívás
 - Az “adattár” megközelítés nem működik
 - Kézi forrásleírás, kezelés korlátai

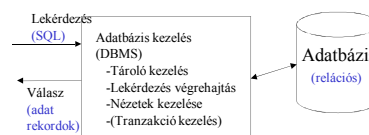


5

Miért nem csak

adatbázisok
elosztott adatbázisok

- **Közös séma hiánya**
 - Források heterogén sémákkal (és fogalmakkal, ontológiákkal) rendelkeznek
 - Félig-strukturált források
- **Régi források**
 - Nem relációs sémák
 - Eltérő elérési módok
- **Független források**
 - Nincs közös adminisztráció
 - Nem kezelt forrás tartalmi átfedések
- **Nehezen előre jelezhető viselkedés**
 - Lekérdezés végrehajtás bonyolult
- **Általában csak olvashatóak**
 - Ez lehet szerencsés is
 - Terjednek a tranzakció kezelési megoldások a weben



6

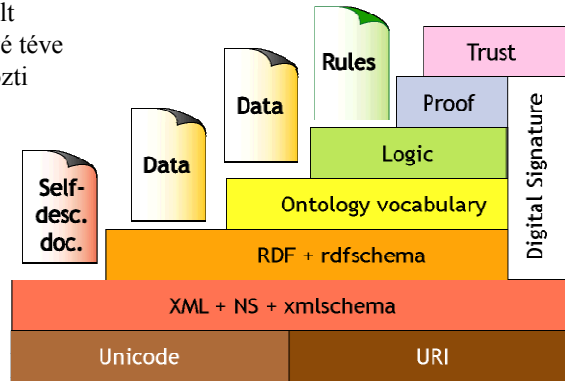
Az egyik irány

A szemantikus web koncepció

„A szemantikus web egy kiterjesztése a jelenlegi webnek, amelyben az információknak jól definiált jelentést adhatunk, lehetővé téve a gépek és felhasználók közötti jobb együttműködést.,,

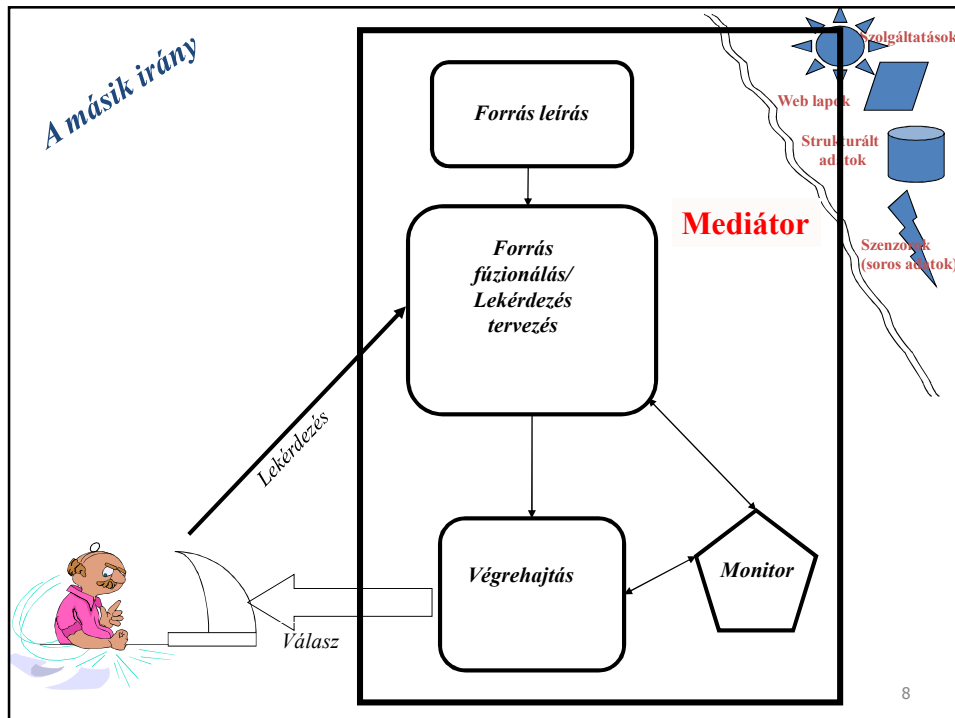
„A szemantikus web elérhetővé teszi a tudást, mint a web a hypertextet.”

