A Quick tour to blank nodes

http://www.ics.forth.gr/isl/bnodeland

Lantzaki Christina

February 2013

Outline

- Blank nodes in RDF
- Blank nodes in various formats
- Blank nodes in SPARQL
- Blank nodes in published data
- Tasks sensitive on blank nodes
- Algorithms for bnode matching

Blank nodes in RDF

Blank nodes

- neither URIs nor literals
- no global identity
- as subjects or objects
- blank node identifiers have only local scope

Blank nodes in RDF

Blank nodes

- play the role of existential variables in RDF

Blank nodes in RDF

Blank nodes

- describe multi-components structures (e.g. RDF containers)

```
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc:<http://purl.org/dc/elements/1.1/> .
@prefix ex:<http://example.org/stuff/1.0/>.

<http://www.csd.uoc.gr/~hy561>
                                dc:title "Web Data Management" ;
                                ex:professor  _:b ;
                                ex:students   _:students .
_:b
     ex:fullName "Yannis Tzitzikas" ;
     ex:homePage <http://www.ics.forth.gr/~tzitzik/> .


_:students rdf:type rdf:Bag ;
           dc:hasMember _:s1 ;
           dc:hasMember _:s2 ;
           dc:hasMember _:s3 ;
           dc:hasMember _:s4 ;
           dc:hasMember _:s5 .
```

Blank nodes in RDF

Blank nodes

- describe complex attributes, whose identity is unknown, but their attributes are known

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://example.org/stuff/1.0/>.

<http://www.csd.uoc.gr/~hy561>
                                dc:title "Web Data Management" ;
                                ex:professor  _:b ;
                                ex:students   _:students .

_:b
     ex:fullName "Yannis Tzitzikas" ;
     ex:homePage <http://www.ics.forth.gr/~tzitzik/> .
     ex:hasAddress  _:ad

_:ad  rdf:type ex:Address ;
      ex:street "Knossou" ;
      ex:number "122";
      ex:postalcode "71409" ;
      ex:city "Heraklion" .
```

Blank nodes in RDF

Blank nodes

- describe reification (provenance information)

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://example.org/stuff/1.0/>.
@prefix prov: <http://www.w3.org/ns/prov#>.

<http://www.csd.uoc.gr/~hy561>
                                dc:title "Web Data Management" ;
                                ex:professor  _:b ;
                                prov:generatedBy _:a1 .
_:b  ex:fullName "Yannis Tzitzikas" .

_:a1 rdf:type prov:Event ;
     prov:creator _:b ;
     prov:atTime "Tuesday 11 February, 06:51:00 CST" .

_:a2 rdf:type prov:Event ;
     rdf:type prov:Update;
     prov:ActionOver _:a1;
     prov:creator _:b ;
     prov:atTime "Monday 17 February, 08:12:00 CST" .
```

Blank nodes in various formats

Blank nodes can be expressed with **blank node identifiers**

- always locally scoped to the RDF graph
- not persistent or portable

Turtle / N-Triples / N3

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://example.org/stuff/1.0/>.
<http://www.csd.uoc.gr/~hy561>
                        dc:title "Web Data Management" ;
                        ex:professor  _:b .
 _:b
   ex:fullName "Yannis Tzitzikas" ;
   ex:homePage <http://www.ics.forth.gr/~tzitzik/> .
```

Blank nodes in various formats

Blank nodes can be expressed with **blank node identifiers**

- always locally scoped to the RDF graph
- not persistent or portable

RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:ex="http://example.org/stuff/1.0/">
<rdf:Description rdf:about="http://www.csd.uoc.gr/~hy561"
   dc:title="Web Data Management">
<ex:professor rdf:nodeID="_:b"/>
</rdf:Description>
<rdf:Description rdf:nodeID="_:b" ex:fullName="Yannis Tzitzikas">
   <ex:homePage rdf:resource="http://www.ics.forth.gr/~tzitzik/"/>
</rdf:Description>
</rdf:RDF>
```

