

RDF图数据库的随机差分测试

杨睿, 郑莹莹, 汤磊, 窦文生, 王伟, 魏峻

Randomized Differential Testing of RDF Stores

The 45th IEEE/ACM International Conference on Software Engineering (ICSE Demo 2023)

联系方式: 杨睿, yangrui22@otcaix.iscas.ac.cn

RDF Graph

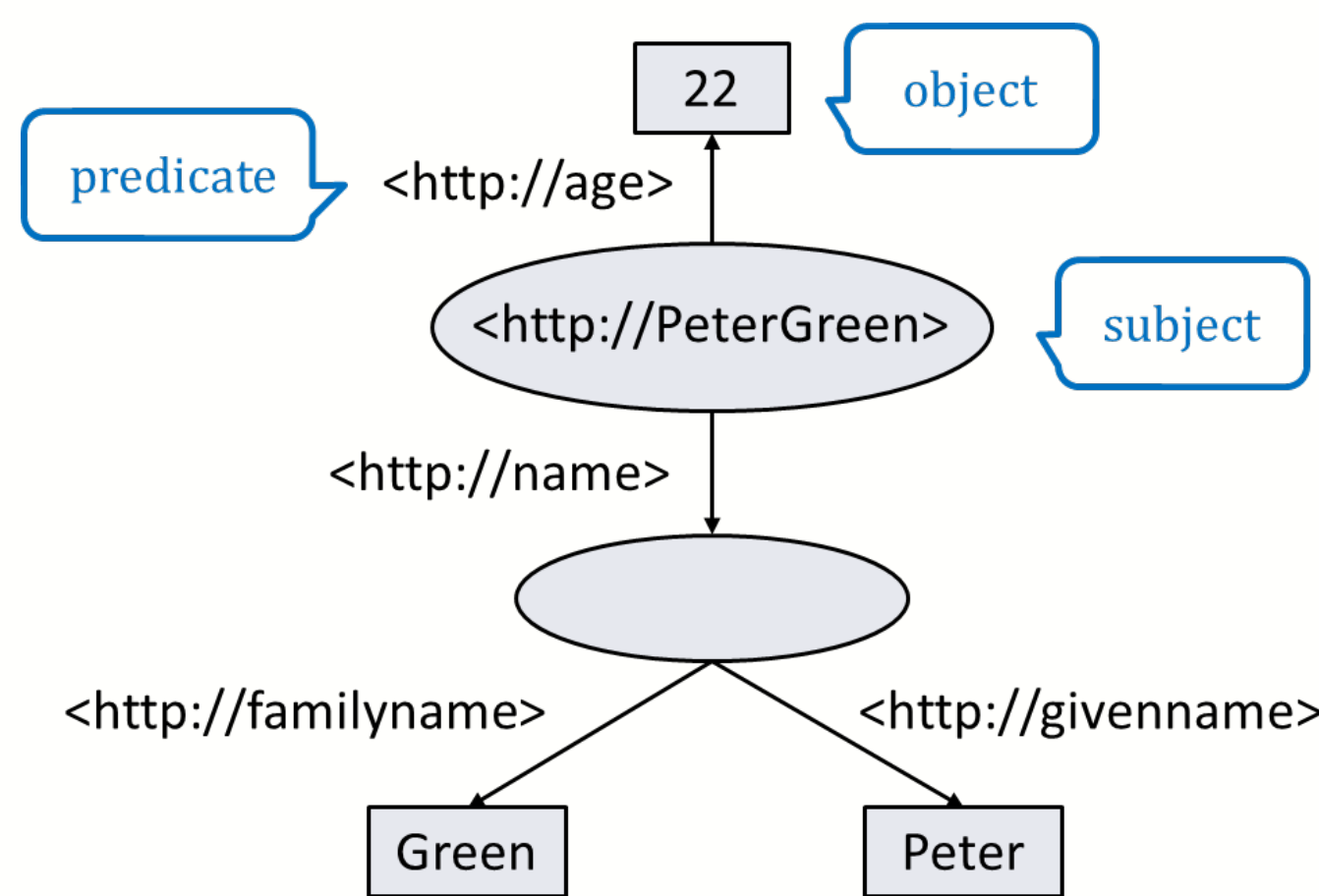
RDF triple: <subject, predicate, object>.

