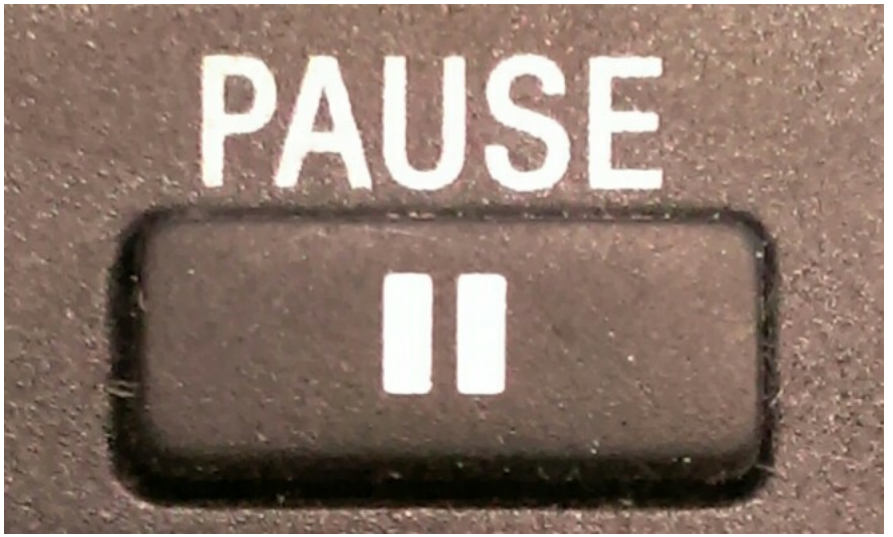
A Context-Based Semantics for SPARQL Property Paths over the Web

Olaf Hartig and Giuseppe Pirrò

Puzzle

Give me the URL of every Web document that mentions its own URL somewhere in its text.

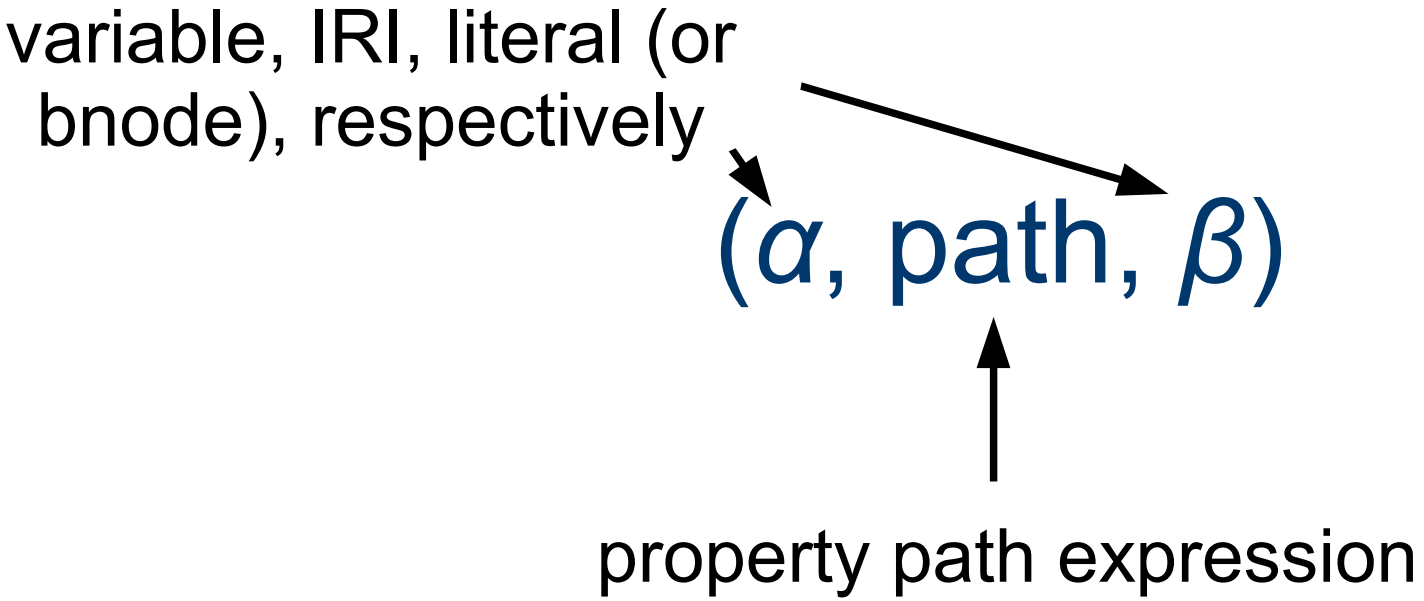


Disclaimer:
This work is not about reasoning.