Blank nodes in various formats

Blank nodes can be expressed through **nesting**

Turtle / N3

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://example.org/stuff/1.0/>.
<http://www.csd.uoc.gr/~hy561> dc:title "Web Data Management" ;
  ex:professor [ex:fullName "Yannis Tzitzikas" ;
                ex:homePage <http://www.ics.forth.gr/~tzitzik/>].
```

Blank nodes in various formats

Blank nodes can be expressed through **nesting**

RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
                 xmlns:dc="http://purl.org/dc/elements/1.1/"
                 xmlns:ex="http://example.org/stuff/1.0/">
<rdf:Description rdf:about="http://www.csd.uoc.gr/~hy561"
dc:title="Web Data Management">
<ex:professor>
<rdf:Description ex:fullName="Yannis Tzitzikas">
   <ex:homePage rdf:resource="http://www.ics.forth.gr/~tzitzik/"/>
</rdf:Description>
</ex:professor>
</rdf:Description>
</rdf:RDF>
```

Blank nodes in SPARQL

- Blank nodes in graph patterns:
  - act as non-distinguished variables
  - do not act as references to specific blank nodes
  - cannot appear in a SELECT clause
  - indicated by either the label form, such as "_:abc", or the abbreviated form "[]"

Query: Get the classes of the instances that are blank nodes

```
SELECT DISTINCT ?concept
WHERE {[] rdf:type ?concept.
   }
```

Blank nodes in SPARQL

- The same blank node label cannot be used in two different basic graph patterns in the same query.
- A blank node which appears more than once in the result set indicates the same blank node
- The labels of the blank nodes in the results only indicate whether RDF terms in the solutions are the same or different
- There need not be any relation between a label _:a in the result set and a blank node in the data graph with the same label

Running Example

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://example.org/stuff/1.0/>.
@prefix prov: <http://www.w3.org/ns/prov#>.

<http://www.csd.uoc.gr/~hy561>
                                dc:title "Web Data Management" ;
                                ex:professor  _:b ;
                                prov:generatedBy _:a1 .
_:b rdf:type Person
    ex:fullName "Yannis Tzitzikas" .

_:a1 rdf:type prov:Event ;
     prov:creator _:b ;
     prov:atTime "Tuesday 11 February, 06:51:00 CST" .

_:a2 rdf:type prov:Event ;
     rdf:type prov:Update;
     prov:ActionOver _:a1;
     prov:creator _:b ;
     prov:atTime "Monday 17 February, 08:12:00 CST" .
```

Blank nodes in SPARQL

- The isBlank operator returns *true* if the RDF term is a blank node. Otherwise, it returns *false*.
  - Data: running example
  - Query: get all the triples that contain two blank nodes

```
SELECT *
WHERE { ?x ?y ?z . FILTER isBlank(?x) . FILTER isBlank(?z)
   }
```

  - Result

?x	?y	?z
_:a	prov:creator	_:b
_:c	prov:ActionOver	_:a
_:c	prov:creator	_:b

Blank nodes in SPARQL

- Data: running example
- Query: get the professor of the subject hy561

```
PREFIX ex: <http://example.org/stuff/1.0/>.
SELECT ?person ?name
WHERE { <http://www.csd.uoc.gr/~hy561> ex:professor ?person .
        ?person ex:fullName ?name .
   }
```

- Result

?person	?name
_:s	"Yannis Tzitzikas"

Blank nodes in SPARQL

- Data: running example
- Query: get the creator and the time of the events over the digital object http://www.csd.uoc.gr/~hy561

```
PREFIX ex: <http://example.org/stuff/1.0/>.
PREFIX  prov: <http://www.w3.org/ns/prov#>.
SELECT ?person ?name ?time
WHERE { <http://www.csd.uoc.gr/~hy561> ex:professor ?person .
        ?event prov:createdBy ?person .
        ?event prov:atTime ?time .
       ?person ?ex:fullName ?name .
   }
```

- Result

?person	?name	?time
_:c	"Yannis *Tzitzikas*"	"Tuesday 11 February, 06:51:00 CST"
_:c	"Yannis *Tzitzikas*"	"Monday 17 February, 08:12:00 CST"

Blank nodes in published data

Blank nodes prevail in published data

Blank node prevalence

Dataset	#blank nodes	%blank nodes
hi5.com foaf	148,409,536	87.5
livejournal.com	8,892,569	58.0
ontologycentral.com	2,882,803	86.0
opiumfield.com	1,979,915	17.4
freebase	1,109,485	15.6
vox.com	843,503	58.0
rdfabout.com	464,797	41.7
opencalais.com	160,441	44.9
soton.ac.uk	117,390	19.1
bbc.co.uk	101,899	7

Blank nodes in published data

See for yourself
Get the number of blank nodes through a SPARQL endpoint.

EXAMPLE

Project: `LOD cloud cache`
SPARQL endpoint: `http://lod.openlinksw.com/sparql/`
Date: `03 March 2014`

