

# Semantic Web, Ontologies and Databases-An Overview

---

Eduardo Blanco

Luis Daniel Ibáñez

Amadís Martínez

Héctor Rodríguez

Elsa Tovar

María Esther Vidal

# Agenda

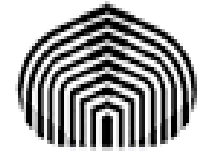
---

- Introduction
  - María Esther Vidal
- An Efficient Representation of RDF documents
  - Amadís Martínez
- Techniques to Compose Web Services
  - Eduardo Blanco
- Representing Active Knowledge in Ontologies
  - Elsa Tovar
- Authority Flow based Metrics to Rank Database objects.
  - Luis Daniel Ibáñez y Héctor Rodríguez



# Grupo de Bases de Datos

Departamento de Computación y Tecnología de la Información



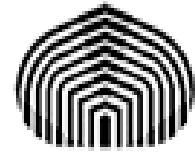
Universidad Simón Bolívar

- **People:**
  - Faculties: five
  - Teaching Assistants: five
  - PhD students: six
  - Master students: twelve
  - Undergraduates: six
- **Courses:**
  - Undergraduate: five
  - Graduate: five
- **Research projects:**
  - International Projects: four
  - Local Projects: five
- **Several projects** with the industry



# Grupo de **Bases de Datos**

Departamento de Computación y Tecnología de la Información



Universidad Simón Bolívar

Some research projects:

- Information Extraction
  - Soraya Abad
- Fuzzy Databases
  - Leonid Tineo
- Top-K and Skyline Techniques
  - Marlene Goncalves and María Esther Vidal
- Query Optimization and Evaluation techniques
  - Edna Ruckhaus and María Esther Vidal

# Some Projects on Query Optimization and Evaluation

---

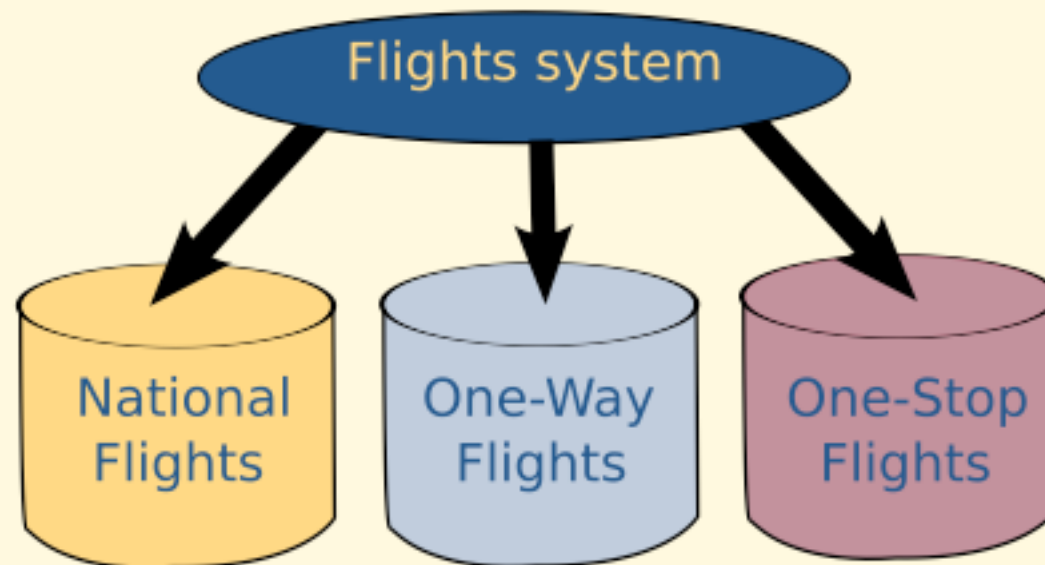
- Propositional Encoding of Query Rewriting Problems
  - Yolifé Arvelo and Blai Bonet and María Esther Vidal [Arvelo & Bonet & Vidal 2006]
- Query Optimization
  - Ontology conjunctive queries and SPARQL queries
- Techniques to store and query RDF efficiently
- Extensions to ontology languages to represent active knowledge
- Cost-based Web Service Composition Techniques
- Ranking Techniques
  - Approximate solutions

# Compilation of Query-rewriting Problems into Tractable Fragments of Propositional Logic- [Arvelo & Bonet & Vidal AAAI06]

---

- We consider the problem of query rewriting using materialized views
- This problem appears frequently in the context of Data Integration, Web Infrastructures and Query Optimization.
- This problem is in general intractable and existing algorithms do not scale