A Context-Based Semantics for SPARQL Property Paths over the Web

Olaf Hartig and Giuseppe Pirrò

Property Path Pattern (PP Pattern)



Property Path Expression

variable, IRI, literal (or bnode), respectively

$(\alpha, \text{path}, \beta)$

path :=	p	(IRI)
	^path	(inverse)
	path / path	(concatenation)
	(path | path)	(alternative)
	(path)*	(recursive concatenation)

Example

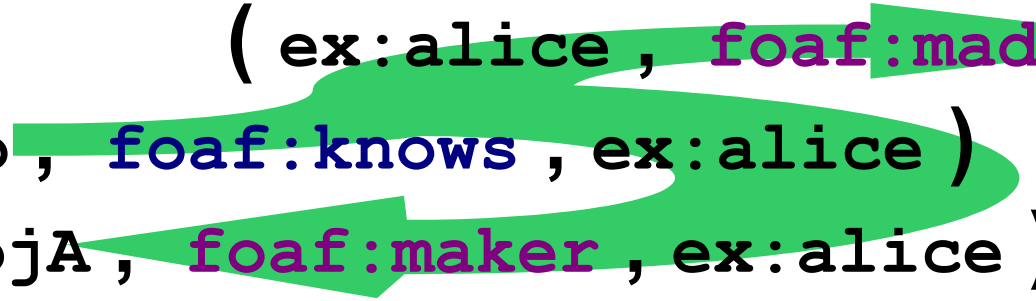
(ex:bob , foaf:knows / (foaf:made | ^foaf:maker) , $?x$)

Data:

(ex:alice , foaf:made , a:doc1)

(ex:bob , foaf:knows , ex:alice)

(e:objA , foaf:maker , ex:alice)



path := p | (IRI)

^path | (inverse)

path / path | (concatenation)

ath) | (alternative)

(recursive concatenation)

μ_1 = { $?x$ → a:doc1 }

μ_2 = { $?x$ → e:objA }

Example

$P =$

$(\texttt{ex:bob}, \texttt{foaf:knows} / (\texttt{foaf:made} \mid \texttt{\^{}foaf:maker}), ?x)$

$$G = \{ (\texttt{ex:alice}, \texttt{foaf:made}, \texttt{a:doc1}), (\texttt{ex:bob}, \texttt{foaf:knows}, \texttt{ex:alice}), (\texttt{e:objA}, \texttt{foaf:maker}, \texttt{ex:alice}) \}$$

$$[\![P]\!]_G = \{ \mu_1, \mu_2 \}$$

$$\mu_1 = \{ ?x \rightarrow \texttt{a:doc1} \}$$

$$\mu_2 = \{ ?x \rightarrow \texttt{e:objA} \}$$

Example

P =

(ex:bob , foaf:knows / (foaf:made | ^foaf:maker) , $?x$)

(ex:alice , foaf:made , a:doc1)

(ex:bob , foaf:knows , ex:alice)

(e:objA , foaf:maker , ex:alice)

Example

P =

(ex:bob , foaf:kno

(ex:alice , foaf:made , a:doc1)

