

# Understanding Billions of Triples with Usage Summaries

Shahan Khatchadourian

Mariano P. Consens

{shahan, consens}@cs.toronto.edu



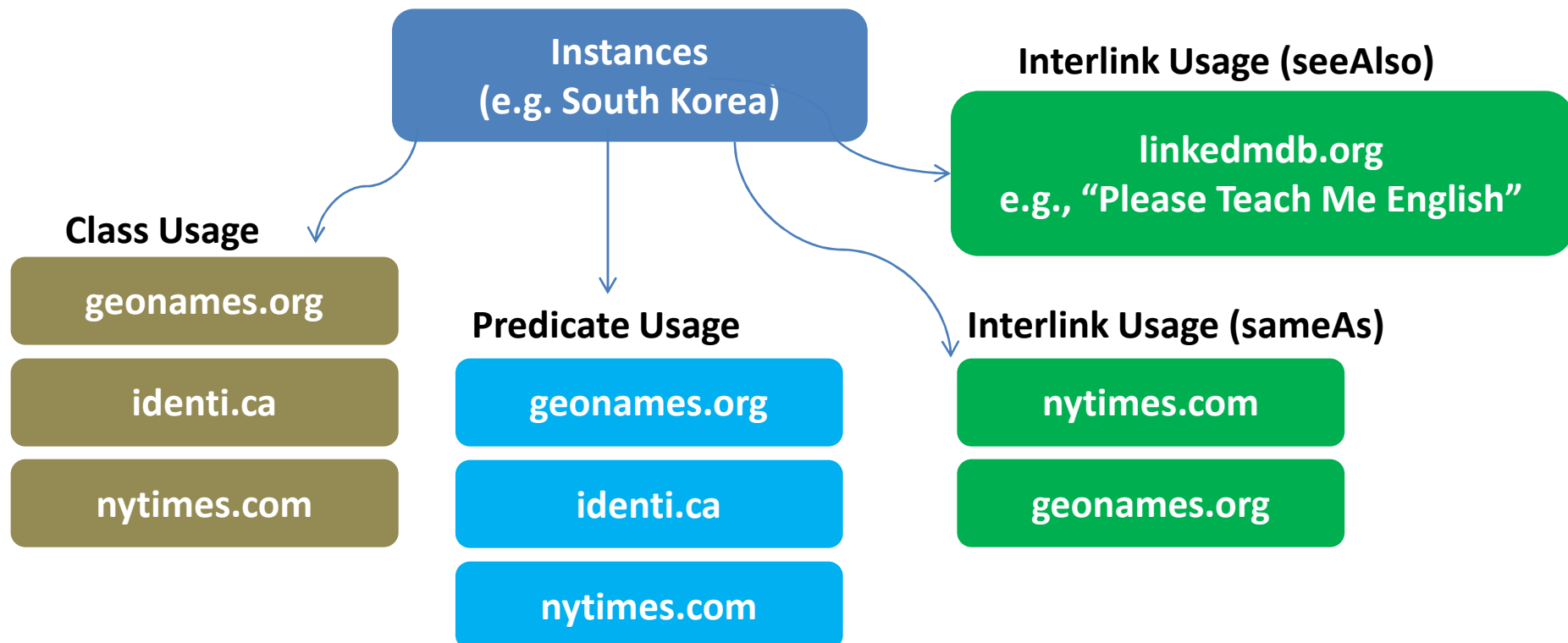
UNIVERSITY OF  
TORONTO

# Motivation and Contributions

- Summaries help to understand how published entities are **used together** by capturing **unique descriptions** (e.g., how classes and predicates interact).  
Understanding facilitates consumption of linked data.
- We provide a **flexible** mechanism to generate and compare summaries, with comprehensive support for **different usages** and **description patterns** (i.e., patterns for interlinking or inferencing)
- We provide a **scalable** approach. We summarize and compare the **entire BTC 2010** and **BTC 2011** datasets.
  - Previous analysis of BTC dataset [3,4] were limited to a domain or subset

# BTC Usage Summaries



ExpLOD [1] provides a mechanism to create comprehensive **usage summaries**, where an entity's usage is characterized as in [2].