Basic element

IRIs: <http://PeterGreen> <http://age> <http://name>, <http://familyname>, <http://givenname>, ...

Blank nodes: b1, b2, ...

Literals: "Green", "Peter", 22, ...

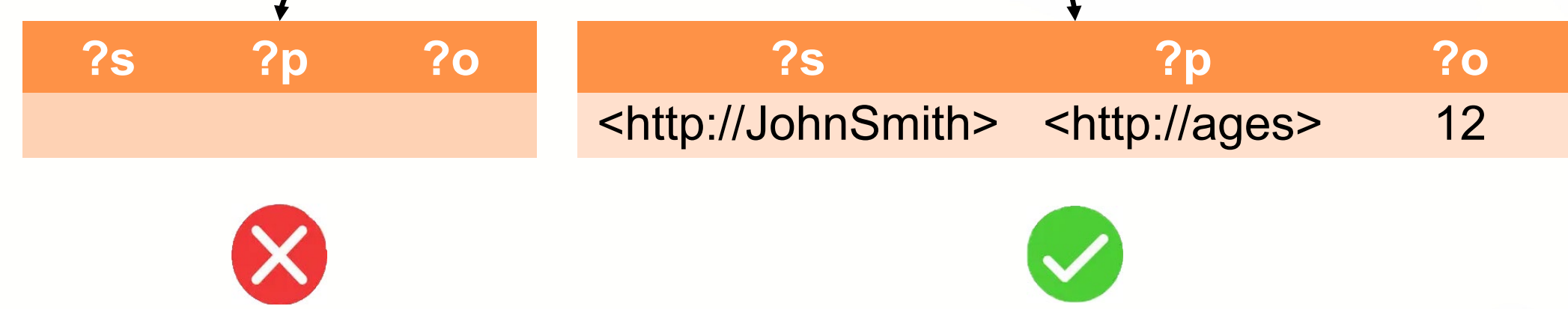


Logic Bugs in RDF Stores

Incorrect implementations can introduce logic bugs.

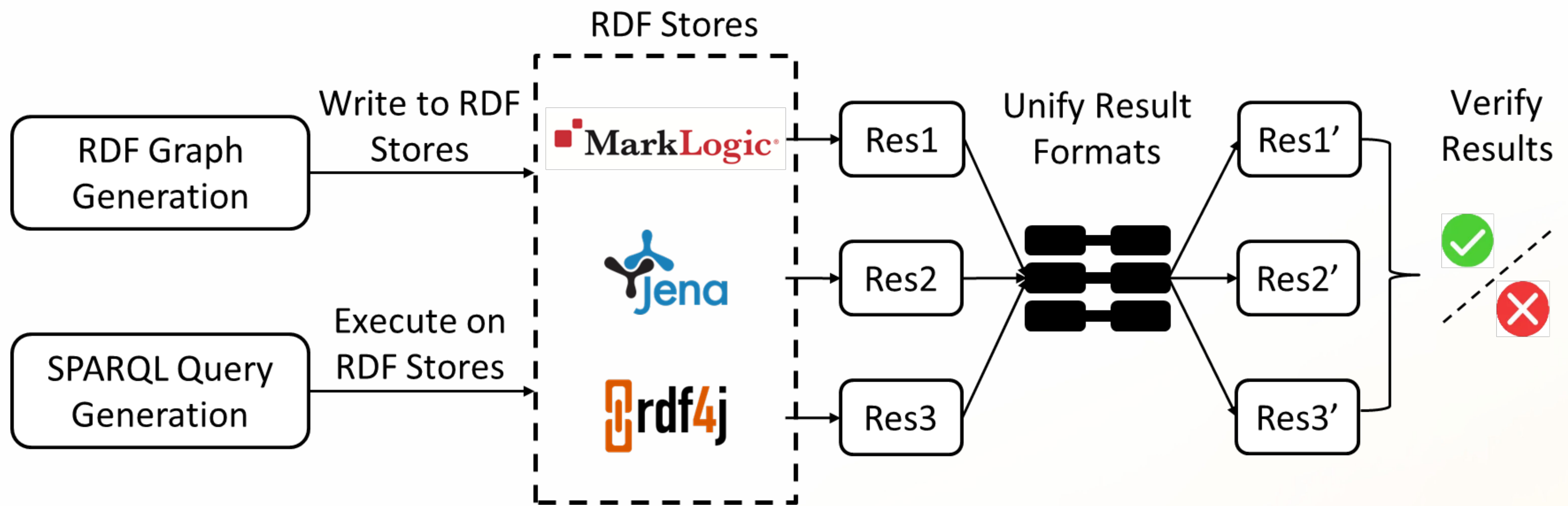
RDF Graph Data <http://JohnSmith> <http://ages> 12.

```
SPARQL Query
SELECT *
WHERE { ?s ?p ?o .
        FILTER ( 80596426678 * 1719307142 )
}
```



