Querying from a Graph Database Perspective: the case of RDF

Renzo Angles and Claudio Gutierrez

Department of Computer Science

Universidad de Chile

{rangles, cgutierr}@dcc.uchile.cl

Agenda

- Motivations, Problem and Proposal
- Database Models
- Graph Database Models
- RDF as data model
- Querying RDF data
- Graph Query Languages

  versus RDF Query Languages

Motivations

- Increasing amount of information with graph structure (Web, genome, complex nets, etc).
- Database models and query languages should support its underlying structure.
- RDF is a proposal of data model with inherent graph structure.

Problem

- Current RDF query languages do not ***support*** common graph queries in current applications.
- Example:

  What is the Erdös distance between author X and author Y?

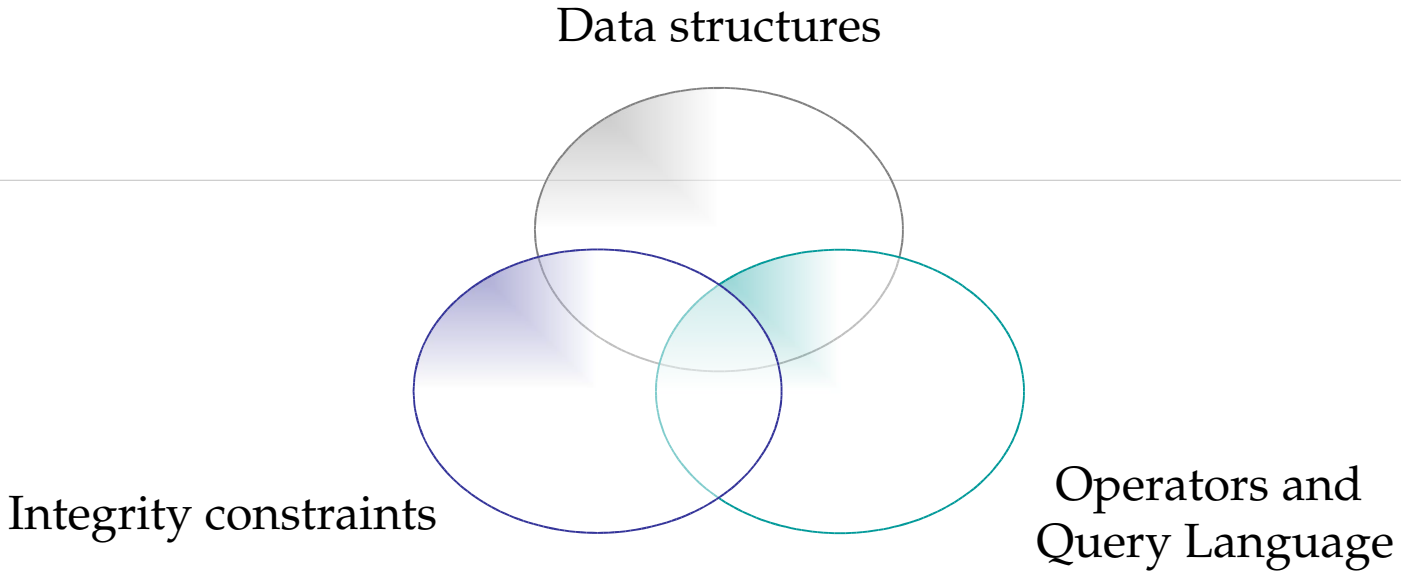
  This expression represents a query about the distance between nodes in a collaboration graph

A language is said to ***support*** a feature if it provides facilities that make it convenient (reasonable easy, safe and efficient) to use the feature [Stroustrup-1988].

