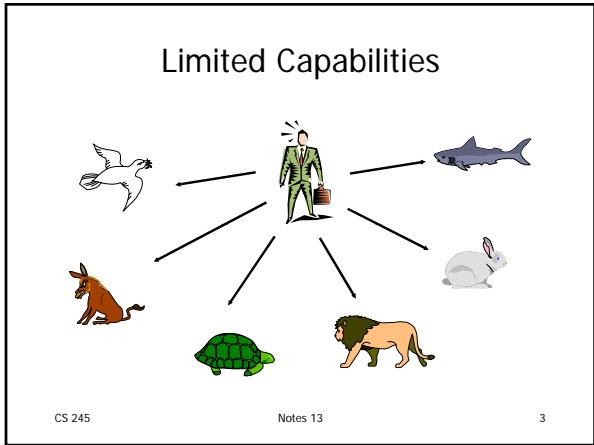
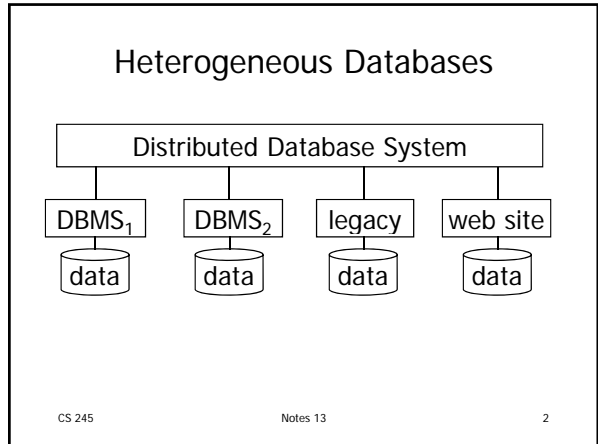


CS 245: Database System Principles

Notes 13: Coping with Limited Capabilities of Sources

Hector Garcia-Molina

CS 245 Notes 13 1



Example: Amazon.com

author:

title:

subject:

format:

price:

must specify at least one of these

this attribute not returned

menu of choices

cannot query on this attribute

CS 245 Notes 13 4

Example: BarnesAndNoble.com

author:

title:

subject:

format:

price:

must specify at least one of these

Menu of choices

can query if one of other attributes specified

CS 245 Notes 13 5

Why Limited Capabilities?

- Search forms
- Security
- Indexes
- Legacy

CS 245 Notes 13 6

Capability vs. Content

- Capability description
 - Can only search for subject = "art," "history," "science"
- Content description
 - Source only contains subject = "art," "history," "science"

CS 245 Notes 13 7

Outline

- Describing source capabilities
- Extending source capabilities
- How mediators cope with limited capabilities
- Mediator capabilities
- Other topics

CS 245 Notes 13 8

Describing Query Capabilities

R(X, Y, ... Z)

Adornments:

- **f**: may or may not specify
- **u**: cannot be specified
- **b**: must be specified
- **c[S]**: specified from list S
- **o[S]**: optional, chose from S

CS 245 Notes 13 9

Describing Query Capabilities

R(X, Y, ... Z)

Adornments:

- **f**: may or may not specify
- **u**: cannot be specified
- **b**: must be specified
- **c[S]**: specified from list S
- **o[S]**: optional, chose from S

With output restriction

- **f'**
- **u'**
- **b'**
- **c'[S]**
- **o'[S]**

CS 245 Notes 13 10

Example

- Relation R(X, Y, Z)
- Description Templates: bu'f, uf'c[z₁, z₂]
- Answerable queries: R(x₁, Y, Z), R(X, Y, z₁)
- Unanswerable queries:
R(X, y₁, Z), R(X, Y, z₃)

CS 245 Notes 13 11

Other Description Mechanisms

- Tsimmis
 - query templates
- Information Manifold
 - capability records (# bound attrs, conditions ok,...)
- Disco
- Garlic
 - black box
- Context-free grammars

CS 245 Notes 13 12

Extending Source Capabilities

Query: author="Freud" AND price > 10

Source: R(author, price, ...)
Template: b, u, ...

