Urban Data Consistency in RDF: A Case Study of Curitiba Transportation System

Mirian Halfeld-Ferrari
Université d'Orléans - France

Carmem S. Hara
Federal University of Paraná (UFPR) - Brazil

Nádia P. Kozievitch
Federal University of Technology (UTFPR) - Brazil

Flavio R. Uber
State University of Maringá (UEM)/Federal University of Paraná (UFPR) - Brazil

LADaS/VLDB - August 2018

Preliminaries

- Urban Mobility
  - Heterogeneous data sources:
    traffic, public transportation, security
- Trend towards freely available data
  - Solution: Linked Data / RDF
- RDF consistency
- BNS System

Curitiba Transportation System

Tube

Constraint: every express line stop must be a fast boarding stop

Terminal

Curitiba Transportation System

Tube

Constraint: every express line stop must be a fast boarding stop

Terminal

s1 is a street stop

s1

Curitiba Transportation System

Tube

Constraint: every express line stop must be a fast boarding stop

Terminal

s1 is a street stop

s1

- Update: “s1 is a new stop for express line 302”

Curitiba Transportation System

Tube

Constraint: every express line stop must be a fast boarding stop

Terminal

s1 is a street stop

s1

- Update: “s1 is a new stop for express line 302”
- Update must be rejected OR
  s1 is not in the class of street stop

BNS System

- Detects RDF constraint violations
- Either refuses the update, or computes a set of compensation actions (side effects) in order to guarantee constraints satisfaction
- BNS can be applied on a urban setting and supports updates in Curitiba Transportation System

Summary

Summary

Data Importing

code character varying(25)	name character varying(100)	category character varying(100)	color character varying(
Z01	UNIBRASIL / TUBO DET	CONVENCIONAL	AMARELA
203	STA. CÂNDIDA / C. RA	EXPRESSO	VERMELHA
302	CENTENÁRIO / RUI BAR	EXPRESSO	VERMELHA
303	CENTENÁRIO / C. COMP	EXPRESSO	VERMELHA

URBS database: Route Table

line_code character v	address character varying(254)	num integer	lat double precisio
302	Estação Tubo Antônio Meirelles	108149	25.44701536525
302	Estação Tubo Urbano Lopes	108157	25.44080042455
302	Terminal Oficinas - 303 - Cente	105601	-25.451172
302	Estação Tubo Teófilo Otoni	108147	25.45731211574
302	Estação Tubo Del. Amazor Preste	108156	-25.4417096055
302	Estação Tubo Antônio Meirelles	108150	25.44697059844
302	Estação Tubo Cajuru	108146	-25.460427198
302	Estação Tubo Hospital Cajuru	108162	25.43702795934

URBS database: Stop Table

Data Importing

code character varying(25)	name character varying(100)	category character varying(100)	color character varying('
Z01	UNIBRASIL / TUBO DET	CONVENCIONAL	AMARELA
203	STA. CÂNDIDA / C. RA	EXPRESSO	VERMELHA
302	CENTENÁRIO / RUI BA	EXPRESSO	VERMELHA
303	CENTENÁRIO / C. COMP	EXPRESSO	VERMELHA

URBS database: Route Table

line_code character v	address character varying(254)	num integer	lat double precisio
302	Estação Tubo Antônio Meirelles	108149	25.44701536525
302	Estação Tubo Urbano Lopes	108157	25.44080042455
302	Terminal Oficinas - 303 - Cente	105601	-25.451172
302	Estação Tubo Teófilo Otoni	108147	25.45731211574
302	Estação Tubo Del. Amazor Preste	108156	-25.4417096055
302	Estação Tubo Antônio Meirelles	108150	25.44697059844
302	Estação Tubo Cajuru	108146	-25.460427198
302	Estação Tubo Hospital Cajuru	108162	25.43702795934

URBS database: Stop Table

Data Importing

code character varying(25)	name character varying(100)	category character varying(100)	color character varying(
Z01	UNIBRASIL / TUBO DET	CONV	
203	STA. CÂNDIDA / C. RA	EXPR	
302	CENTENÁRIO / RUI BAR	EXPR	
303	CENTENÁRIO / C. COMP	EXPR	

URBS database: R

subject character varying(200)	predicate character varying(200)	object character varying(200)
302	ExpressLineStop	Estação Tubo Praça Eufrasio Correia
302	ExpressLineStop	Estação Tubo Germânia
302	ExpressLineStop	Estação Tubo Profª. Maria Aguiar Teixeir
302	ExpressLineStop	Estação Tubo Profª. Maria Aguiar Teixeir
302	ExpressLineStop	Estação Tubo Praça Eufrasio Correia
302	ExpressLineStop	Estação Tubo Jardim Botânico
302	ExpressLineStop	Terminal Centenário - 303 - Centenário /
302	ExpressLineStop	Estação Tubo Profª. Maria Aguiar Teixeir
302	ExpressLineStop	Estação Tubo Catulo da P. Cearense
302	ExpressLineStop	Estação Tubo Cajuru