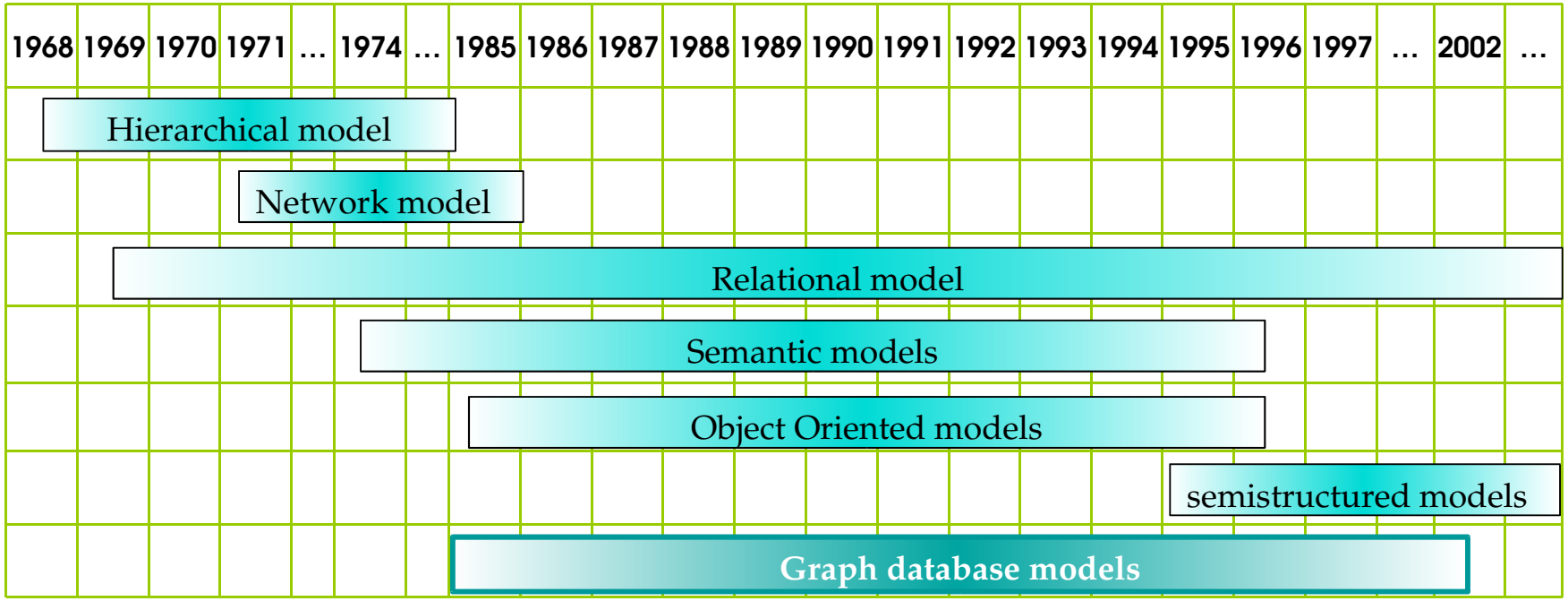
Database Models

A data model is a collection of three components [Codd-1980] :

2) A collection of data structure types
3) A collection of operators or inferencing rules
4) A collection of general integrity rules.



Database Models chronology / Time-line

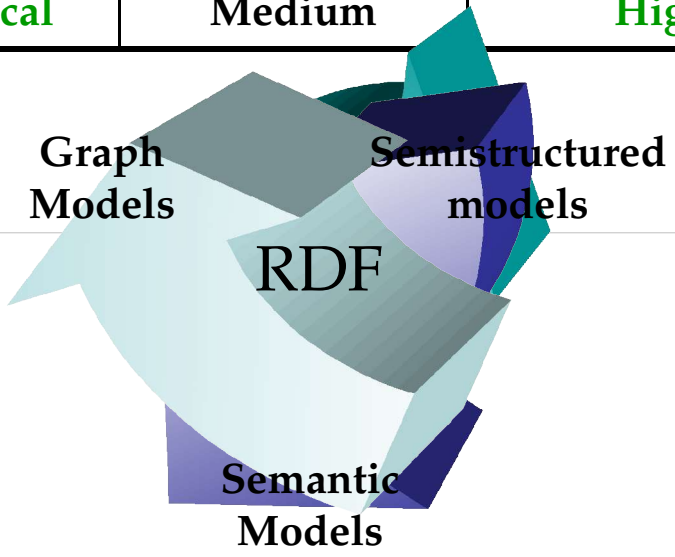


Graph Database models

- Model information whose logical structure is a graph.
- Emphasize the interconnectivity between the data.
- Appeared in the mid-eighties, flourished in the mid-nineties and unexpectedly disappeared with the emergence of semi-structured models.
- Some current application areas: Web, biology (genomic), chemistry (chemical reactions), geographic information (cities, roads, rivers), large linked networks (police, airways), complex nets.