**QUERY:** Find round-trip flights that start in the US



Sources are independent, described as views, incomplete

**QUERY:** Find round-trip flights that start in the US

$$Q(x, y) \text{ :- flight}(x, y), \text{ flight}(y, x), \text{ uscity}(x)$$

Data sources modelled as views:

$$\text{national}(x_1, y_1) \text{ :- flight}(x_1, y_1), \text{ uscity}(x_1), \text{ uscity}(y_1)$$
$$\text{oneway}(x_2, y_2) \text{ :- flight}(x_2, y_2)$$
$$\text{onestop}(x_3, z_3) \text{ :- flight}(x_3, y_3), \text{ flight}(y_3, z_3)$$

- **ASSUMPTION:** Views may be incomplete

- Then, the solution is the **collection** of rewritings:

$$R_1(x, y) \text{ :- oneway}(x, y), \text{ oneway}(y, x), \text{ national}(x, w)$$

$$R_2(x, y) \text{ :- oneway}(x, y), \text{ oneway}(y, x), \text{ national}(w, x)$$

$$R_3(x, y) \text{ :- national}(x, y), \text{ national}(y, x)$$

$$R_4(x, y) \text{ :- oneway}(x, y), \text{ national}(y, x)$$

$$R_5(x, y) \text{ :- national}(x, y), \text{ oneway}(y, x)$$

- Observe that there is no rewriting using  $\text{onestop}(x, y)$

# Existing Algorithms

---

- Bucket Algorithm [Levy & Rajaraman & Ullman 1996]
- Inverse rules Algorithm [Duscka & Genesereth 1997]
- MinCom [Pottinger & Halevy 2001]

- Exploit independences to decompose into smaller subproblems and then combine solutions
- Solutions to subproblems are called MCDs

MCD	View	Mapping	Covered subgoals
$M_1$	national	$\{X \rightarrow X_1, Y \rightarrow Y_1\}$	$\{0\}$
$M_2$	national	$\{X \rightarrow Y_1, Y \rightarrow X_1\}$	$\{1\}$
$M_3$	national	$\{X \rightarrow X_1\}$	$\{2\}$
$M_4$	national	$\{X \rightarrow Y_1\}$	$\{2\}$
$M_5$	oneway	$\{X \rightarrow X_2, Y \rightarrow Y_2\}$	$\{0\}$
$M_6$	oneway	$\{X \rightarrow Y_2, Y \rightarrow X_2\}$	$\{1\}$

# MiniCom

---

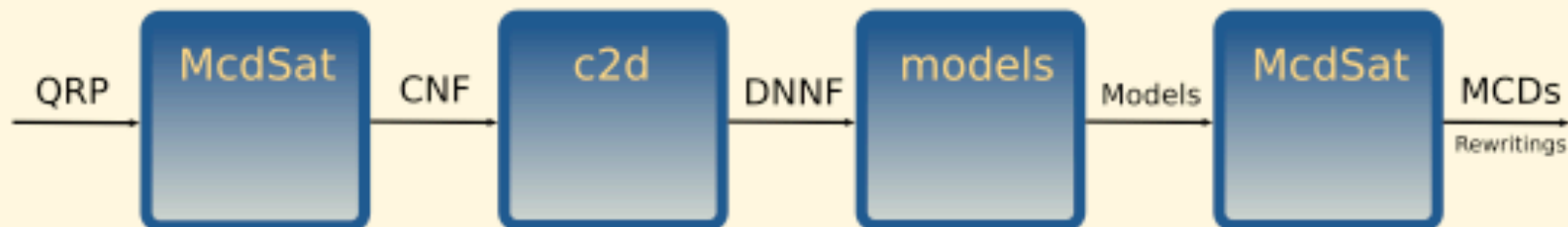
- Generate all MCDs (performs blind search)
- Rewritings generated greedily as combination of MCDs
  - Cover disjoint subsets of subgoals in the query
  - Cover all subgoals in the query

# MCDSat Approach

---

- Given a query  $Q$  and a set of views  $V$
- Build a propositional theory such that its models correspond to the MCDs
- Generating MCDs is a problem of model enumeration
- The theory was extended to produce query rewritings

- MCDSAT translates QRP into a propositional theory  $T$
- $T$  is compiled into d-DNNF using Darwiche's `c2d` compiler
- Models are **obtained** from the d-DNNF and transformed into MCDs or rewritings



- `c2d` and `models` are off-the-shelf components
- MCDSAT written in scripting language