RDF triples

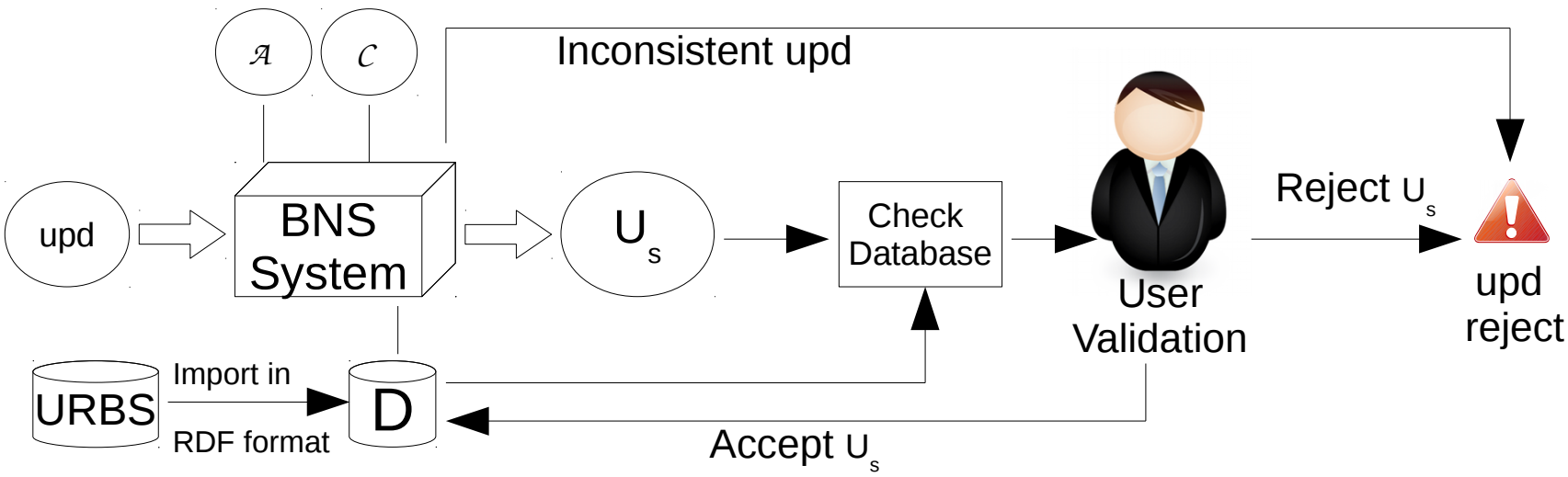
(busline,"ExpressLineStop",stop)

line_code character v	address character varying(254)		
302	Estação Tubo Antônio Meirelles	108149	25.44701536525
302	Estação Tubo Urbano Lopes	108157	25.44080042455
302	Terminal Oficinas - 303 - Cente	105601	-25.451172
302	Estação Tubo Teófilo Otoni	108147	25.45731211574
302	Estação Tubo Del. Amazor Preste	108156	-25.4417096055
302	Estação Tubo Antônio Meirelles	108150	25.44697059844
302	Estação Tubo Cajuru	108146	-25.460427198
302	Estação Tubo Hospital Cajuru	108162	25.43702795934