Data models comparison

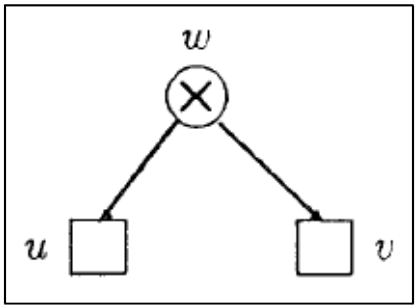
MODEL	LEVEL	COMPLEXITY OF DATA	CONNECTIVITY	TYPE OF DATA
Semantic	User	Simple / medium	High	Homogeneous
Semistructured	Logical	Medium	Medium	Heterogeneous
Graph	Logical	Medium	High	Heterogeneous
Semistructured	Logical	Medium	Medium	Heterogeneous
Graph	Logical	Medium	High	Heterogeneous

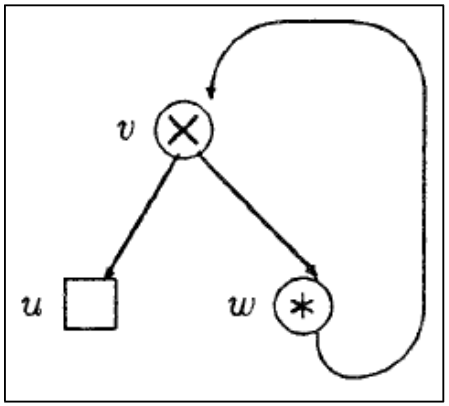


Graph Database models (cont)

Logical Data Model

G.M. Kuper and M.Y. Vardi (1984)





Schemas

$I(u)$

l	$val(l)$
1	Rehoboam
2	Solomon
3	David

$I(v)$

l	$val(l)$
4	Solomon
5	David
6	Batsheba
7	Jesse

$I(w)$

l	$val(l)$
8	(1, 4)
9	(2, 5)
10	(2, 6)
11	(3, 7)

$I(u)$

l	$val(l)$
1	Rehoboam
2	Solomon
3	David
4	Batsheba
5	Jesse

$I(v)$

l	$val(l)$
6	(1, 11)
7	(2, 12)
8	(3, 13)
9	(4, 14)
10	(5, 14)

$I(w)$

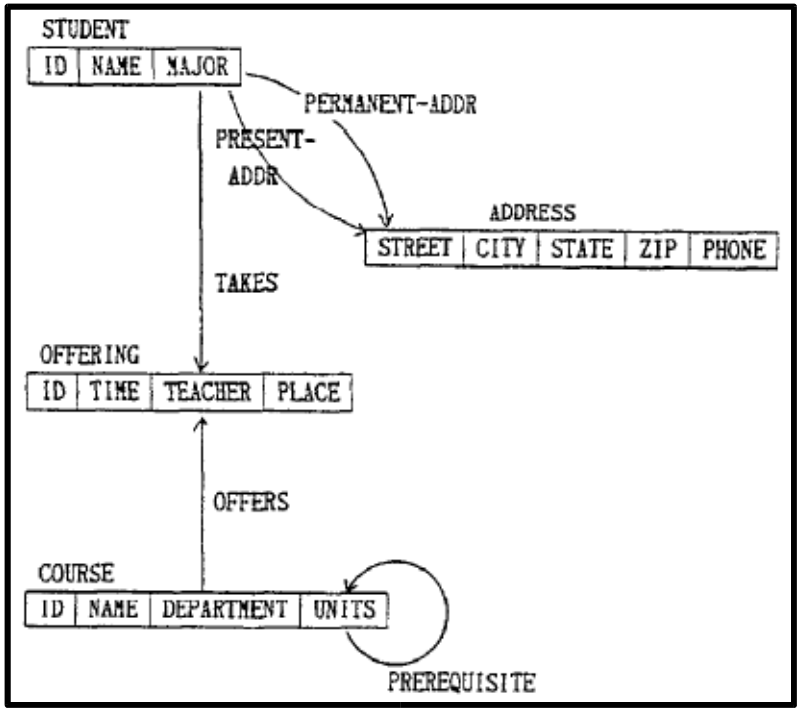
l	$val(l)$
11	{7}
12	{8, 9}
13	{10}
14	$\emptyset$