CS 245 Notes 13 13

Extending Source Capabilities

Query: author="Freud" AND price > 10

Wrapper Filter: price > 10

Source Query: author="Freud"

Source: R(author, price, ...)
Template: b, u, ...

CS 245 Notes 13 14

Another Example

Query: (author = "Freud" OR author = "Jung") AND price < 10

R(author, price, ...)
No disjunctive conditions;
Price can only be specified with author

CS 245 Notes 13 15

Another Example

Query: (author = "Freud" OR author = "Jung") AND price < 10

Union Operation

Q1: author = "Freud" AND price < 10
Q2: author = "Jung" AND price < 10

R(author, price, ...)
No disjunctive conditions;
Price can only be specified with author

CS 245 Notes 13 16

Extending Source Capabilities

- General scheme:
 - try many query rewritings
 - check if query fragments supported by source
 - check if wrapper can combine answer fragments
 - do all this very efficiently!! [See ICDE99 paper]
- Tsimmis, Info Manifold: no disjunctive queries
- DISCO: no query splitting
- Garlic: only CNF queries

CS 245 Notes 13 17

Mediator Processing

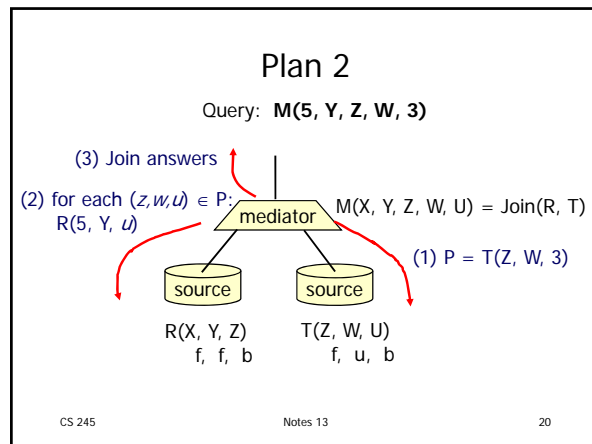
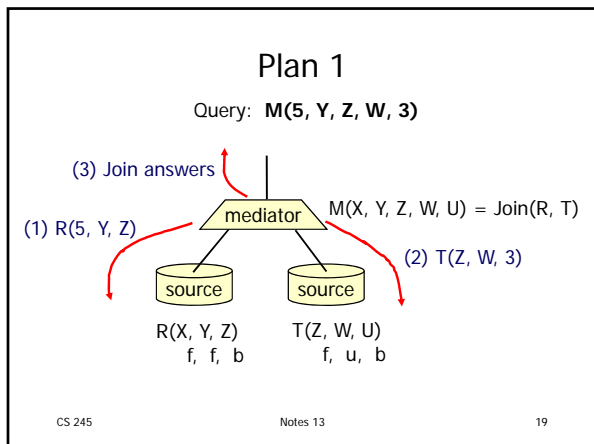
Query: M(5, Y, Z, W, 3)

M(X, Y, Z, W, U) = Join(R, T)

R(X, Y, Z)
f, f, b

T(Z, W, U)
f, u, b

CS 245 Notes 13 18



- ### Mediator Plan Generation
- Need feasible and efficient plan
 - Search space is huge
 - Tsimmis, Info Manifold, Garlic:
 - exponential algorithms
 - Polynomial algorithms:
 - often find optimal or near-optimal plan
 - bounded performance
 - [See ICDT99 Paper]
- CS 245 Notes 13 21

- ### Conclusion
- Not all sources are created equal!
 - Need to
 - describe what sources can do
 - efficiently process queries with limited sources
 - describe what mediators can do
 - exploit content information
 - deal with unavailable sources
- CS 245 Notes 13 22