URBS database: Stop Table

Summary

Process Overview

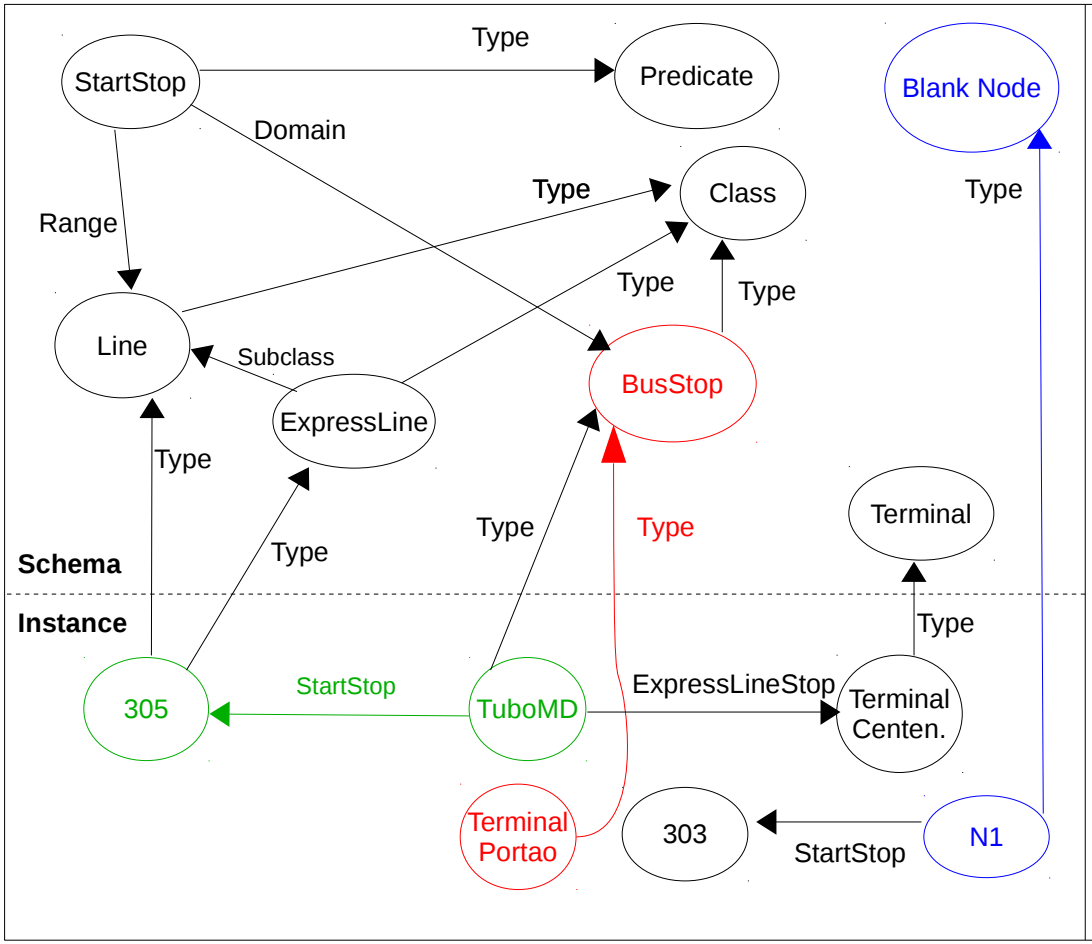


Summary

RDF - Triples

	RDFS Triples	Ground Facts
Schema	(A,rdf:type,rdfs:Class)	Cl(A)
	(A,rdf:type,rdf:Property)	Pr(A)
	(A,rdfs:subClassOf,rdfs:Resource)	Csub(A, rdfs:Resource)
	(A,rdfs:subClassOf,B)	Csub(A,B)
	(A,rdfs:subPropertyOf,B)	Psub(A,B)
	(A,rdfs:domain,rdfs:Resource)	Dom(A,rdfs:Resource)
	(A,rdfs:domain,B)	Dom(A,B)
	(A,rdfs:range,rdfs:Resource)	Rng(A,rdfs:Resource)
	(A,rdfs:range,rdfs:Literal)	Rng(A,rdfs:Literal)
	(A,rdfs:range,B)	Rng(A,B)
Instance	(A,rdf:type,rdfs:Resource)	Ind(A)
	(A,rdf:type,B)	CI(A,B)
	(A,C,B)	PI(A,B,C)

RDF - Graph



Csub(ExpressLine, Line)
CI(TerminalCenten, Terminal)
CI(TerminalPortao, BusStop)
CI(TuboMD, BusStop)
CI(395,ExpressLine)
PI(305, Tubo MD, StartStop)
PI(N1, 303, StartStop)
Pr(StartStop)
Dom(StartStop, BusStop)
Rng(StartStop, Line)
BN(N1)
Cl(Line)
Cl(ExpressLine)

Application Constraints

- Defined on instances $CI(X, Y)$ and properties $PI(X, Y, Z)$
- No negative atom in the body of the rule
- Format and restrictions avoid *blank nodes* propagation
- Types:
  - Type 1:

    $CI(X_1, c_1) \rightarrow CI(X_1, c_2)$ $\quad$ $PI(X_1, X_2, p_1) \rightarrow PI(X_1, X_2, p_2)$

    $CI(X_1, c_1) \rightarrow \neg CI(X_1, c_2)$ $\quad$ $PI(X_1, X_2, p_1) \rightarrow \neg PI(X_1, X_2, p_2)$

  - Type 2:

    $CI(X_1, c_1) \rightarrow PI(X_1, X_2, p_1)$ $\quad$ $CI(X_1, c_1) \rightarrow \neg PI(X_1, X_2, p_1)$

    $CI(X_2, c_1) \rightarrow PI(X_1, X_2, p_1)$ $\quad$ $CI(X_2, c_1) \rightarrow \neg PI(X_1, X_2, p_1)$

  - Type 3:

    $PI(X_1, X_2, p_1) \rightarrow CI(X_1, c_1)$ $\quad$ $PI(X_1, X_2, p_1) \rightarrow \neg CI(X_1, c_1)$

    $PI(X_1, X_2, p_1) \rightarrow CI(X_2, c_1)$ $\quad$ $PI(X_1, X_2, p_1) \rightarrow \neg CI(X_2, c_1)$

Examples of Application Constraints - Type 1

- $CI(X_1, c_1) \rightarrow CI(X_1, c_2)$
  $PI(X_1, X_2, p_1) \rightarrow PI(X_1, X_2, p_2)$
  $CI(X_1, c_1) \rightarrow \neg CI(X_1, c_2)$
  $PI(X_1, X_2, p_1) \rightarrow \neg PI(X_1, X_2, p_2)$
  - $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $PI(X_1, X_2, ExpressLineStop) \rightarrow \neg PI(X_1, X_2, StreetLineStop)$

