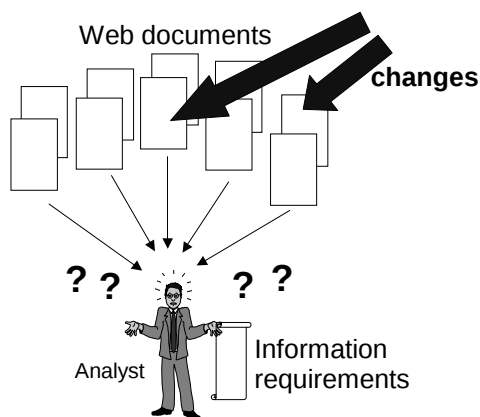


Managing Source Schema Evolution in Web Warehouses

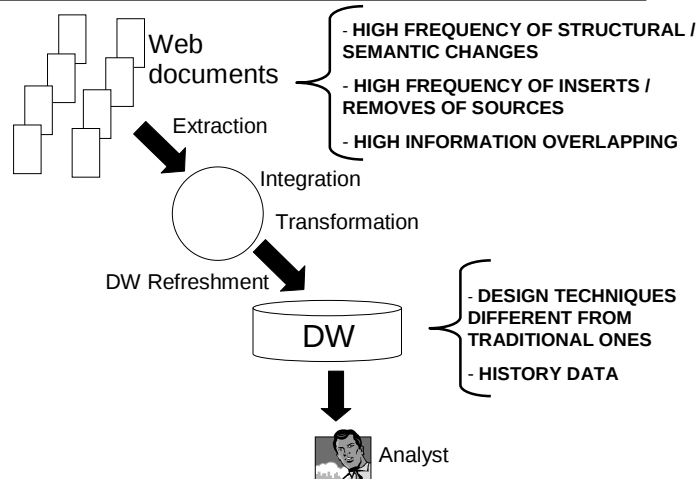
Adriana Marotta, Regina Motz, Raúl Ruggia

Instituto de Computación, Facultad de Ingeniería
Universidad de la República. Montevideo, Uruguay.
e-mail: [adriana,rmotz,ruggia]@fing.edu.uy

Motivation



The Problem



WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

3

Presentation Outline

- Main issues to address
- Our goal
- Proposed architecture
- Changes
- Evolution propagation
- Example cases
- Conclusion and future work

WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

4

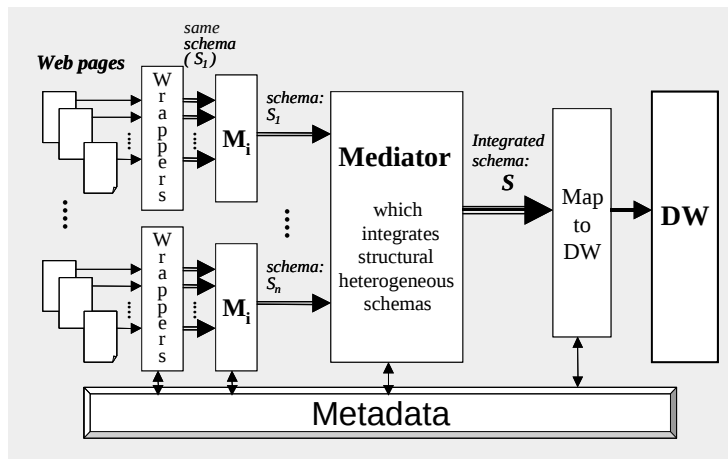
Main issues to address

- Extraction of data from Web sources and translation to a common model
- Reconciliation of information extracted from different Web sources
- DW design
- Maintenance of the consistency between DW and Web sources

Our goal

- Providing a framework where:
 - information extraction, information integration, DW design, consistency maintenance between DW and Web can be solved.
- Semi-automatic mechanism for propagating the changes of Web sources to the DW.