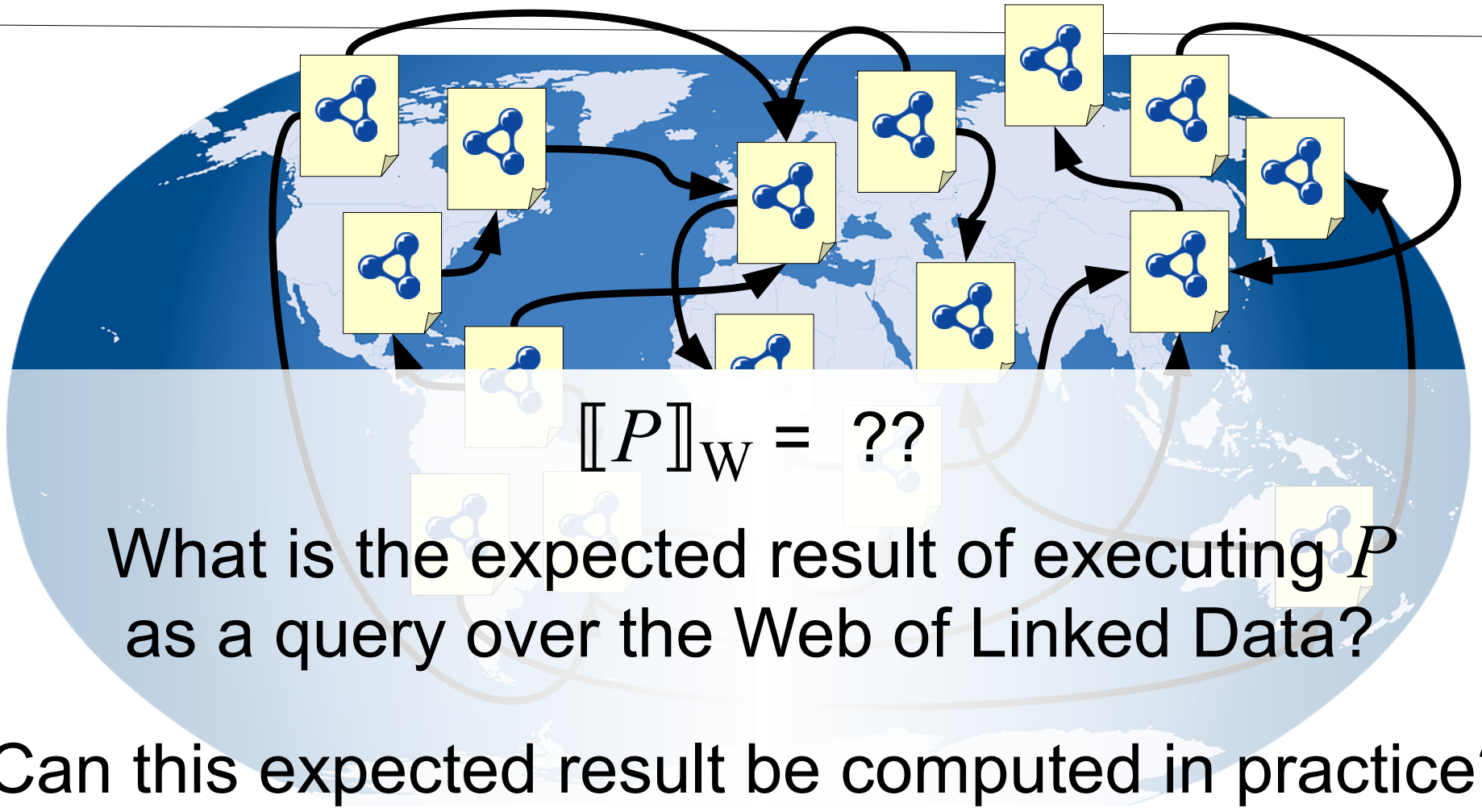
(ex:bob , foaf:knows , ex:alice)

(e:objA , foaf:maker , ex:alice)

Example

P =

(**ex:bob** , **foaf:knows** / (**foaf:made** | **^foaf:maker**) , $?x$)



$[\![P]\!]_{\mathrm{W}}$ = ??

What is the expected result of executing P as a query over the Web of Linked Data?

Can this expected result be computed in practice?