Examples of Application Constraints - Type 2

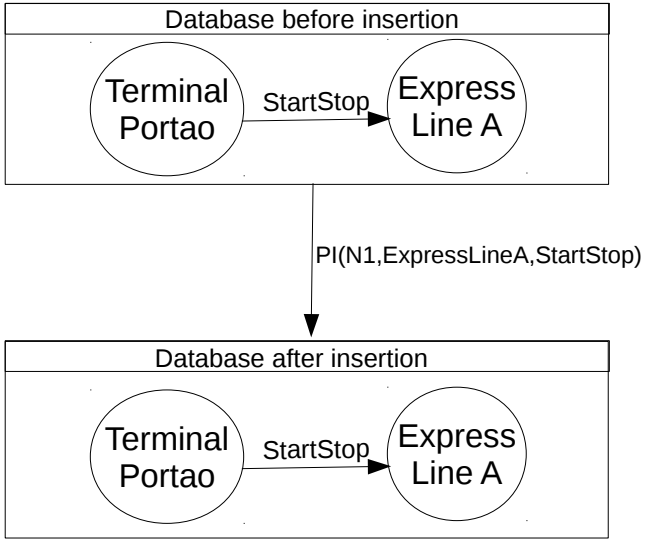
- $CI(X_1, c_1) \rightarrow PI(X_1, X_2, p_1)$
  $CI(X_1, c_1) \rightarrow \neg PI(X_1, X_2, p_1)$
  $CI(X_2, c_1) \rightarrow PI(X_1, X_2, p_1)$
  $CI(X_2, c_1) \rightarrow \neg PI(X_1, X_2, p_1)$
  - $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$
  - $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, EndStop)$

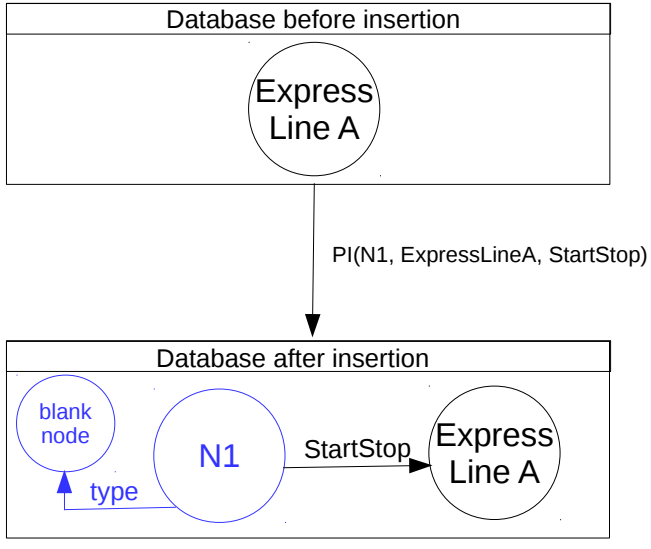
Examples of Application Constraints - Type 3

- $PI(X_1, X_2, p_1) \rightarrow CI(X_1, c_1)$
  $PI(X_1, X_2, p_1) \rightarrow \neg CI(X_1, c_1)$
  $PI(X_1, X_2, p_1) \rightarrow CI(X_2, c_1)$
  $PI(X_1, X_2, p_1) \rightarrow \neg CI(X_2, c_1)$
  - $PI(X_1, X_2, ExpressLineStop) \rightarrow CI(X_2, ExpressLine)$

Example of Property Insertion with Blank Node

- Insertion of $PI(N1, ExpressLineA, StartStop)$ = If there exists express line stop X in database D associated to *ExpressLineA* by property *StartStop* nothing is inserted. Otherwise, $\{PI(N1, ExpressLineA, StartStop), BN(N1)\}$ are inserted.





Summary

Example of Side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{CI(Portao, Terminal)\}$
  - upd + side effects = $\{CI(Portao, Terminal), CI(Portao, FastBoarding)\}$
- $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$
  - $upd = \{CI(303, ExpressLine)\}$
  - upd + side effects = $\{CI(303, ExpressLine), PI(N1, 303, StartStop)\}$

Example of Side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{CI(Portao, Terminal)\}$
  - upd + side effects = $\{CI(Portao, Terminal),\ CI(Portao, FastBoarding)\}$
- $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$
  - $upd = \{CI(303, ExpressLine)\}$
  - upd + side effects = $\{CI(303, ExpressLine), PI(N1, 303, StartStop)\}$

Example of Side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{CI(Portao, Terminal)\}$
  - upd + side effects = $\{CI(Portao, Terminal), CI(Portao, Terminal)\}$
- $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$
  - $upd = \{CI(303, ExpressLine)\}$
  - upd + side effects = $\{CI(303, ExpressLine), PI(N1, 303, StartStop)\}$