RD² Overview

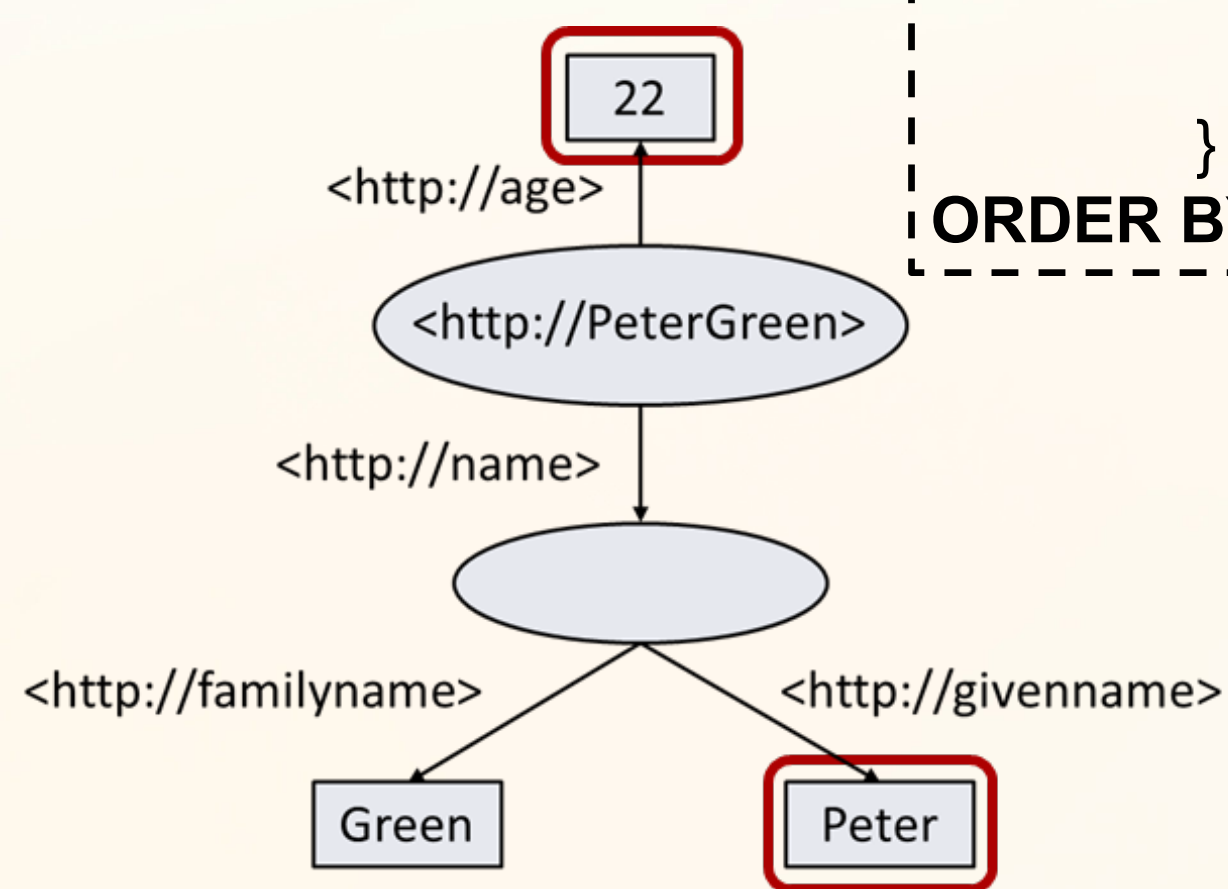
Goal: find logic bugs in RDF stores.



