Maintenance of Knowledge Tags in Heterogeneous Web Content: The Repository

Karol Rástočný

Supervisor: Prof. Mária Bieliková

Aim: Lightweight Semantics over the Web

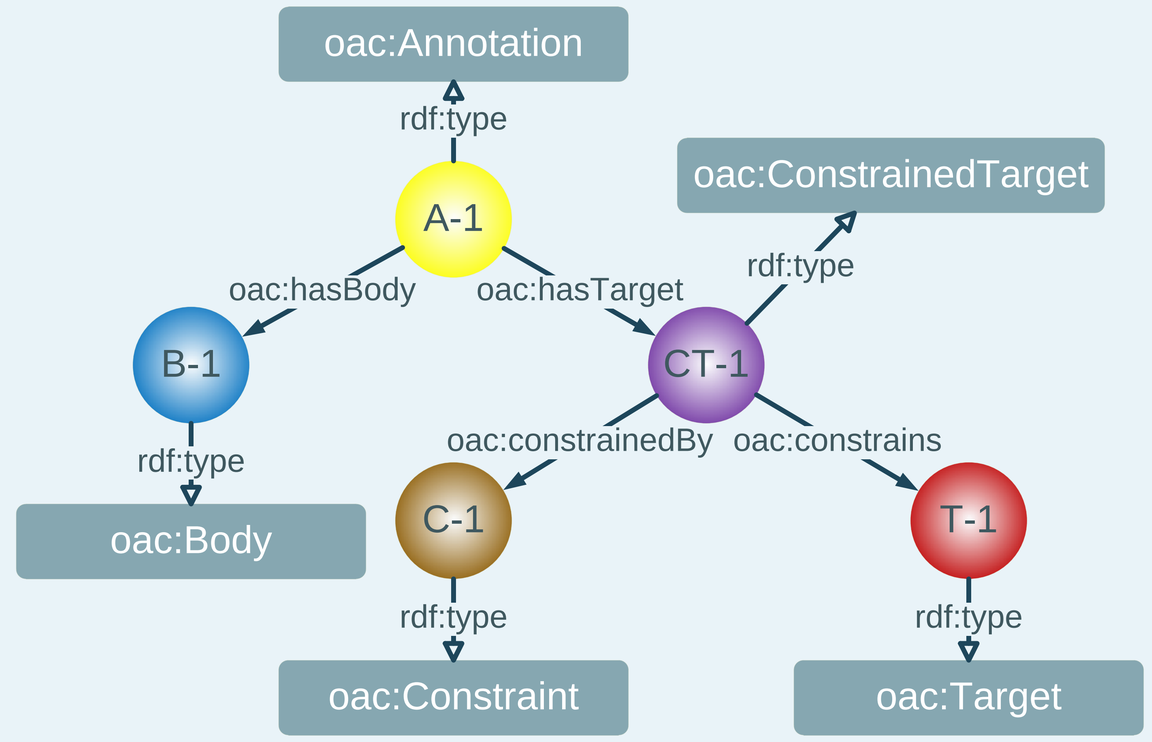
Based on cooperation of web information systems

1st step – shared repository for knowledge tags
2nd step – automatic maintenance of knowledge tags

The Knowledge Tag

- Lightweight annotation, which assigns short metadata to an information artifact
- Semantic relation to a tagged content
- Machine readable

Basic Structure of the Open Annotation Model



Knowledge Tag in Triple Stores

A-1	rdf:type	oac:Annotation
A-1	oac:hasBody	B1
A-1	oac:hasTarget	CT-1
B-1	rdf:type	oac:Body
CT-1	rdf:type	oac:ConstrainedTarget
CT-1	oac:constrainedBy	C-1
CT-1	oac:constrains	T-1
C-1	rdf:type	oac:Constraint
T-1	rdf:type	oac:Target

Triple Stores

+ Attributes accessibility
+ Inference possibilities
• Objects accessibility
- Scalability

Document Stores

+ Objects accessibility
+ Scalability
• Attributes accessibility
- Inference possibilities

Knowledge Tag in Document Stores

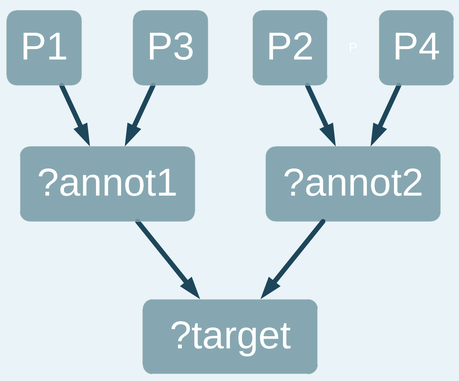
```
{
  _id:A-1, type:'oac:Annotation',
  hasBody:{ _id:A-1, type:'oac:Body' },
  hasTarget:{
    _id:CT-1, type:'oac:ConstrainedTarget',
    constrainedBy:{ _id:C-1, type:'oac:Constraint' },
    constrains:{ _id:T-1, type:'oac:Target' }
  }
}
```