Example of Side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{CI(Portao, Terminal)\}$
  - upd + side effects = $\{CI(Portao, Terminal), CI(Portao, FastBoarding)\}$
- $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$
  - $upd = \{CI(303, ExpressLine)\}$
  - upd + side effects = $\{CI(303, ExpressLine), PI(N1, 303, isMember)\}$

Example of side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{ \neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects = $\{\neg CI(Portao, FastBoarding), \neg CI(Portao, Terminal)\}$
- $PI(X_1, X_2, ExpressLineStop) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects= $\{\neg CI(Portao, FastBoarding), \neg PI(Portao, N1, ExpressLineStop)\}$

Example of side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects = $\{\neg CI(Portao, FastBoarding), \neg CI(Portao, Terminal)\}$
- $PI(X_1, X_2, ExpressLineStop) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects= $\{\neg CI(Portao, FastBoarding), \neg PI(Portao, N1, ExpressLineStop)\}$

Example of side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects $=$
    $\{\neg CI(Portao, FastBoarding), \neg CI(Portao, Terminal)\}$
- $PI(X_1, X_2, ExpressLineStop) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects$=$
    $\{\neg CI(Portao, FastBoarding), \neg PI(Portao, N1, ExpressLineStop)\}$

Example of side Effects

- $CI(X_1, Terminal) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects = $\{\neg CI(Portao, FastBoarding), \neg CI(Portao, Terminal)\}$
- $PI(X_1, X_2, ExpressLineStop) \rightarrow CI(X_1, FastBoarding)$
  - $upd = \{\neg CI(Portao, FastBoarding)\}$
    1. $\neg upd \rightarrow \{CI(Portao, FastBoarding)\}$
    2. $upd \leftarrow \neg <$ *left side of rule* $>$
  - upd + side effects= $\{\neg CI(Portao, FastBoarding), \neg\ PI(Portao, N1, ExpressLineStop)\}$

Example of Side Effects

- $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$
  - $upd = \{\neg\ PI(TuboMD, 303, StartStop)\}$
  - upd + side effects = $\{\neg PI(TuboMD, 303, StartStop), PI(N1, 303, StartStop)\}$
  - 303 has another start stop but it is an express line

Summary

Related Work

- Consistency in RDF
  - Frommhold [SEMANTICS 2016] - version control of blank nodes (different semantics)
  - Language to update RDF database
  - Magiridou [ISWC 2005] - control frequent updates on RDF
  - Consistency maintenance
  - Flouris [KIS 2013] -Application/RDF constraints
- Constraints on urban transportation
  - Li and Tong [Socio-Economic Planning Sciences]
  - Yang [Transportation Research Part B: Methodological]
  - da Silva [ITSC]
  - de Bona [Mathematical Problems in Engineering]
  - Kozievitch [Annals of Data Science]

Flouris et al.(FKAC)