# Interlinking Usage Summary Example

## Bisimulation labels (BCN)

/C#<fu-berlin.de>, /P/NI#<fu-berlin.de>,  


/P/<owl:sameAs>#<fu-berlin.de>, /P/<rdfs:seeAlso>#linkedmdb.org  
  
  


**Class usage: C**   **Graph: fu-berlin.de**

1 quad   <persons:Tony+Vitale> <rdf:type> <foafPerson> <persondoc:Tony+Vitale> .

**Predicate usage: P/NI**   **Graph: fu-berlin.de**

2 quads

<persons:Tony+Vitale> <rdfs:label> "Tony Vitale" <books:1449951546> .  
 <persons:Tony+Vitale> <foaf:name> "Tony Vitale" <persondoc:Tony+Vitale> .

**Interlink usage: P/<owl:sameAs>**   **Graph: fu-berlin.de**

1 quad   <persons:Tony+Vitale> <owl:sameAs> <dblp:283446> <persondoc:Tony+Vitale> .

**Interlink usage: P/<rdfs:seeAlso>**   **Graph: linkedMDB.org**

1 quad   <persons:Tony+Vitale> <rdfs:seeAlso>  
 <http://data.linkedmdb.org/sparql?query=DESCRIBE+%3Chttp%3A%2F%2Fwww  
 4.wiwiss.fu-berlin.de%2Fbookmashup%2Fpersons%2FTony%2BVitale%3E>  
 <http://data.linkedmdb.org/data/writer/3297>.

(BTC 2011 has 9 more instances with this interlink usage neighbourhood)

# Comprehensive Usage Summaries

LOD Semi-Structure
  Pattern Usage
  Ontology Usage

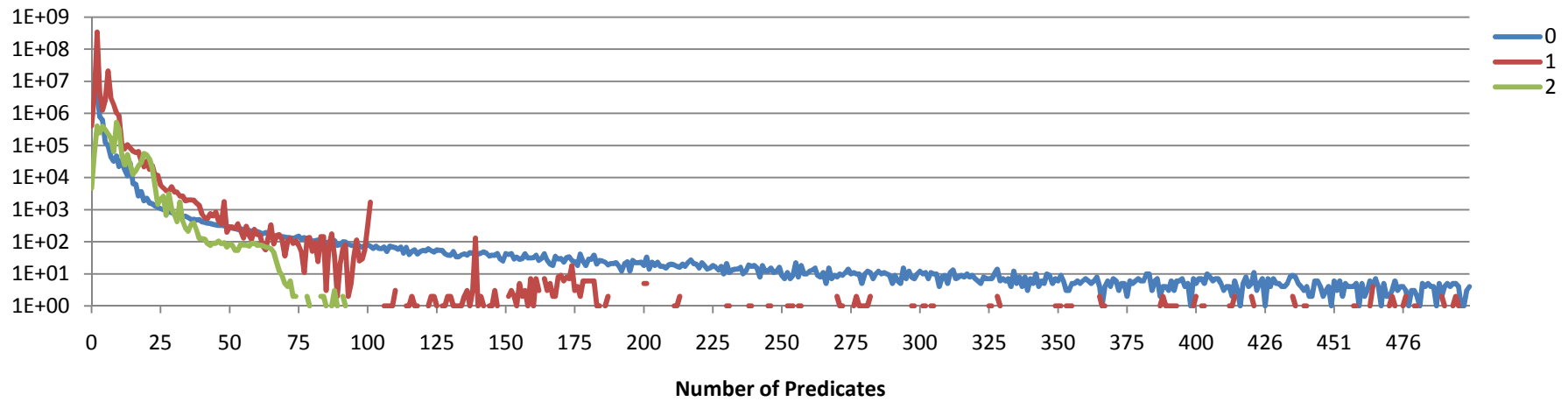
Usage Summary	Bisimulation Label		
	usage	graph	entity
Graph	x	x	none
CPO	x	none	class or predicate
<b>Complete</b>	x	x	class or predicate
Index	x	x	rdf:Seq or rdf:List
Inference	x	x	skos:broader or skos:narrower
Interlink	x	x	owl:sameAs, rdf:seeAlso, or skos:exactMatch
Topic	x	x	like foaf:topic or foaf:PrimaryTopic
FOAF	x	x	foaf:*
SKOS	x	x	skos:*