Instances

Graph Database models (cont)

G-Base

H. S. Kunii (1987)

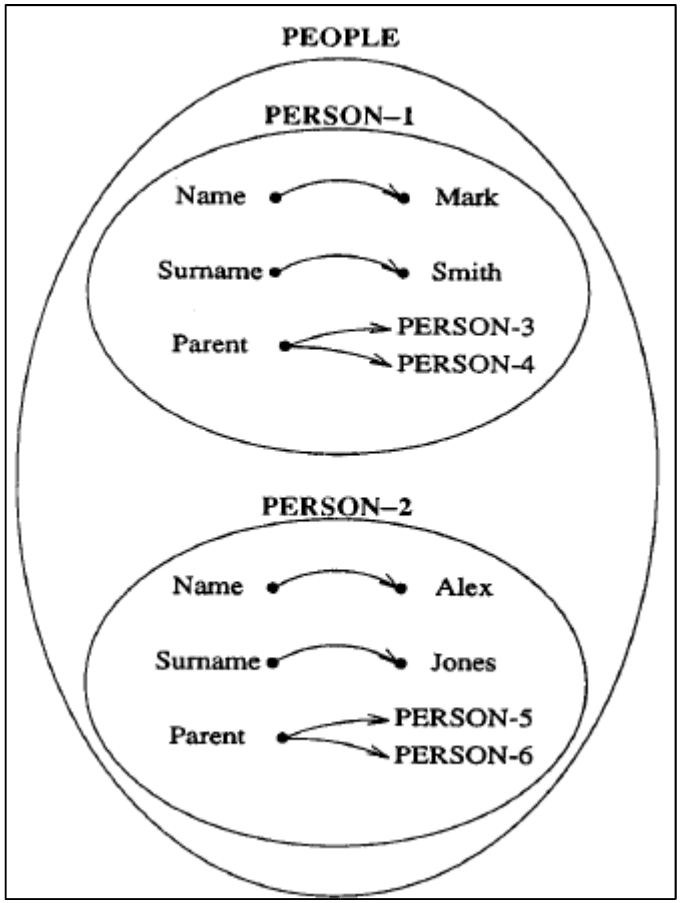


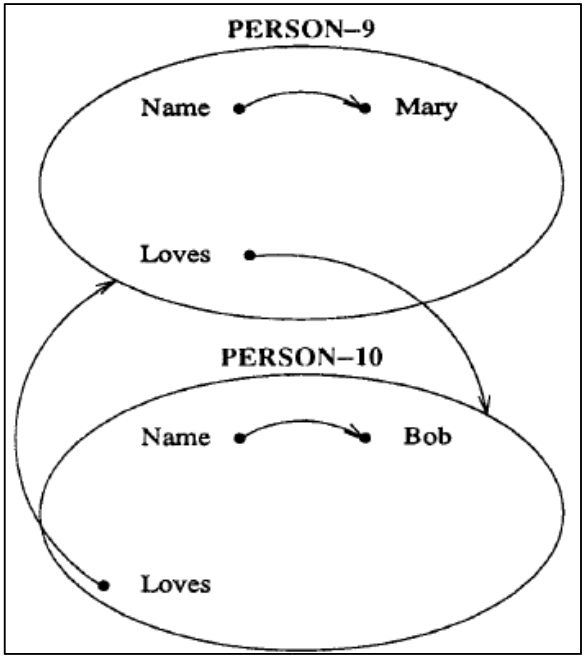
Schema

Graph Database models (cont)