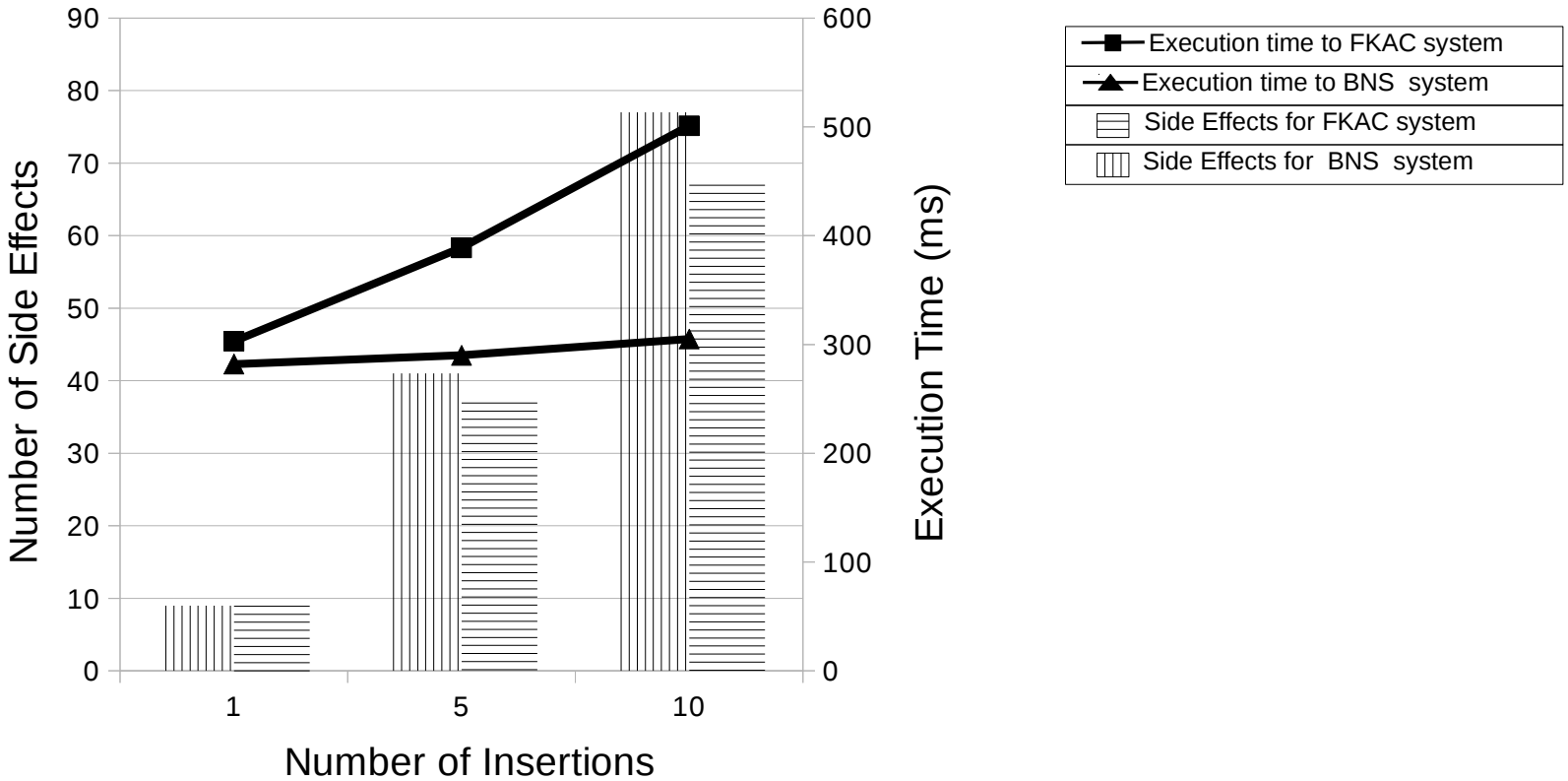
- Most similar to our proposal
  - Application and RDF semantic constraints
  - Minimal change
- Differences
  - Consider schema and instance operations
  - Cascade operations and arbitrary instantiation
  - Not deterministic - DELTA

Summary

Experiments

- Objective
  - Comparing our proposal (BNS system) with FKAC system based on number of side effects and execution time
- Implementation
  - SML#
  - PostgreSQL
- Dataset
  - URBS - Curitiba Urbanization Company

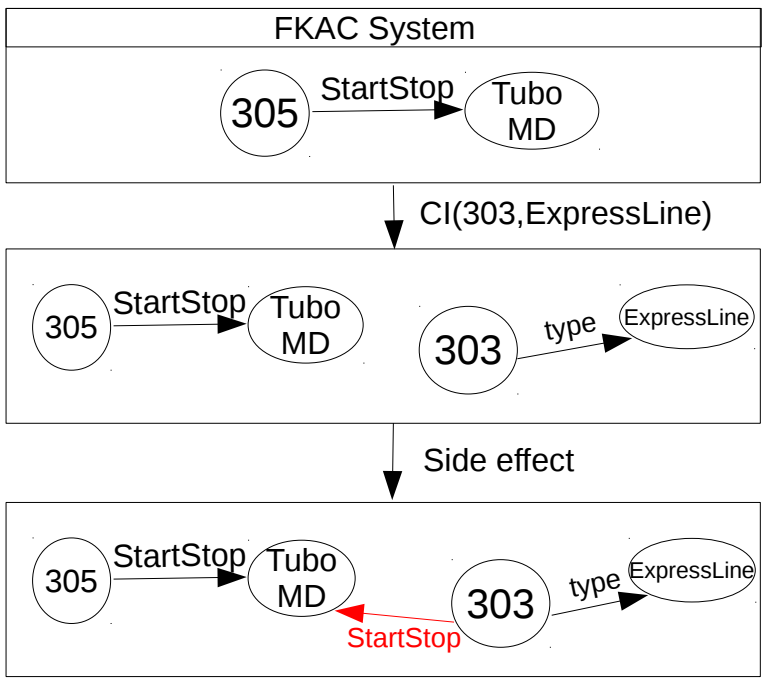
Experimental Study - Insertion Operation