Usage Summary	Complete Usage Neighbourhoods		Instances	
	2010	2011	2010	2011
<b>Index</b>	12,324	60	710,972	769,138
<b>Inference</b>	120	256	68,672	842,938
<b>Interlink</b>	2,824	2,212	69,563,476	363,104,656
<b>Topic</b>	2,380	6,223	222,267	1,071,213

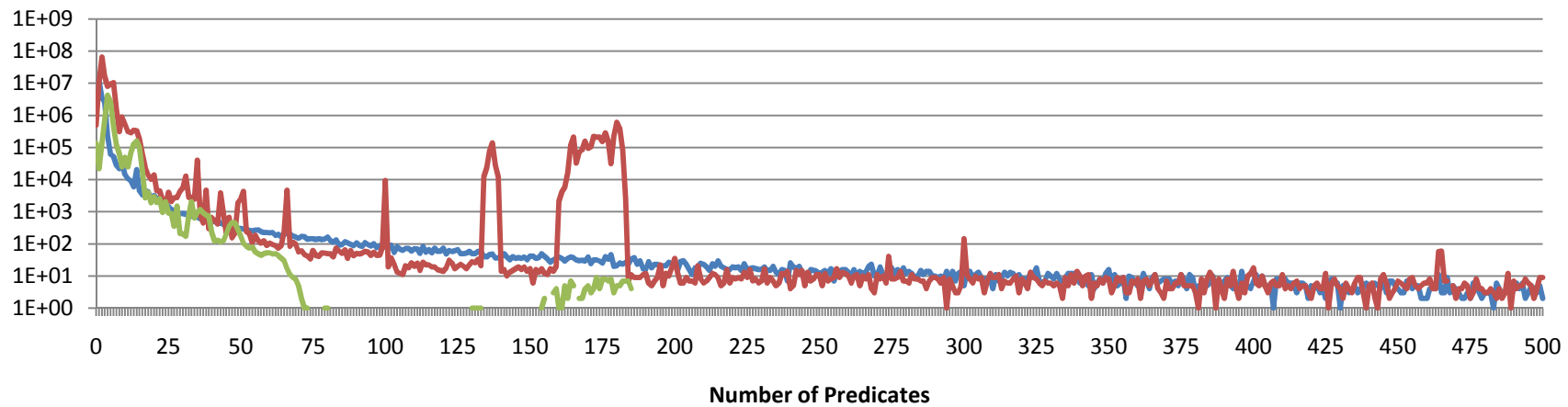


# Instance Counts in Class and Predicate Usage Neighbourhoods (Tails)

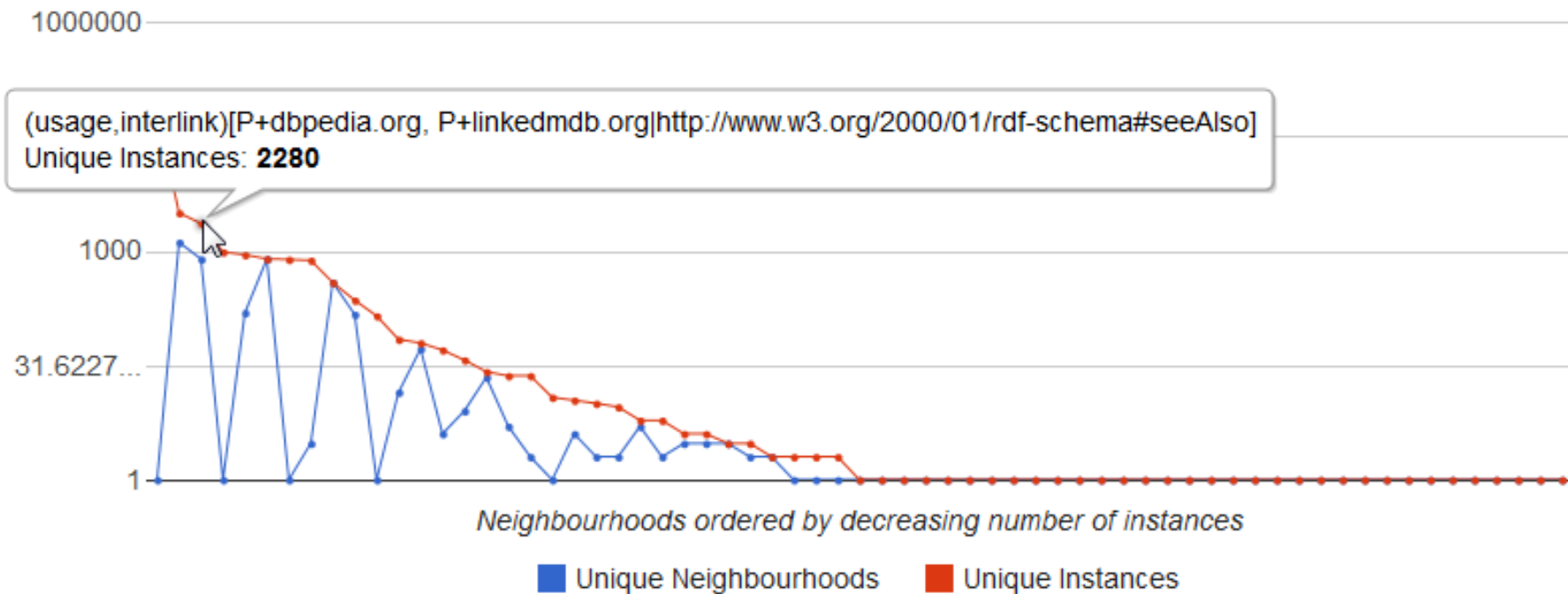
## Tail of BTC 2011



## Tail of BTC 2010



# Interactive Online Exploration Tool



	BCN	Unique Neighbourhoods	Unique Instances	Instance Quads
1	(usage,interlink)[P+linkedmdb.org http://www.w3.org/2000/01/rdf-schema#seeAlso]	1	63584	<a href="#">download</a>
2	(usage,interlink)[C+linkedmdb.org, P+linkedmdb.org, P+linkedmdb.org http://www.w3.org/2002/07/owl#sameAs]	1289	3129	<a href="#">download</a>
3	(usage,interlink)[P+dbpedia.org, P+linkedmdb.org http://www.w3.org/2000/01/rdf-schema#seeAlso]	778	2280	<a href="#">download</a>
4	(usage,interlink)[P+dbpedia.org http://www.w3.org/2002/07/owl#sameAs, P+linkedmdb.org http://www.w3.org/2000/01/rdf-schema#seeAlso]	1	962	<a href="#">download</a>
5	(usage,interlink)[P+linkedmdb.org, P+linkedmdb.org http://www.w3.org/2002/07/owl#sameAs]	154	885	<a href="#">download</a>



# References

Interactive Online Exploration Tool available at <http://www.cs.toronto.edu/~shahan/swc2011/>

1. S. Khatchadourian and M. P. Consens. ExpLOD: Summary-based exploration of interlinking and RDF usage in the linked open data cloud. In ESWC, pages 272–287, 2010.
2. L. Ding and T. Finin. Characterizing the semantic web on the web. In ISWC, pages 242–257, 2006
3. C. Böhm, J. Lorey, D. Fenz, E. Kny, M. Pohl, and F. Naumann. Creating void descriptions. Semantic Web Challenge 2010.
4. G. T. Williams, J. Weaver, M. Atre, and J. A. Hendler. Scalable reduction. Semantic Web Challenge 2009.
5. L. Dodds and I. Davis. Linked data patterns, <http://patterns.dataincubator.org/book>, Sept. 2011.