Hypernode

M. Levene and A. Poulovassilis (1990)



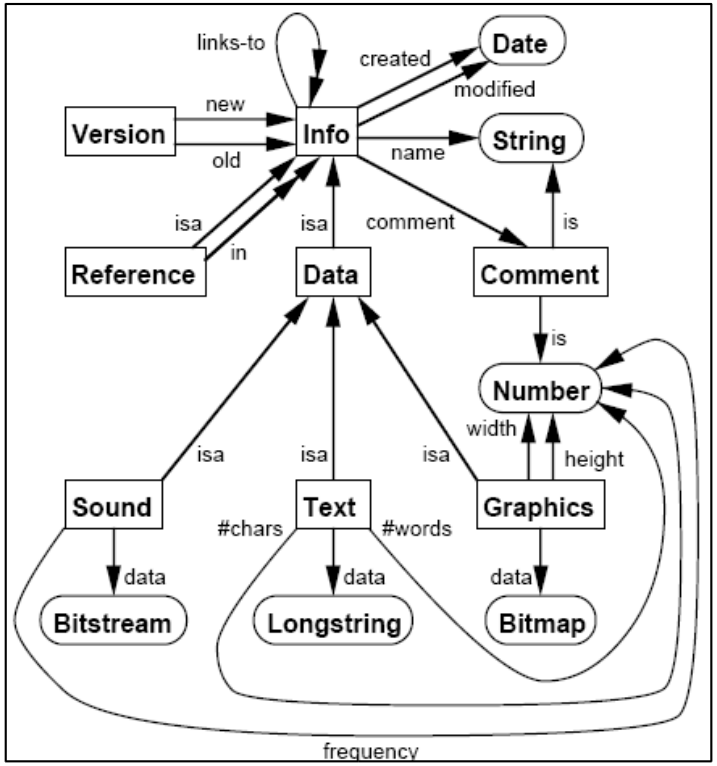


Instances

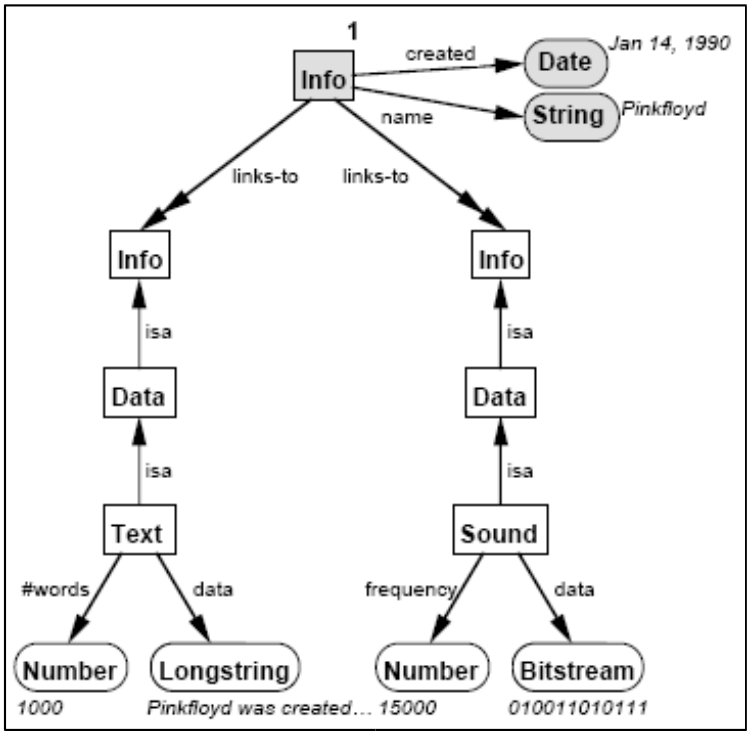
Graph Database models (cont)