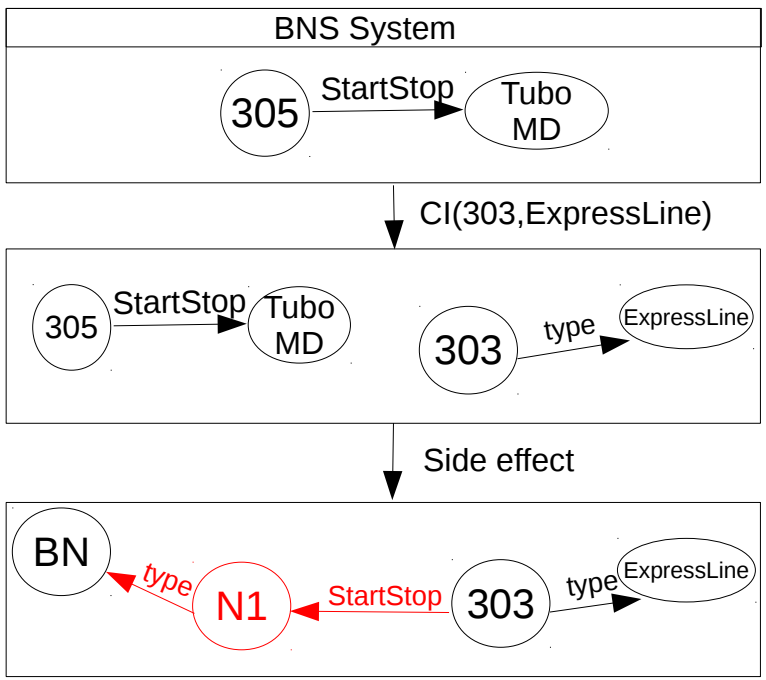


- Set of side effects is larger for BNS system
- Execution time is larger for FKAC system

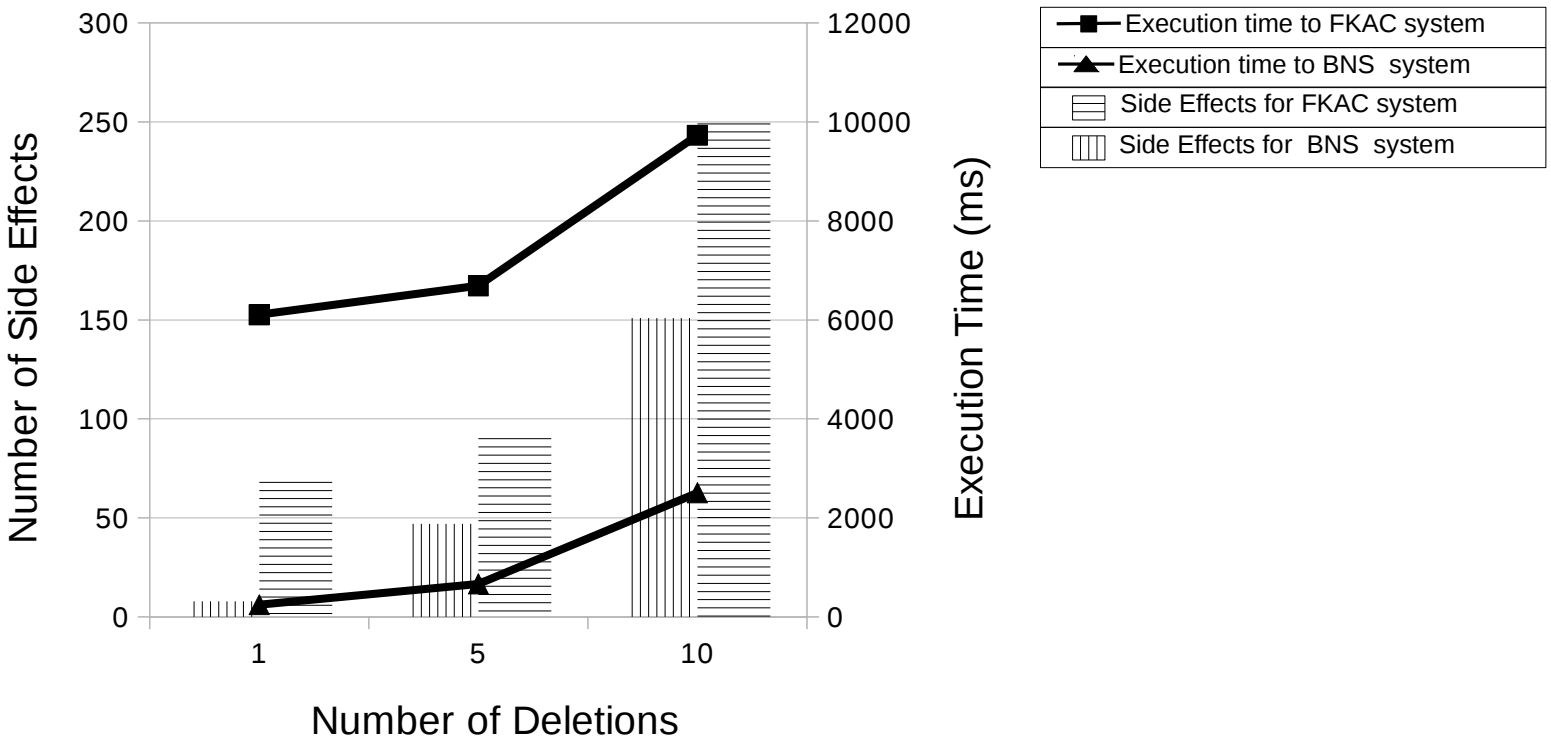
Experimental Study - Insertion Operation

