SPARQL Basic Graph Pattern Optimization Using Selectivity Estimation

Markus Stocker, Andy Seaborne,
Abraham Bernstein, Christoph Kiefer and Dave Reynolds

WWW2008, April 24, 2008, Beijing

Universität Zürich

DIS Dynamic and Distributed Information Systems

The Need for Query Optimization

„A simple rearrangement of a SPARQL query leads to an improvement of factor 220”

Jena Developers Mailing List, March 2006 [1]

„The [performance] result of the older version […] is over a hundred times lower“

BigOWLIM 0.9.5 vs. 0.9.6, November 2007 [2]

Query performance

matters!

Our Focus is on ...

Selectivity based
Basic Graph Pattern optimization
for *main memory* query engines

... and not on

Persistent triple stores
index structures
and performance evaluations

An Example

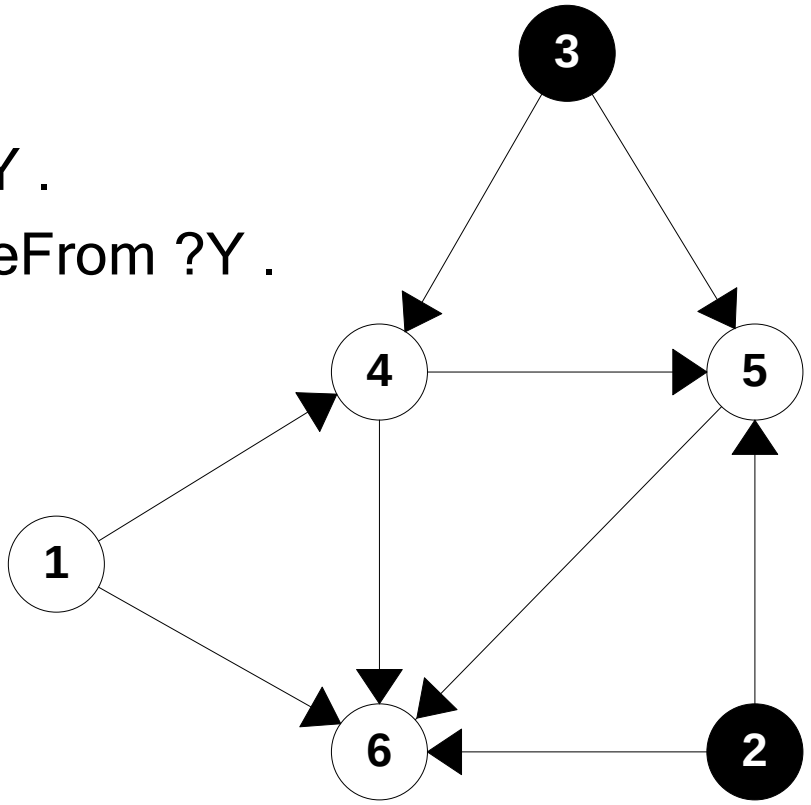
?x rdf:type uv:Person .