GOOD

M. Gyssens, J. Paredaens, J. V. Den Bussche (1990)



Schema



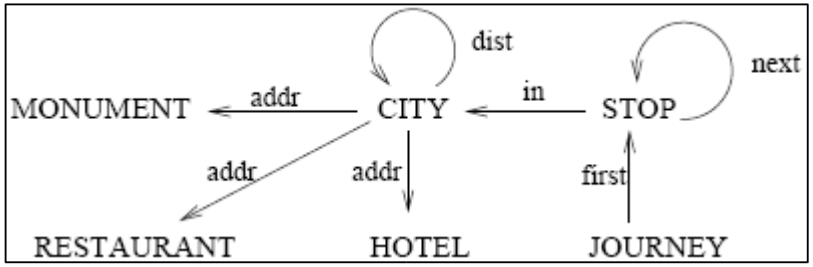
Instance

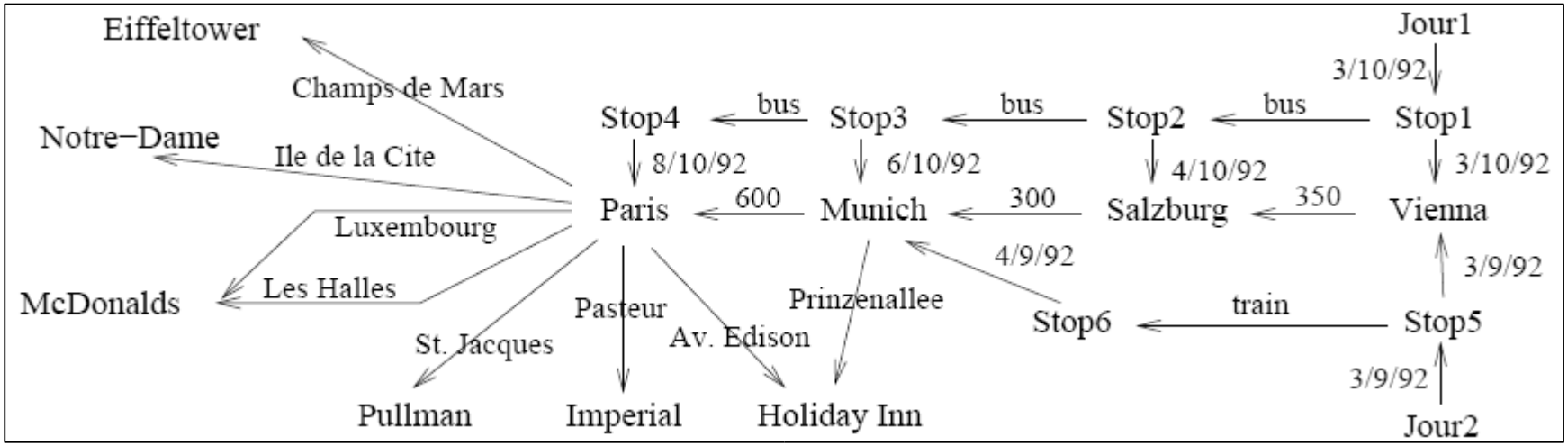
Graph Database models (cont)