Contributions

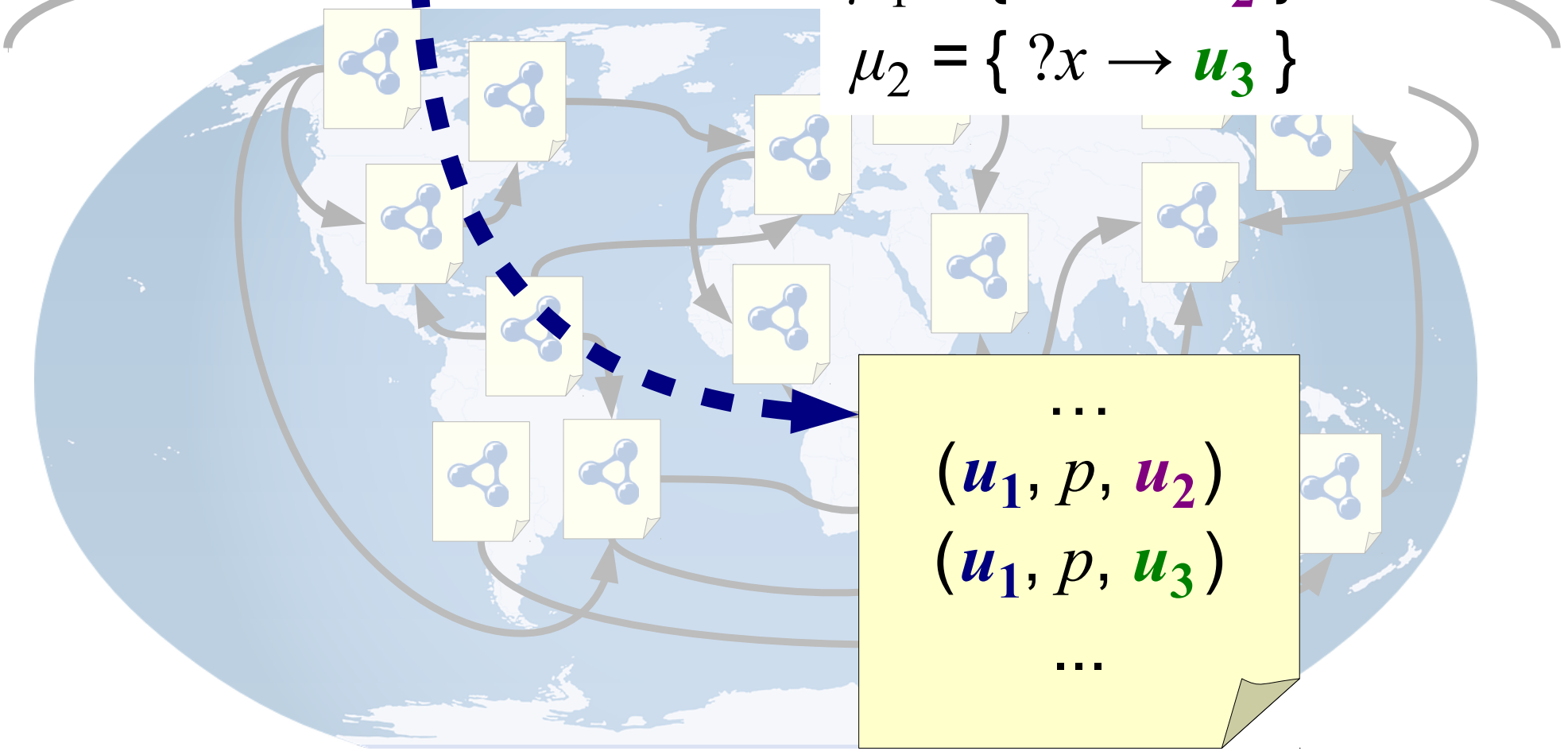
- **Formal definition of a context-based query semantics** to use PP patterns (and SPARQL patterns constructed thereof) as queries over Linked Data in a well-defined manner.
- **Web-safeness property** of such queries (i.e., the fundamental feasibility of executing such a query in practice over the WWW)
- **Decidable syntactic property** to identify queries that are Web safe