?x uv:socialSecurityNumber „555-05-7880“

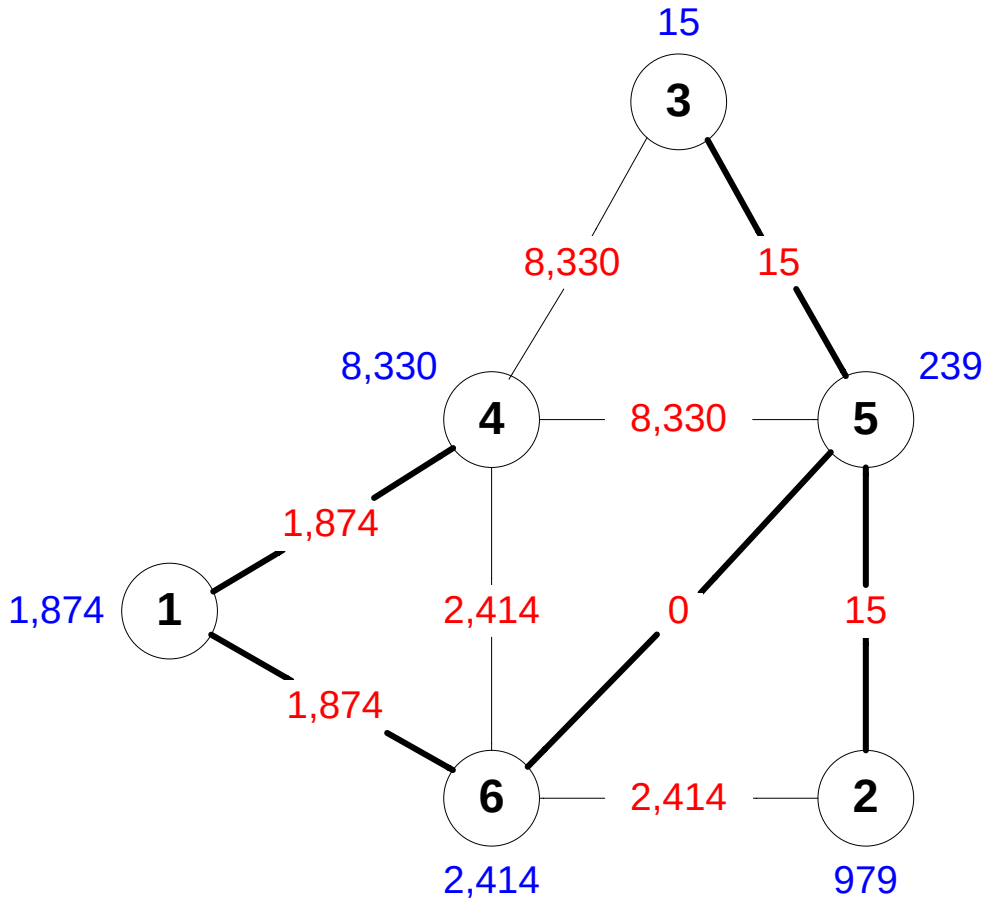
Our Approach

1 ?X rdf:type ub:GraduateStudent .
2 ?Y rdf:type ub:University .
3 ?Z rdf:type ub:Department .
4 ?X ub:memberOf ?Z .
5 ?Z ub:subOrganizationOf ?Y .
6 ?X ub:undergraduateDegreeFrom ?Y .

Original



Our Approach

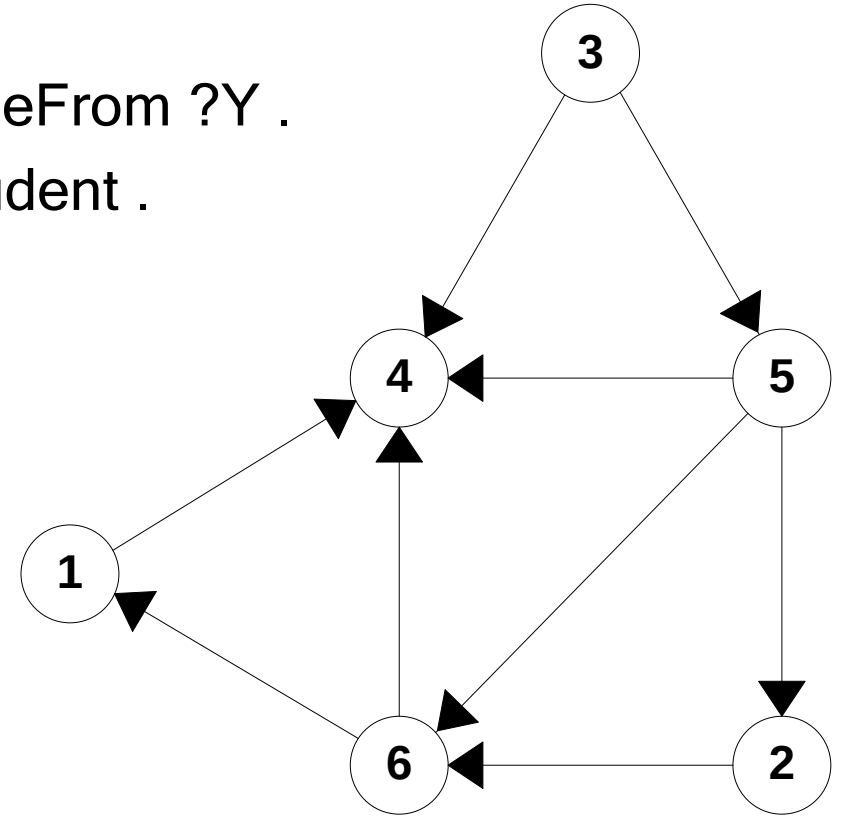


[3, 5, 2, 6, 1, 4]

Our Approach

3 ?Z rdf:type ub:Department .
5 ?Z ub:subOrganizationOf ?Y .
2 ?Y rdf:type ub:University .
6 ?X ub:undergraduateDegreeFrom ?Y .
1 ?X rdf:type ub:GraduateStudent .
4 ?X ub:memberOf ?Z .