Proposed architecture

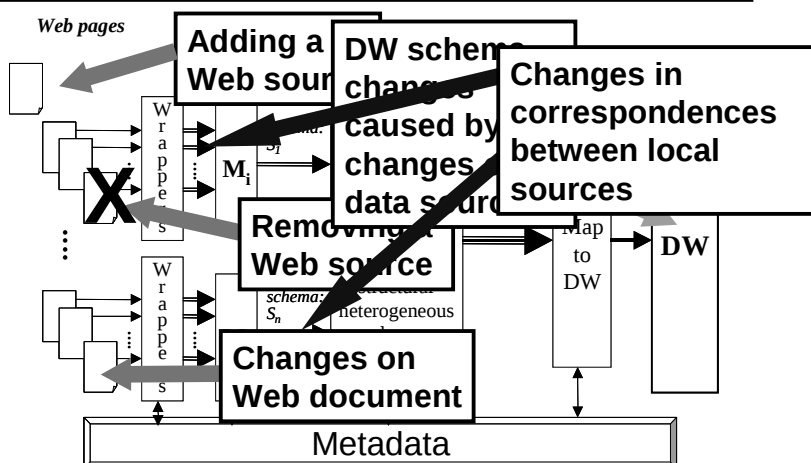


WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

7

Changes

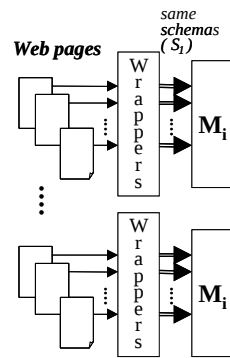


WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

8

Web to local schema propagation



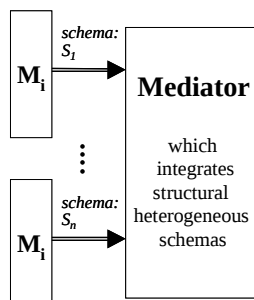
- Web source is added
- OR
- A web source structure is modified
- ↓
- Source schema is added to an existing M_i
 - Instance mapping is updated

WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

9

Local schema to integrated schema propagation



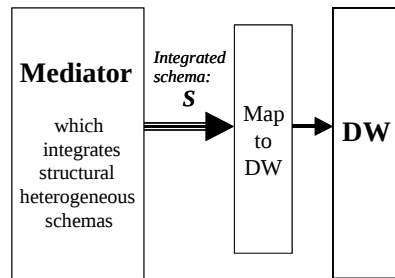
- Correspondences between the local sources changes
- ↓
- Incremental schema integration
 - Actualization of mappings from local-schema to integrated-schema

WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

10

Integrated schema to DW schema propagation



- Integrated schema changes



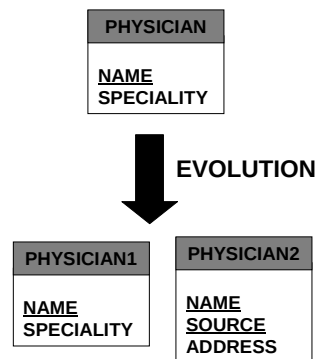
- Automatic propagation of changes to the DW schema based on Metadata

Propagating changes to the DW

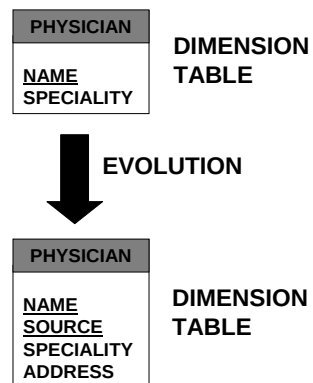
- Metadata-based approach:
 - DW design guidelines
 - Design trace
 - Distinguishing types of attributes and relations according to the Dimensional Model
 - Semantic properties of data
 - Information about Web changes (data availability)

Case 1

Integrated Schema



Data Warehouse Schema



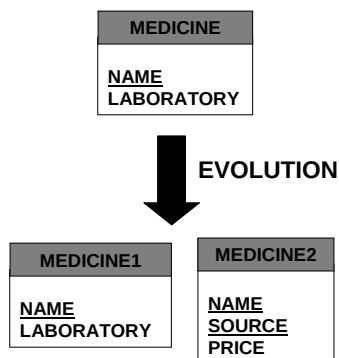
WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