Context-Based Query Semantics

$$[\![(u_1, p, ?x)]\!]_W = \{ \mu_1, \mu_2 \}$$

$$\mu_1 = \{ ?x \rightarrow u_2 \}$$

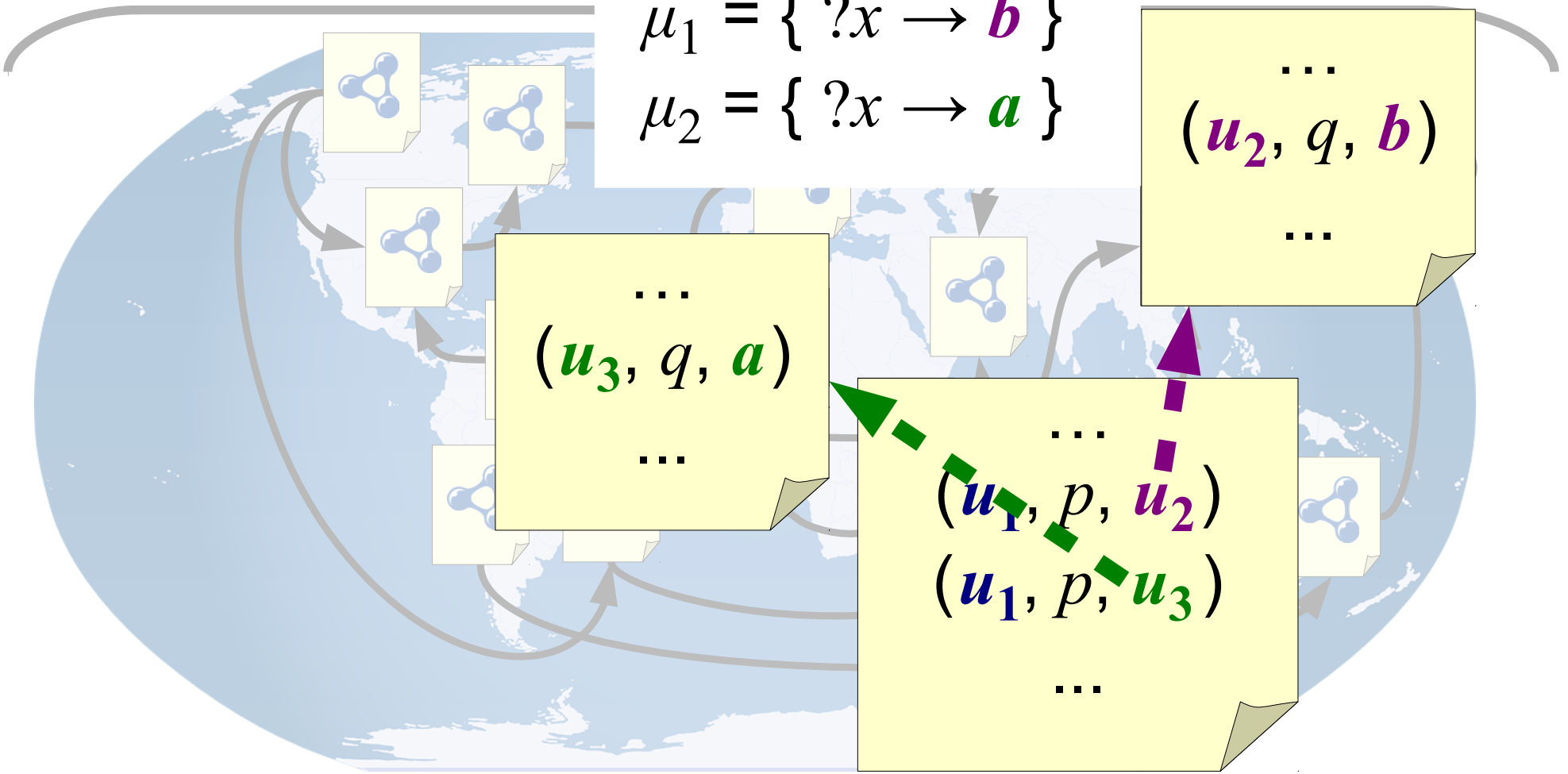
$$\mu_2 = \{ ?x \rightarrow u_3 \}$$



Context-Based Query Semantics

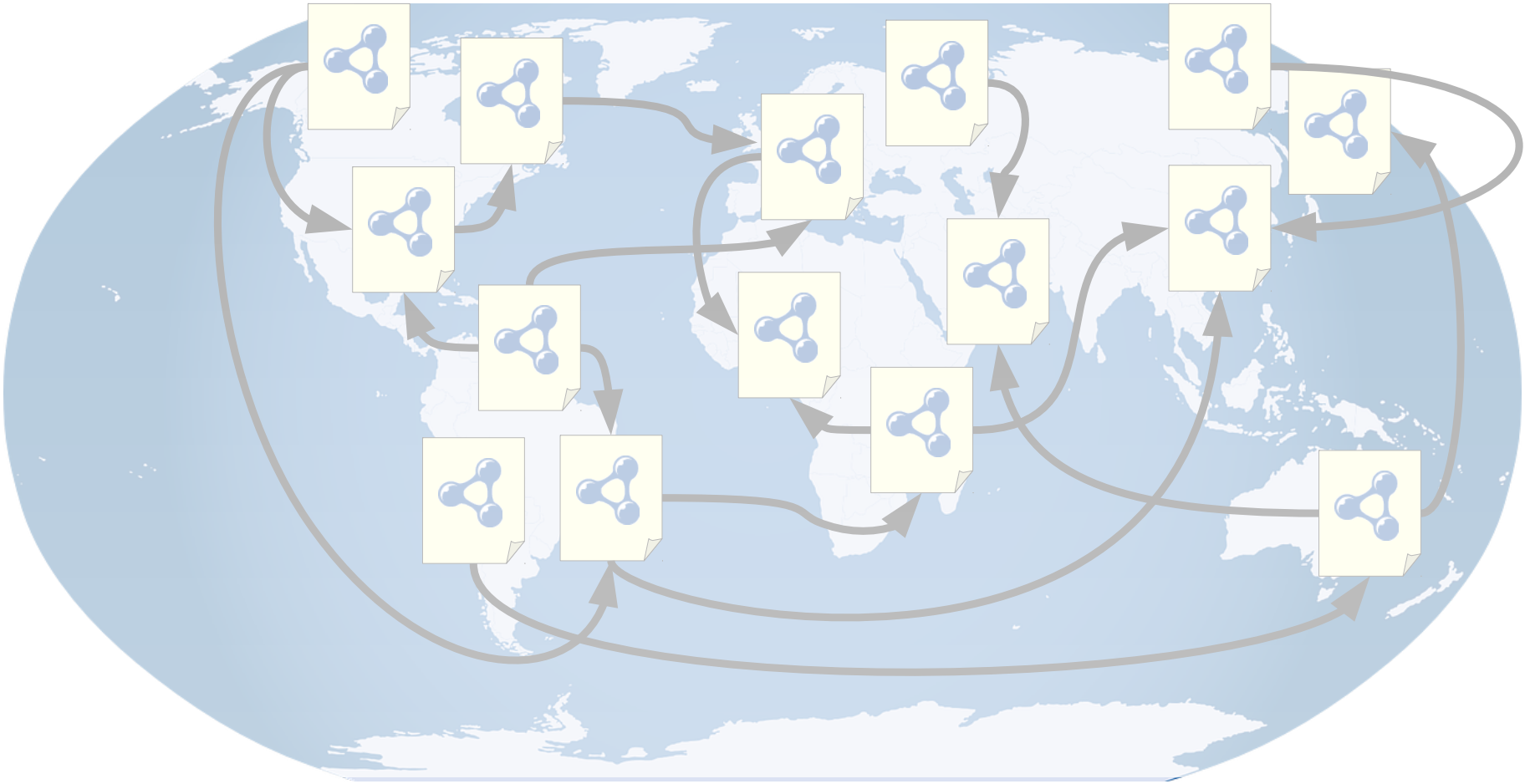
$$[\![(u_1, p \,/\, q\ , ?x)]\!]_W = \{\, \mu_1, \mu_2 \,\}$$

$$\mu_1 = \{\, ?x \rightarrow b \,\}$$
$$\mu_2 = \{\, ?x \rightarrow a \,\}$$



Context-Based Query Semantics

$$[\![(?x, p, a)]\!]_W = \{\ \mu=\{\ ?x \rightarrow u\ \} \mid \text{for every URI } u \text{ s.t. } (u, p, a) \text{ in ctxt}(u)\ \}$$



Main Result

Theorem: A PP-based SPARQL pattern P is Web-safe if $cbvars(P \mid \emptyset) = vars(P)$,

...where $cbvars(P \mid X)$ denotes the set of conditionally bounded variables of P for a given set X of variables (see paper).

- Open problem:
  - Note that the condition is sufficient only.
  - Is there an alternative condition that is sufficient and also necessary (and decidable)?

Summary

- **Formal definition of a context-based query semantics** to use PP patterns (and SPARQL patterns constructed thereof) as queries over Linked Data in a well-defined manner.
- **Web-safeness property** of such queries (i.e., the fundamental feasibility of executing such a query in practice over the WWW)
- **Decidable syntactic property** to identify queries that are Web safe