SPARQL Query Generation

Randomly build a SPARQL query's Abstract Syntax Tree.

```
SELECT ?givenName ?age
WHERE {
  ?person <http://name> ?name .
  ?person <http://age> ?age .
  ?name <http://givenname> ?givenName .
  FILTER ( ! ( ?age = 20 ) )
}
ORDER BY ?givenName
```

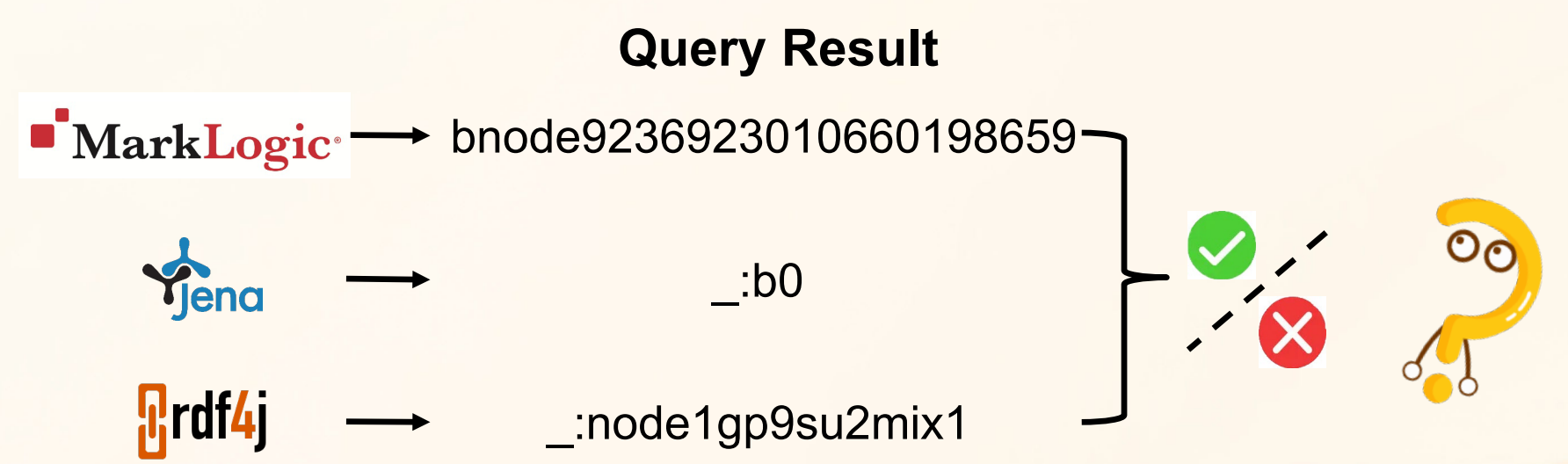


Syntactically correct and valid SPARQL query

Differential Testing

Compare the query results to find discrepancies.

```
SELECT ?name
WHERE {
  ?person <http://name> ?name .
}
```



Different RDF stores have their own storage and query result formats.

Differential Testing

Unify formats of blank node IDs.

MarkLogic actual ID	Jena actual ID	RDF4j actual ID	Unified ID
bnode9236923010660198659	_:b0	_:node1gp9su2mix1	1
bnode11935990903821970957	_:b1	_:node1gp9su2mix2	2
bnode16367137124885678738	_:b2	_:node1gp9su2mix3	3

Bug Detection Result

RD² finds 5 bugs among the three target RDF stores.

RDF Stores	Detected	Confirmed	Intended
MarkLogic	4	1	3
Apache Jena	-	-	-
RDF4j	1	1	-
Total	5	2	3