Gram

B. Amann and M. Scholl (1992)

Schema



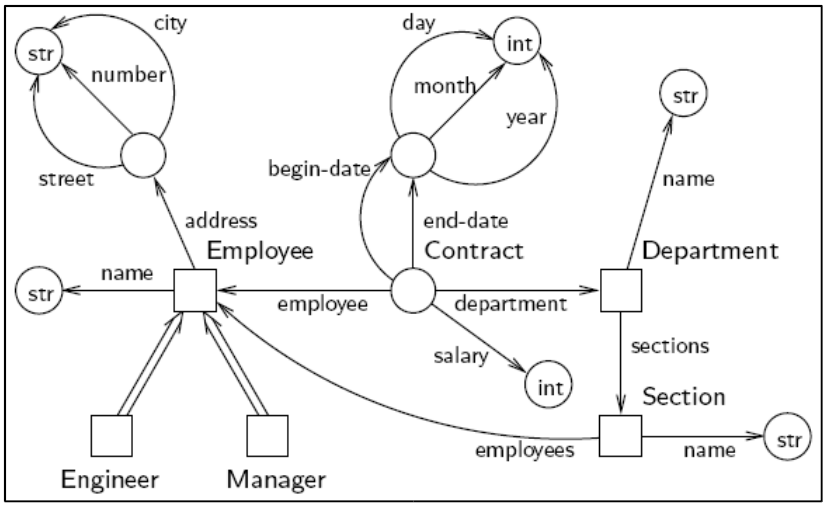


Instance

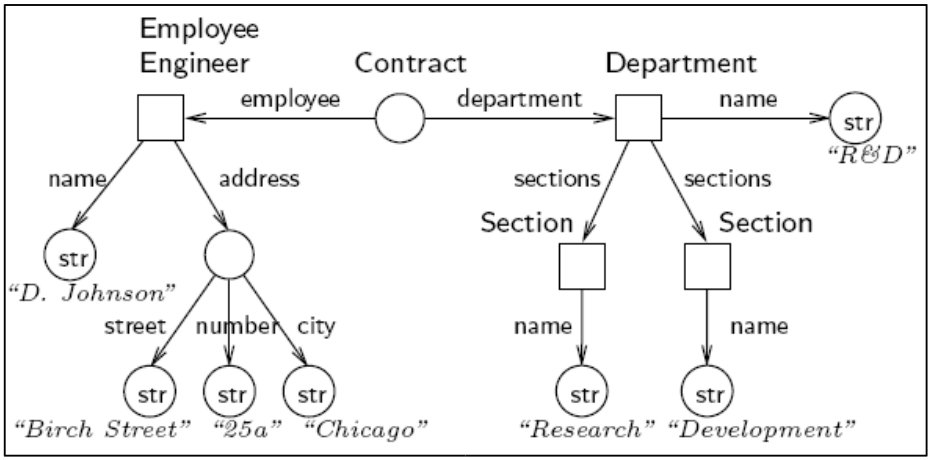
Graph Database models (cont)

GDM

J. Hidders (2002)



Schema

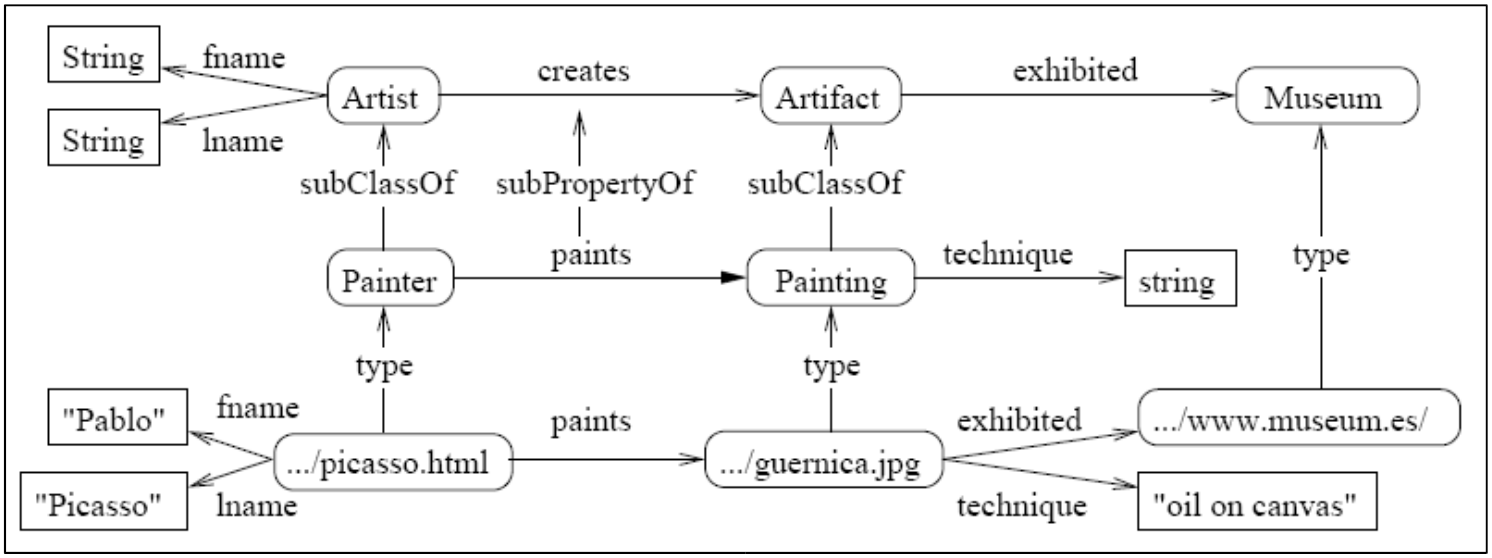


Instance

Graph Database models (cont?)

RDF

W3C (1999)



Schema and Instance

RDF from a database perspective

1) ***Data Structure Types*:** RDF establishes a syntax that represents structures to describe resources and their relations.
2) ***Integrity constraints:*** the RDF specification does not present clear notions of constraints.
3) ***Operators or Inference rules:*** there are proposals of query languages for RDF, but they do not support some frequent queries in current graph applications.

Querying RDF data (cont)

Graph notion	Real-life Application Query
Paths	**Are suspects A and B related?** (relevant paths in a police database)
	What is the shortest route between city A and city B? (Shortest path in a database of roads)
	What is the influence of article D? (transitive closure in database of bibliographic cites)
Distance	**What is the Erdös distance between author X and author Y?** (distance between nodes in a collaboration graph)
Pattern Matching	**Where and how much a motif (pattern) appears?** (Pattern matching in genome data)

Querying RDF data

Graph notion	Real-life Application Query
Adjacency	**All relatives of degree one of Alice** (adjacent nodes in a genealogy database)
	What chemical composes does a given chemical reaction produce? (Adjacent edges in chemical information)
	What cities are near Athens? (neighborhood in a tourism system graph)
Degree of a node	**What is/are the most cited paper/s?** (searching node/s with maximum in-degree in a database of bibliographic cites)

Support of current RDF Query Languages

- None of the current query languages for RDF *support* the previous queries.
- Graph properties like diameter, radius or eccentricity are not supported.
- None of the languages support arbitrary paths.
- Some languages require union to answer simple queries (e.g. Adjacency, degree).
- Aggregate operators are important and some languages do not provide these operators.

PROPERTY		Adjacent Nodes	Adjacent Edges	Degree of a Node	Path	Fixed-length path	Distance	Diameter
RDF Query Language	RQL							
	SeRQL							
	RDQL							
	Triple							
	N3							
	Versa							
	RxPath							
Graph Query Language	G							
	G+							
	Graph Log							
	Gram							
	Graph DB							
	Lorel							
	F-G							

PROPERTY		Adjacent Nodes	Adjacent Edges	Degree of a Node	Path	Fixed-length path	Distance	Diameter
RDF Query Language	RQL							
	SeRQL							
	RDQL							
Graph Query Language								
	Gram							
	Graph DB							
	Lorel							
	F-G							

What is the Erdös distance between author X and author Y?
(distance between nodes in a collaboration graph)

PROPERTY		Adjacent Nodes	Adjacent Edges	Degree of a Node	Path	Fixed-length path	Distance	Diameter
RDF Query Language	RQL							
	SeRQL							
	RDQL							
	Triple							
	N3							
	Versa							
	RxPath							
Graph Query Language	G							
	G+							
	Graph Log							
	Gram							
	Graph DB							
	Lorel							
	F-G							

All relatives of degree one of Alice
(adjacent nodes in a genealogy database)

What chemical composes does a given chemical reaction produce?
(adjacent edges in chemical information)

PROPERTY		Adjacent Nodes	Adjacent Edges	Degree of a Node	Path	Fixed-length path	Distance	Diameter
RDF Query Language	RQL							
	SeRQL							
	RDQL							
	Triple							
	N3							
	Versa							
	RxPath							
Graph Query Language	G							
	G+							
	Graph Log							
	Gram							
	Graph DB							
	Lorel							
	F-G							

What is/are the most cited paper/s?
(search node/s with maximum in-degree in a database of bibliographic cites)

Thanks for your attention

Questions? Comments?

Renzo Angles (rangles@dcc.uchile.cl)

Claudio Gutierrez (cgutierr@dcc.uchile.cl)

Department of Computer Science

Universidad de Chile

References

[Stroustrup-1988] Stroustrup B.: What is Object-Oriented Programming? IEEE Softw. 5 (1988) 10 -20.

[Codd-1980] Codd, E.F.: Data Models in Database Management. In: Proc. Of the workshop on Data abstraction, databases and conceptual modeling, ACM Press (1980) 112-114.