- Arbitrary association in FKAC
  - Not removed in later time
- $CI(X_2, ExpressLine) \rightarrow PI(X_1, X_2, StartStop)$



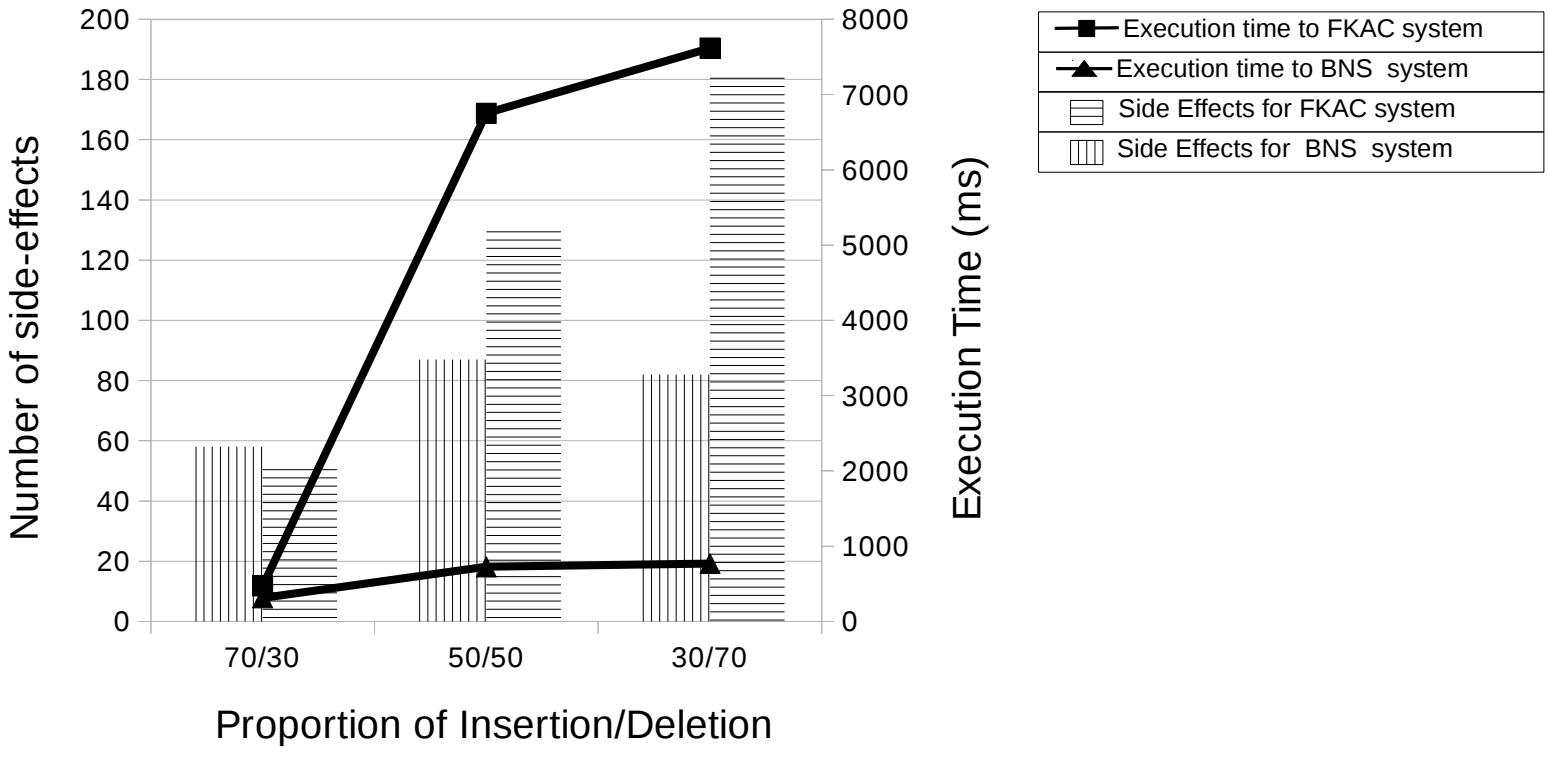


Experimental Study - Deletion Operation



- Set of side effects is larger for FKAC system
  - BNS system reduces cascade deletes

Experimental Study - Mixed Operations



- Number of side effects is larger for FKAC

Summary

Conclusion

- For insertion operations
  - BNS system generated number of side effects larger than FKAC system
  - BNS system gererated semantically meaningful side effects
- For deletion and mixed operations
  - FKAC system generated number of side effects larger than BNS system
- Future Work
  - Integration with ScEx or SHACL for validation of shape of RDF graph after updates
  - More extensive integration (URBS x BNS)

RDF Updates with Constraints

Authors: Flavio R. Uber (UEM/UFPR - Brazil)
Carmem S. Hara (UFPR - Brazil)
Nadia P. Kozievitch (UTFPR - Brazil)
Mirian Halfeld-Ferrari (Université d'Orleans - France)

contact: flavio.uber@gmail.com

THANK YOU