Optimized



Q1 LUBM

Original

```
?X rdf:type          ub:GraduateStudent .
?X ub:takesCourse    ub:GraduateCourse0
```

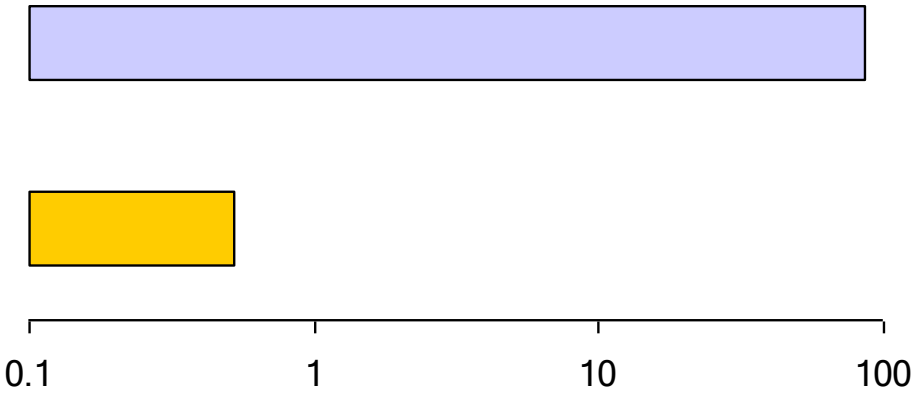
Optimized

```
?X ub:takesCourse    ub:GraduateCourse0 .
?X rdf:type          ub:GraduateStudent
```

Two Orders of Magnitude

Original 85.04 ms

Optimized 0.518 ms



164x

Q9 LUBM

Original

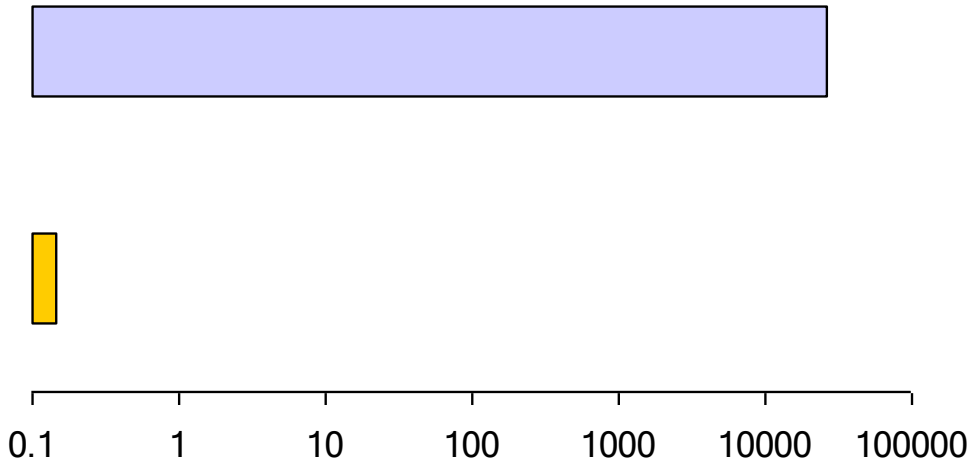
?X	rdf:type	ub:Student .
?Y	rdf:type	ub:Faculty .
?Z	rdf:type	ub:Course .
?X	ub:advisor	?Y .
?Y	ub:teacherOf	?Z .
?X	ub:takesCourse	?Z

Optimized

?Y	rdf:type	ub:Faculty .
?Y	ub:teacherOf	?Z .
?Z	rdf:type	ub:Course .
?X	ub:advisor	?Y .
?X	rdf:type	ub:Student .
?X	ub:takesCourse	?Z