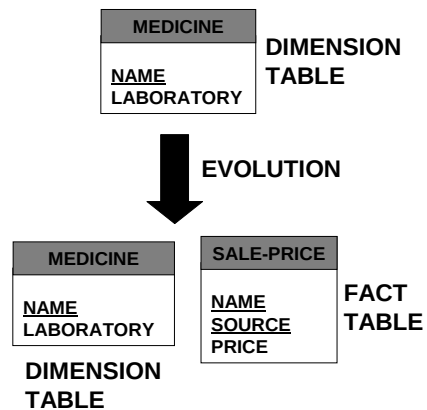
13

Case 2

Integrated Schema



Data Warehouse Schema

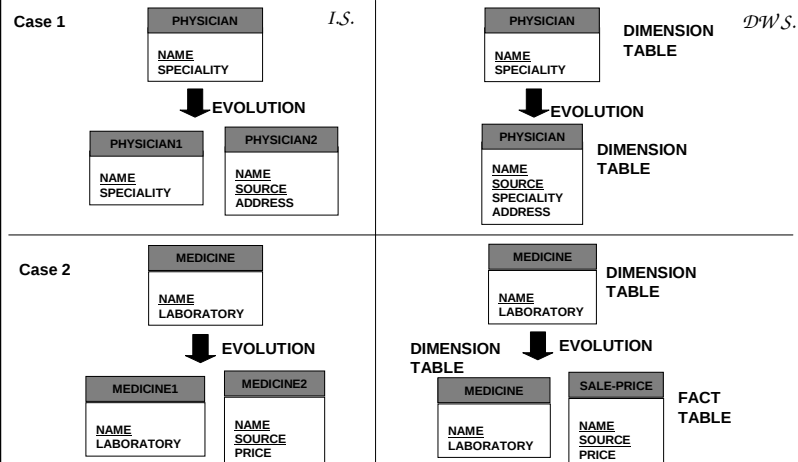


WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

14

Case 1 – Case 2

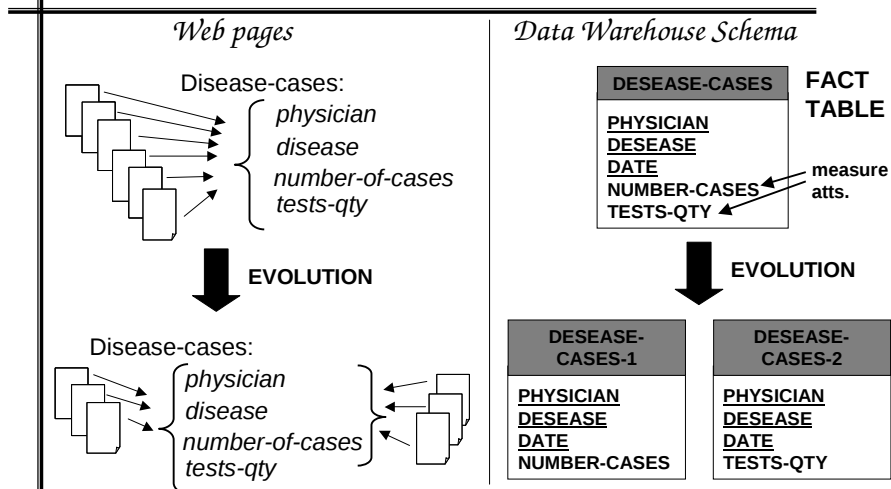


WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

15

Case 3



WIIW 2001

Marotta, Motz, Ruggia - Universidad de la Republica. Uruguay

16

Conclusion

- Framework for semi-automatic propagation of web source changes and addition/deletion of sources to an existing DW.
 - Specialized Mediators $\Rightarrow$ transparency in many cases
 - Metadata allows:
 - Automatic classification of attributes and relations according to Dimensional Model, based on the semantics of data.
 - DW design decisions based on the type of attributes and relations, and on information about the web changes.

Future work

- Refinement of Web evolution cases
- Propagation frequency
- DW loading and refreshment
- Assembling of existing prototypes