Tim Berners-Lee

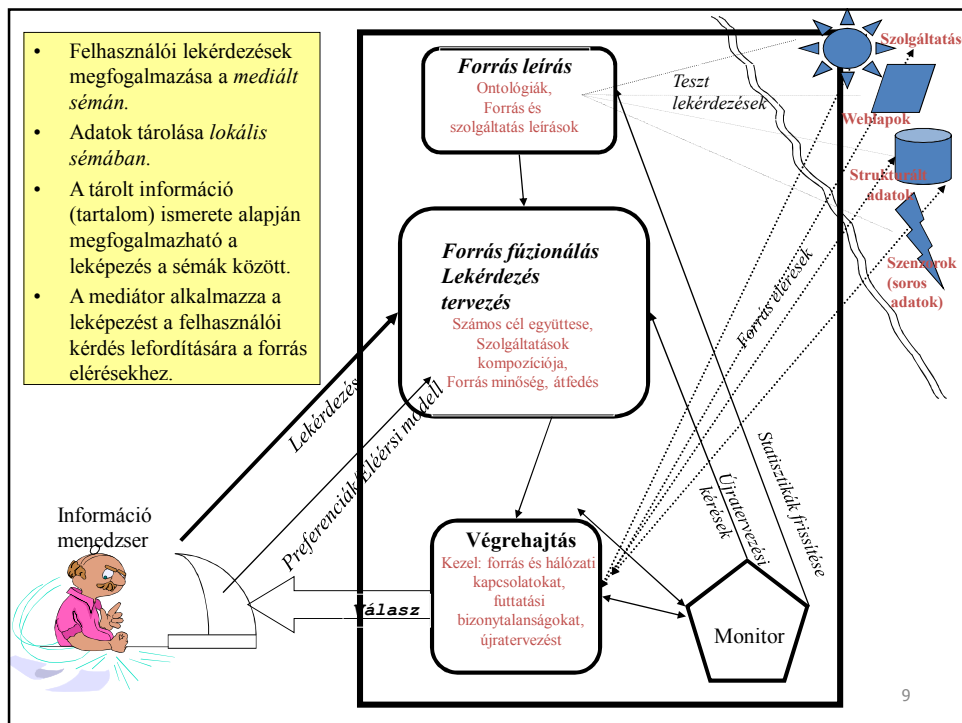


7

A másik irány

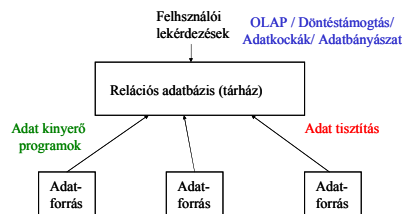
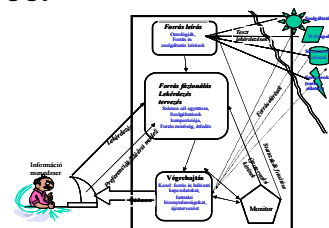


8



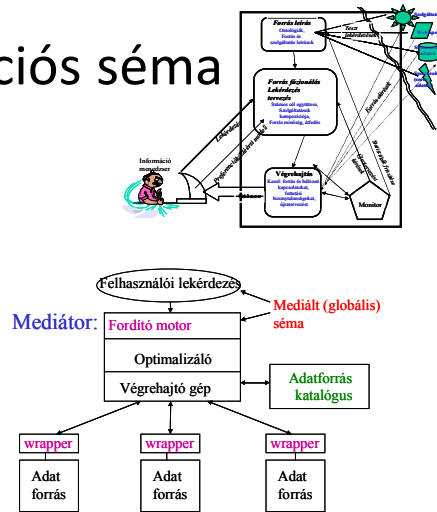
Kis forrás szám melletti integráció

- **Általában ad-hoc programozás:** speciális eset megvalósítása minden esetre, sok konzultáció.
- **Adattárházak:** minden adat periódikus feltöltése az adattárházba.
 - 6-18 hónap bevezetési idő
 - Operációs és döntéstámogatási RDBMS elválasztás. (nem csak adatintegrációra megoldás).
 - Teljesítmény jó,
 - adat lehet, hogy nem friss;
 - Rendszeres adattisztítás szükséges.



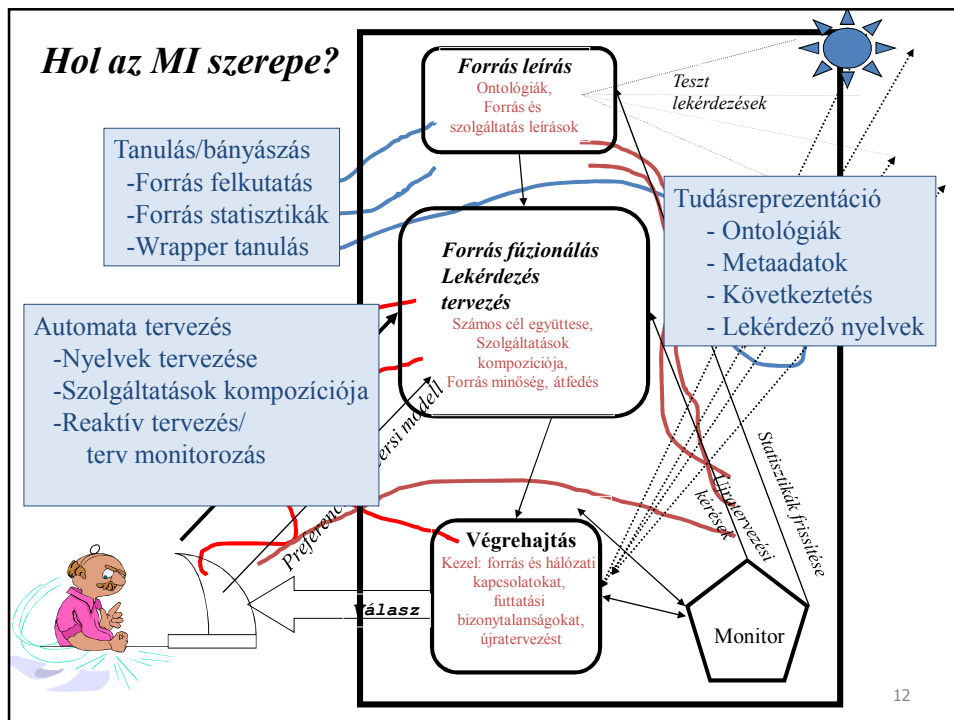
Virtuális integrációs séma

- 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



11

Hol az MI szerepe?



12