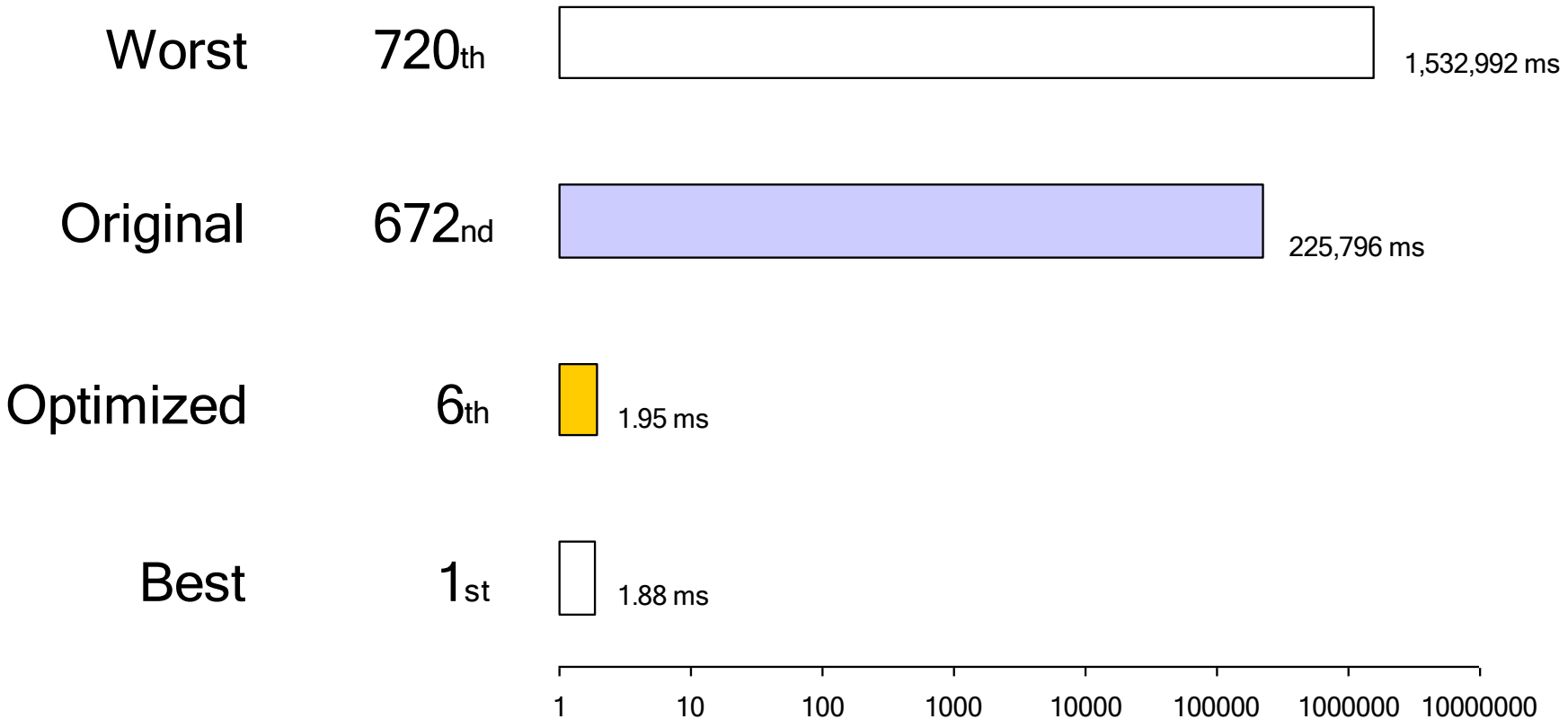
Five Orders of Magnitude

Original	26,342 s
Optimized	0.1442 s



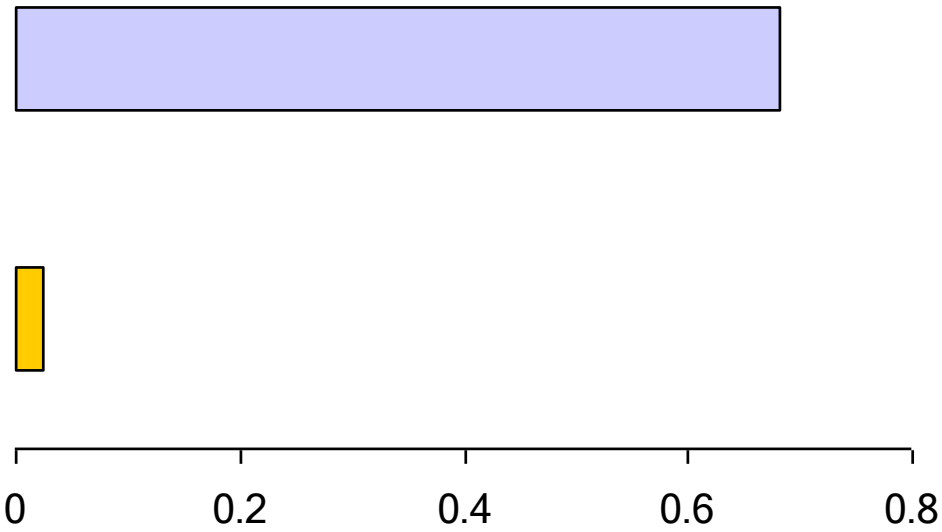
182,676x

Q2 LUBM Plan Space



672nd vs. 6th

Plan	Distance
Original	0.681
Optimized	0.023



833 vs. 14

Limitations and Conclusions

Scaling, SPARQL syntax, Synthetic dataset

Near optimal plan

Thank you

Questions?

References

[1] http://tech.groups.yahoo.com/group/jena-dev/message/21436

[2] A. Kiryakov et al. Measurable Targets for Scalable Reasoning. Technical Report. Ontotext Lab, Sirma Group Corp. November 2007