```
SELECT count(distinct ?x)
WHERE  {?x ?y ?z
        FILTER(isBlank(?x))}
```

Result: 19,203,856

Blank nodes in published data

- The majority of datasets contain blank nodes that form tree structures
- A small percentage contains more complex blank node structures

See for yourself

Get the number of blank nodes that participate in chains of length 3
Project: `LOD cloud cache`
SPARQL endpoint: `http://lod.openlinksw.com/sparql/`
Date: `03 March 2014`

```
select count(?s1)
where { ?s1 ?p1 ?o1 FILTER(isBlank(?s1))FILTER(isBlank(?o1))
                    FILTER(?s1 != ?o1).
        ?o1 ?p2 ?o2 FILTER(isBlank(?o2)) FILTER(?o1 != ?o2).
        ?o2 ?p3 ?o3 FILTER(isBlank(?o3)) FILTER(?o2 != ?o3)}
```

Result: 5,382,889

Blank nodes in published data

- The majority of datasets contain blank nodes that form tree structures
- A small percentage contains more complex blank node structures

See for yourself

Get the number of blank nodes that participate in cycles of length 4
Project: `LOD cloud cache`
SPARQL endpoint: `http://lod.openlinksw.com/sparql/`
Date: `03 March 2014`

```
select count(?s1)
where {?s1 ?p1 ?o1 FILTER(isBlank(?s1)) FILTER(isBlank(?o1))
                   FILTER(?s1 != ?o1) .
       ?o1 ?p2 ?o2 FILTER(isBlank(?o2)) FILTER(?o1 != ?o2).
       ?o2 ?p2 ?o3 FILTER(isBlank(?o3)) FILTER(?o2 != ?o3).
       ?o3 ?p3 ?s1 FILTER(?o3 != ?s1)
}
```

Result: 13,197

Tasks sensitive on blank nodes

The existence of blank nodes requires special treatment in various tasks:

- Equivalence checking
- Comparing RDF Knowledge Bases
- Entailment checking
- Storage of RDF data for versioning
- Synchronization

• All the above tasks require some kind of **matching between the blank nodes** of the working datasets

Algorithms for bnode matching

In [3] authors show how to exploit blank node anonymity in order to **reduce the delta size** when comparing RDF/S Knowledge Bases
Proved that finding the optimal blank node mapping is NP-Hard in the general case (polynomial if there are no directly connected blank nodes)

They present polynomial approximate algorithms for the general case:
- a *Hungarian*-based and
- a *Signature*-based

The algorithms provide a delta of 12.7 to 7,294 times smaller than without blank node matching.

The Signature-based algorithm requires only 10.5 seconds to match 153,600 blank nodes.

References

[1] L. Chen, H. Zhang, Y. Chen, and W. Guo. Blank Nodes in RDF. *Journal of Software, 2012.*

[2] A. Mallea, M. Arenas, A. Hogan, and A. Polleres. On Blank Nodes. In *Procs of the 10th Intern. Semantic Web Conference (ISWC 2011), 2011.*

[3] Y. Tzitzikas, C. Lantzaki, and D. Zeginis. Blank Node Matching and RDF/S Comparison Functions. In *Procs of the 11th Intern. Semantic Web Conference (ISWC), 2012.*

[4] R. Pichler, A. Polleres, F. Wei, and S. Woltran. dRDF: Entailment for Domain- Restricted RDF. In *Extended Semantic Web Conference (ESWC), 2008.*

[5] C. Lantzaki, T. Yannakis, Y. Tzitzikas, and A. Analyti. Generating Synthetic RDF Data with Connected

[6] Blank Nodes for Benchmarking. In *Procs of the Extended Semantic Web Conference (ESWC), 2014.*

[7] J. Munkres. Algorithms for the assignment and transportation problems. *J-SIAM,*5(1), 1957.

[8] *http://www.w3.org/TR/REC-rdf-syntax/*

[9] *http://www.ics.forth.gr/isl/bnodeland/*

[10] *http://www.w3.org/TR/rdf-sparql-query/*

[11] *http://www.csd.uoc.gr/~hy561/*