Distributed SPARQL Processing

- Optimal join tree
- Iterative MapReduce(Filter) phase
  - **Map:** Maps values to attributes in patterns
  - **Reduce:** Joins mapped patterns
  - **Filter:** Filter out incomplete joins
- A conversion of results to output format

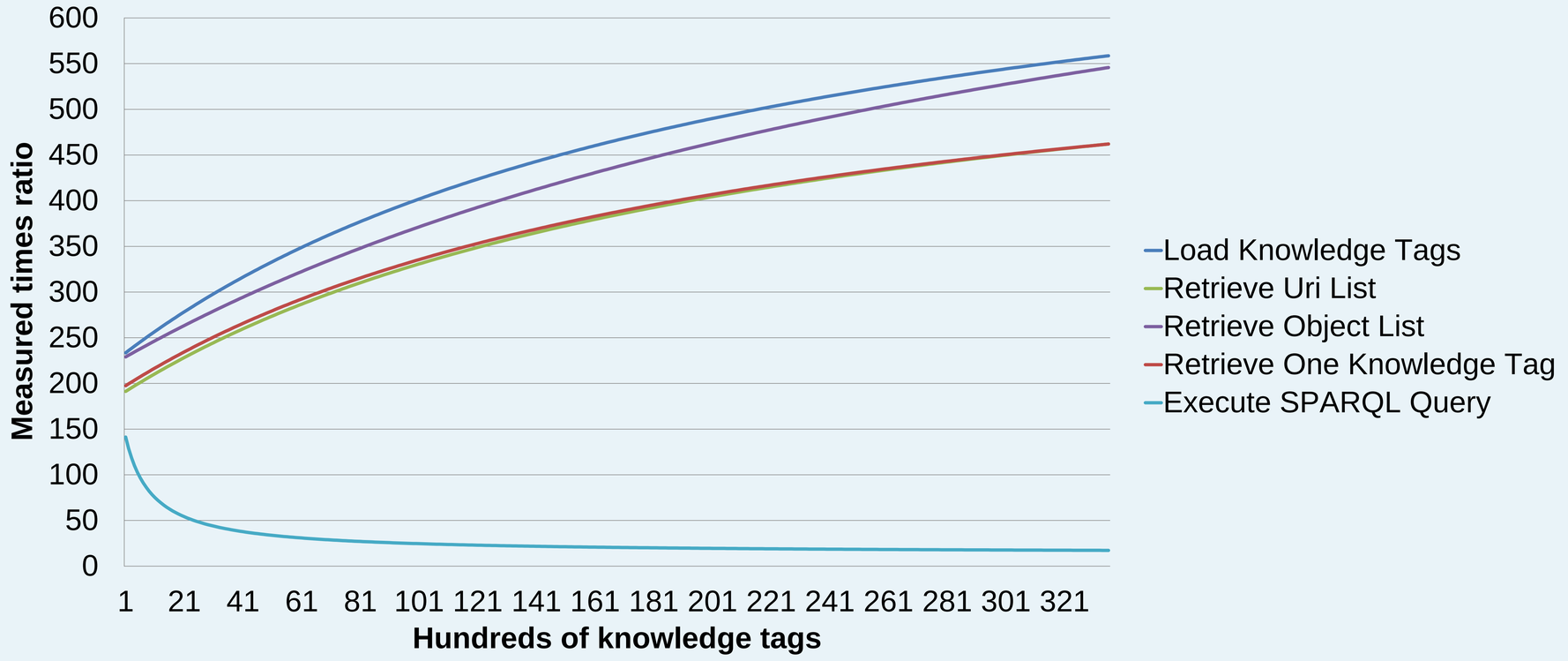
Function	Result
Map	{ key : (annot1[X]), value : ({ (P1), (annot1[X] | target[page.html]) }) }
	{ key : (annot1[X]), value : ({ (P3), (annot1[X] | creator1[John]) }) }
	{ key : (annot1[Y]), value : ({ (P1), (annot1[Y] | target[style.css]) }) }
Reduce	{ key : (annot1[X]), value : ({ (P1|P3), (annot1[X] | creator1[John] | target[page.html]) }
	{ key : (annot1[Y]), value : ({ (P1), (annot1[Y] | target[style.css]) }) }
Finalize	{ key : (annot1[X]), value : ({ (P1|P3), (annot1[X] | creator1[John] | target[page.html]) }

Optimal Join Tree



P1 – ?annot1 oac:hasTarget ?target
P2 – ?annot2 oac:hasTarget ?target
P3 – ?annot1 dcterms:creator ?creator1
P4 – ?annot2 dcterms:creator ?creator2

Comparison of the Realization via MongoDB and Bigdata



STU FIIT

Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislva