

Implementation and Optimization of Queries in XSPARQL

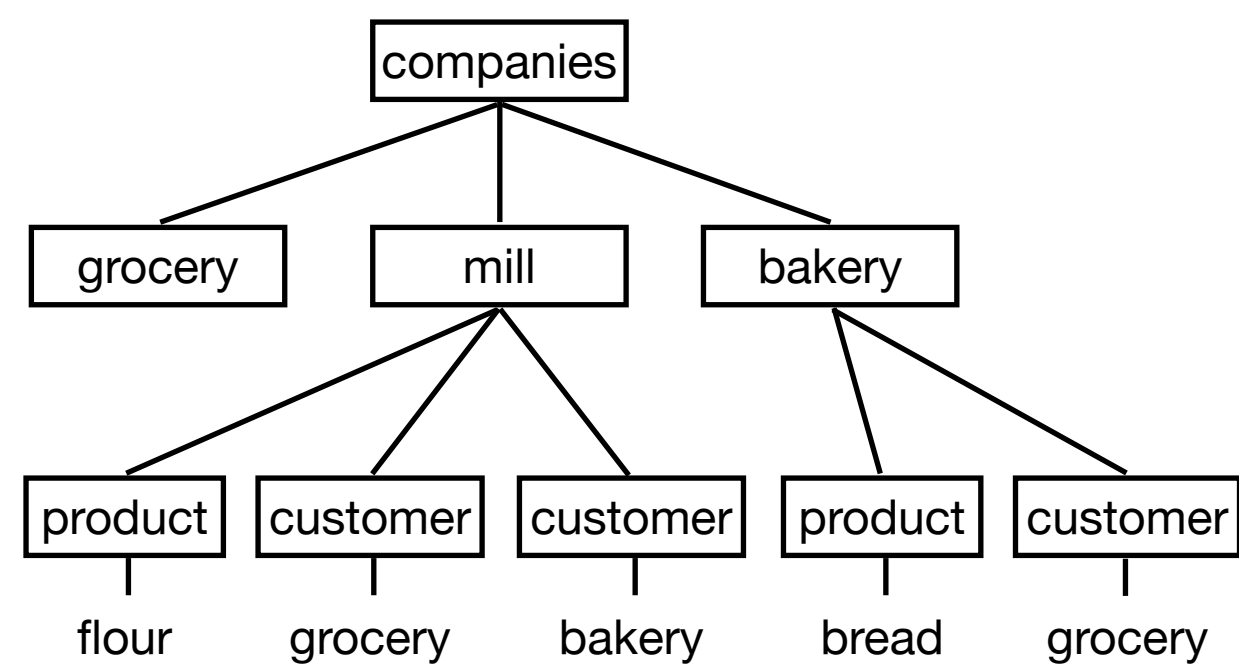
Masterstudium:
Information & Knowledge Management

Stefan Bischof

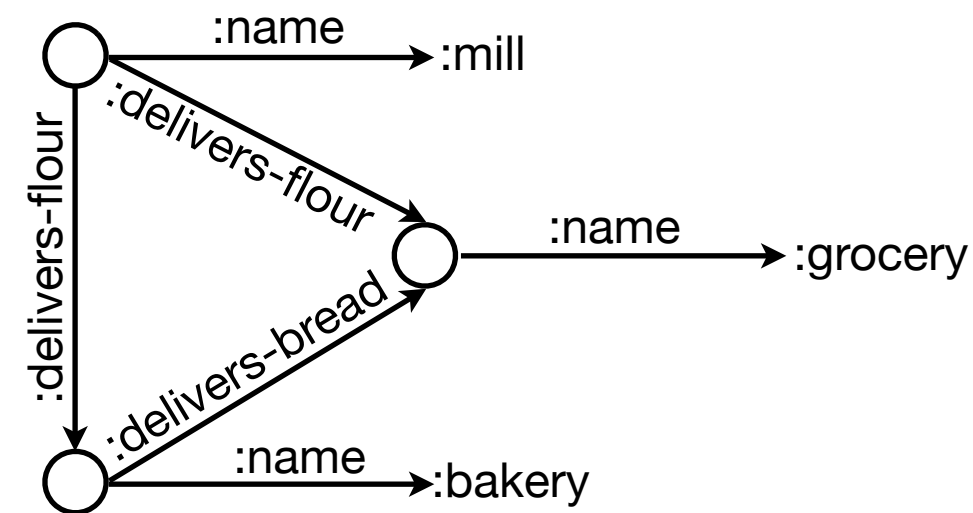
Technische Universität Wien
Institut für Informationssysteme
Arbeitsbereich: Knowledge Based Systems
Betreuer: Univ.-Prof. Dr. Thomas Eiter

Data on the Web: XML and RDF

XML (Extensible Markup Language) is a widely used data exchange format using a tree representation



RDF (Resource Description Framework) is the Semantic Web data format based on directed graphs



XQuery is a functional XML query language

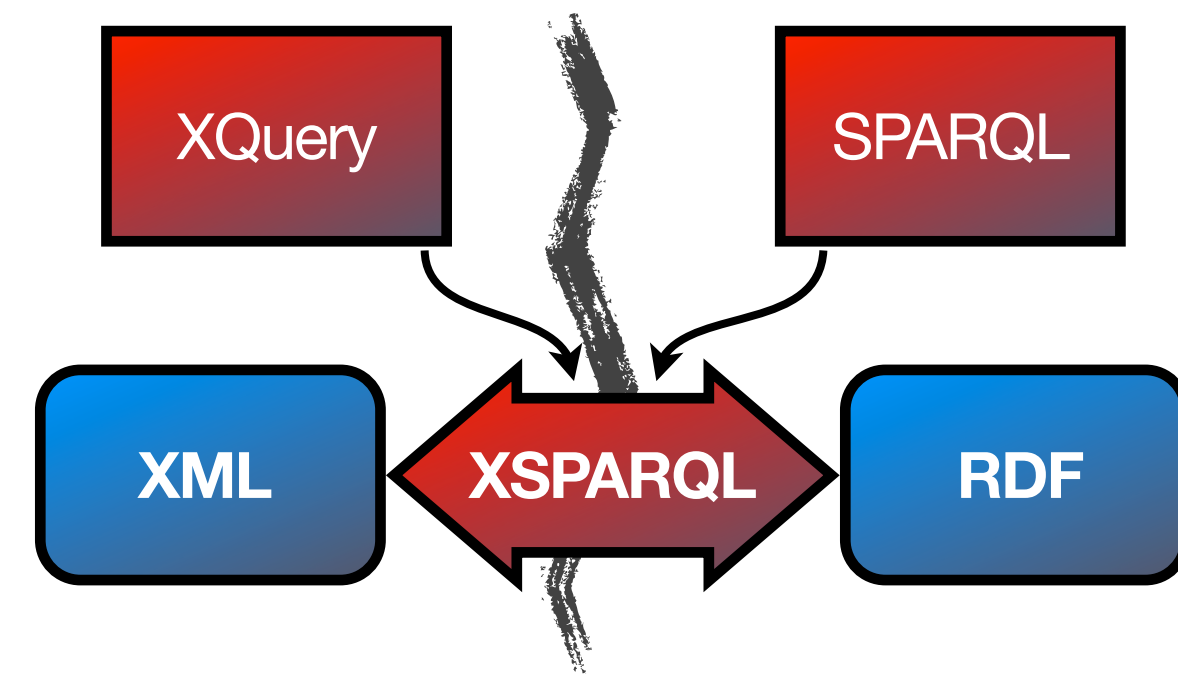
SPARQL uses graph patterns to query RDF

Data Integration and Interoperability

- How to convert data between XML and RDF?
- How to access XML and RDF data sources at the same time?
- How to integrate XML and RDF data?
- XML tools are not suited to process graph based data

XSPARQL: Bridging the gap of XML and RDF

- XSPARQL combines the strengths of both query languages XQuery and SPARQL
- Based on XQuery, imports parts of SPARQL
- Provides XQuery's function library to SPARQL
- Provides SPARQL's graph pattern matching facility to XQuery



Issues of XSPARQL

- XSPARQL semantics shows unintended behavior when reusing variables of SPARQL parts
- Query evaluation can be very slow for SPARQL parts in loops
- Implementation is not suited to implement new features and optimizations easily

Thesis Contributions

- Extend XSPARQL by new features to simplify query authoring
- Improve query evaluation time for XSPARQL queries with XDEP optimization
- Create a new, maintainable, platform independent, implementation

1 XDEP: Dependent Join Optimization

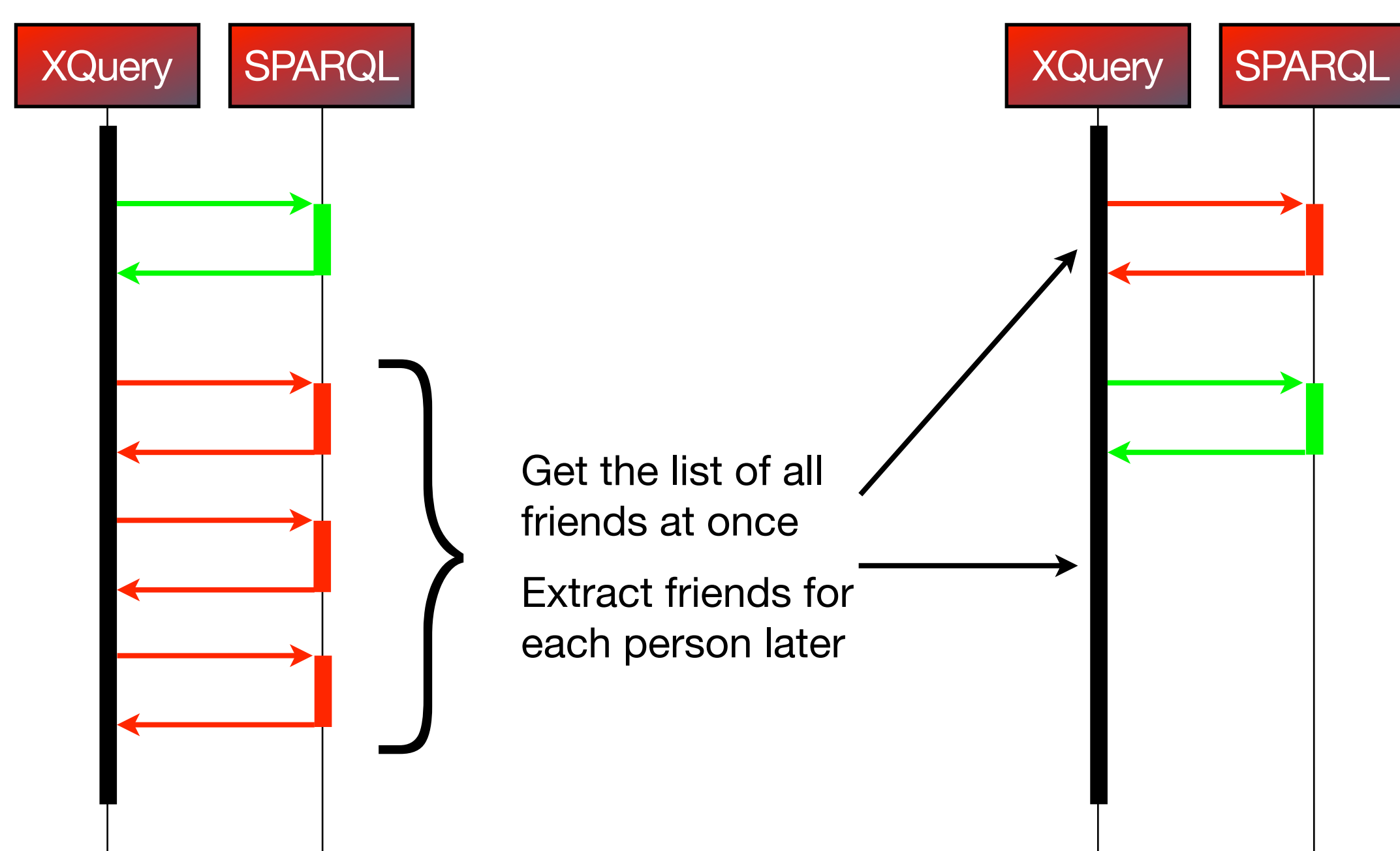
- Goal: Optimize interface between XQuery engine and SPARQL engine
- Minimize communication overhead by reducing the number of interactions
- Standard XSPARQL: call SPARQL engine N times (outer iterations) for inner loop
- XDEP: get the results of all inner queries at once, joins results later in XQuery

Standard XSPARQL

```
Get a list of persons LP
For every person P of LP do
  Print P's name
  Get a list of P's friends LF
  For every person F of LF do
    Print F's name
```

XDEP Optimization

```
Get a list of all friends LAF
Get a list of persons LP
For every person P of LP do
  Print P's name
  Join P with LAF -> LF
  For every person F of LF do
    Print F's name
```



2 New XSPARQL Features

Constructed Dataset

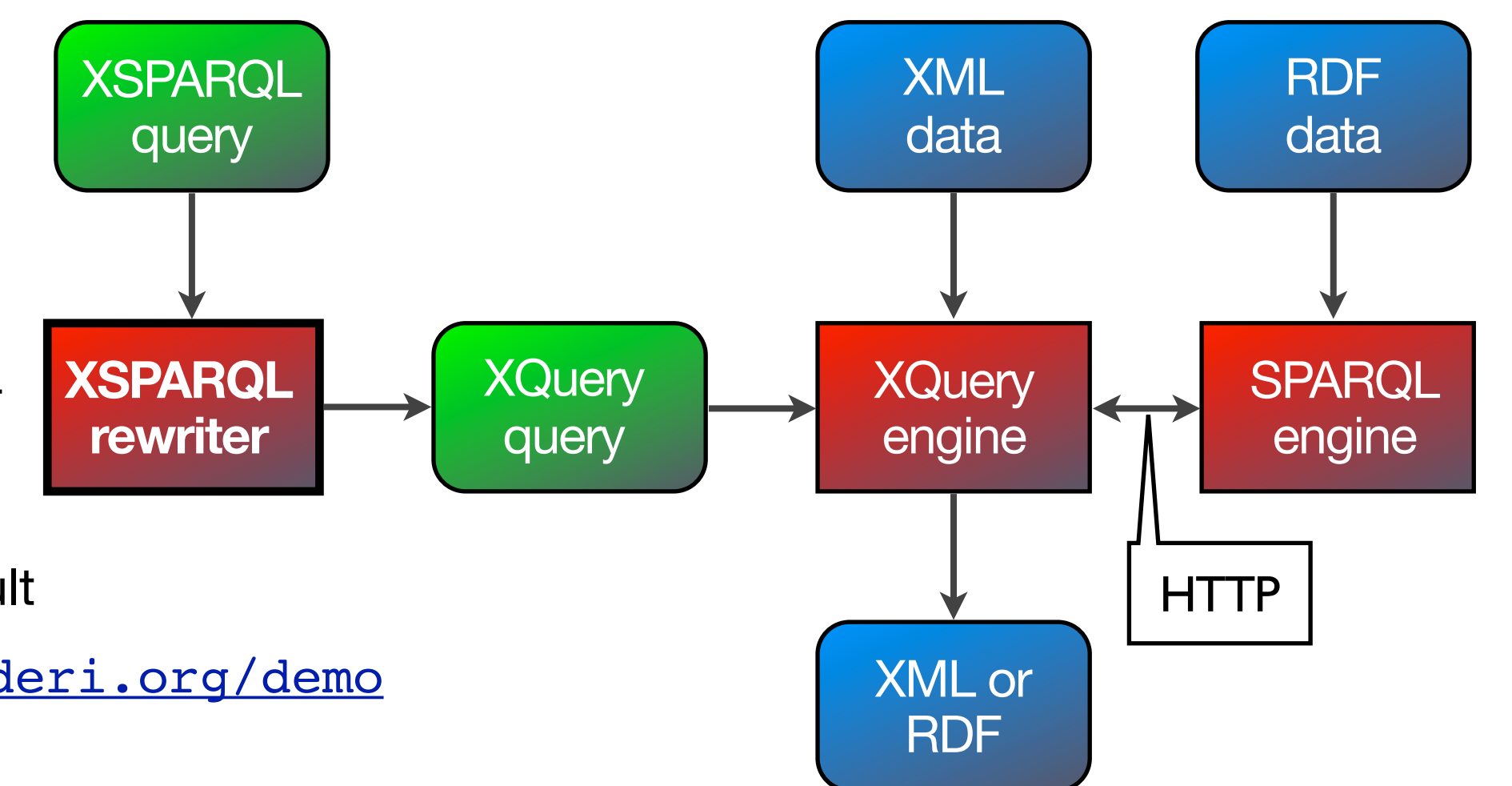
- Create temporary RDF graphs to be used in the same query again
- Allows to query aggregated values or manually optimize queries by preselecting data

Dataset Scoping

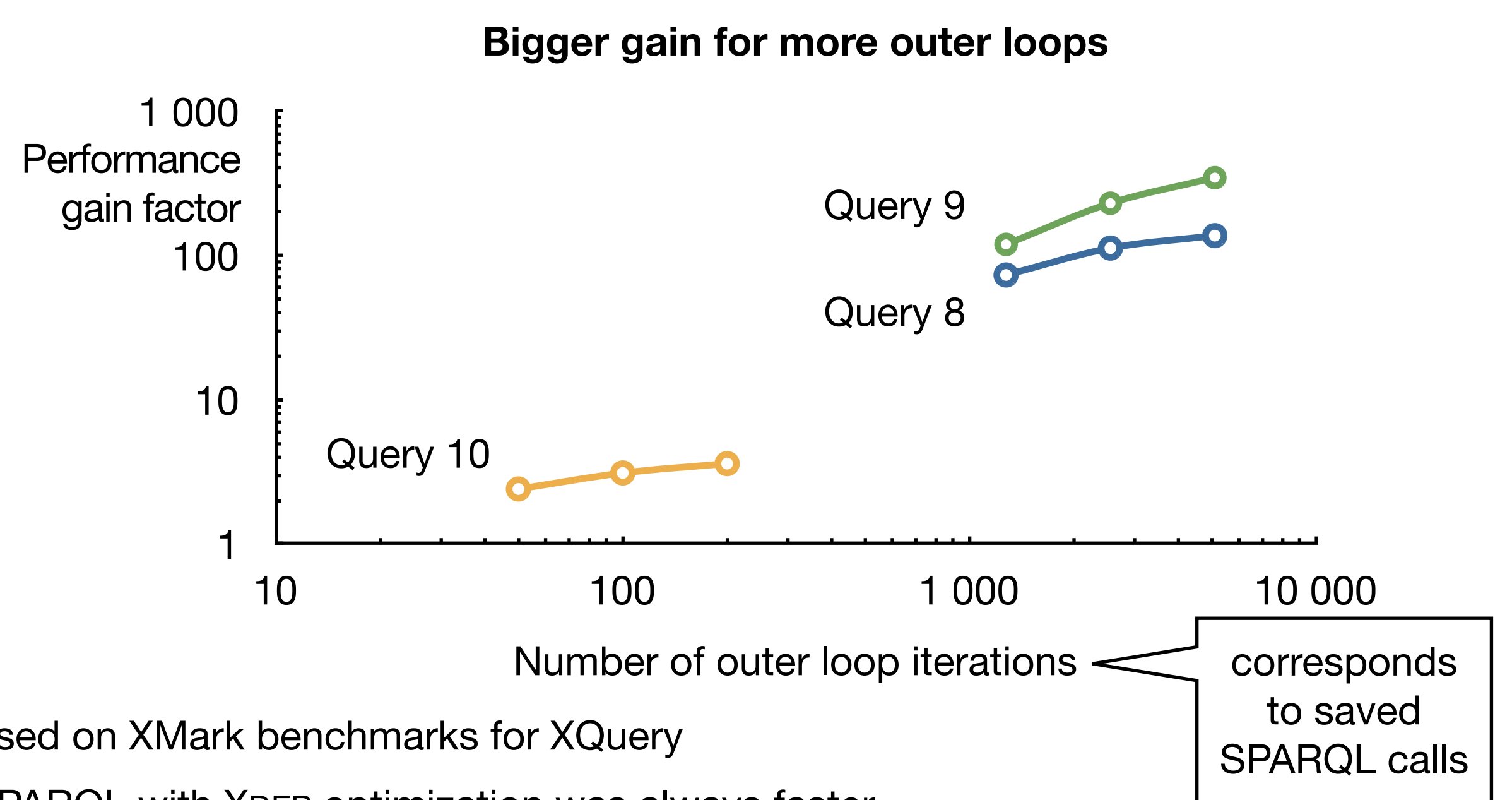
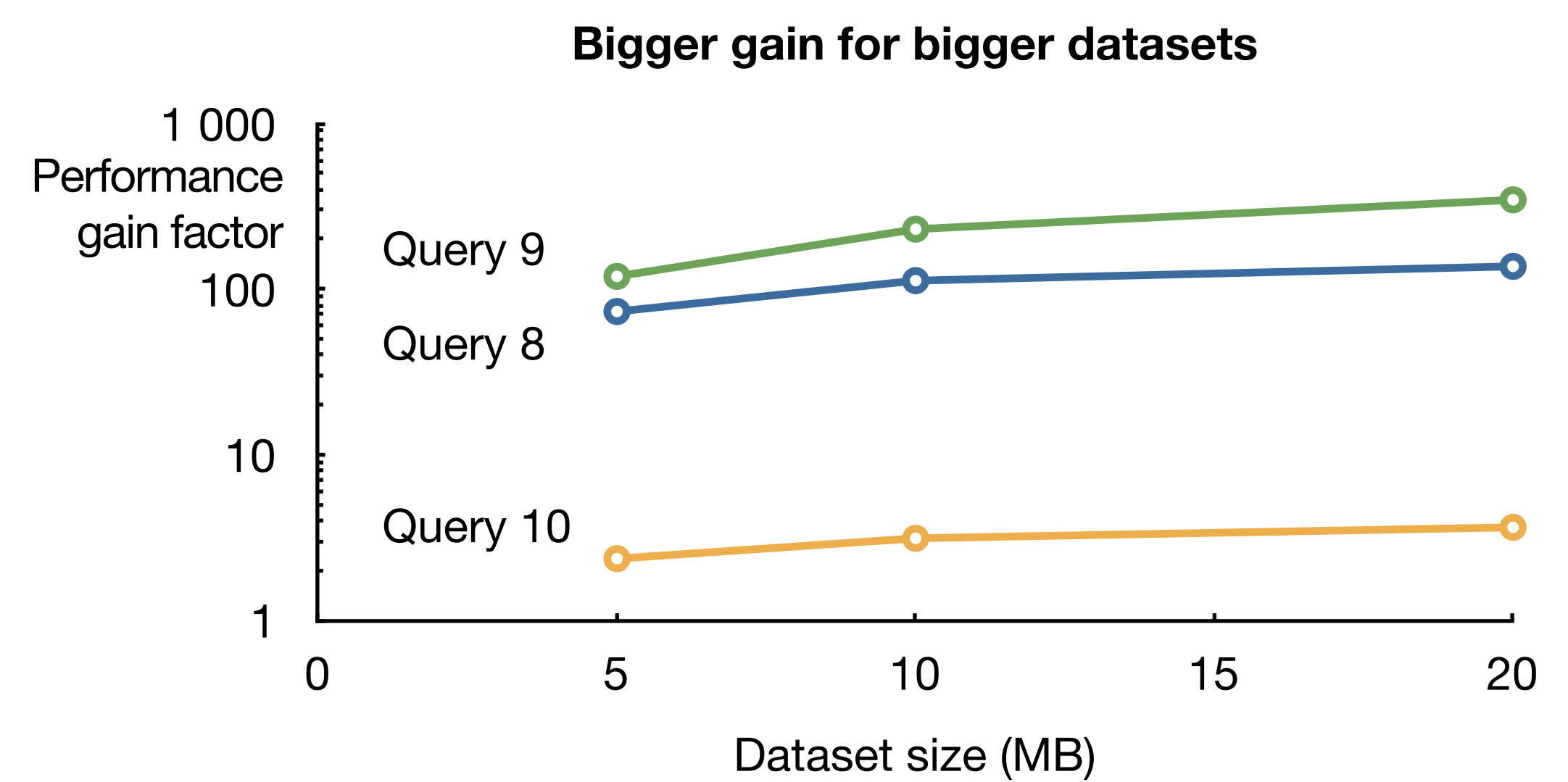
- Provides an optional fix for unintended effects when evaluating complex queries

3 Implementation Architecture

- Rewrite XSPARQL query to XQuery query
- Evaluate XQuery query
- Process XML data directly
- Process RDF: Send SPARQL query to SPARQL engine, return result as XML
- Create final XML or RDF result
- Try it at <http://xsparql.deri.org/demo>



4 Evaluation: Standard XSPARQL vs. XDEP



- Based on XMark benchmarks for XQuery
- XSPARQL with XDEP optimization was always faster

Conclusions

- New features simplify writing complex queries and lead to more predictable results
- New implementation increases maintainability and makes adaption easier
- XDEP optimization targets a performance bottleneck of the architecture
- XDEP performance gain increases with number of saved SPARQL calls

More information at <http://xsparql.deri.org/>

References

- Akhtar W., Kopecký J., Krennwallner T., and Pollers A. *XSPARQL: Traveling between the XML and RDF worlds - and Avoiding the XSLT pilgrimage*. In 5th European Semantic Web Conference (ESWC2008), pages 432–447, 2008.
- Bischof S. *Implementation and Optimization of Queries in XSPARQL*. Master's Thesis, Vienna University of Technology, 2010.
- Bray T., Paoli J., Sperberg-McQueen C. M., Maler E., and Yergeau F. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. <http://www.w3.org/TR/xml/>, November 2008. W3C Recommendation.
- Boag S., Chamberlin D., Fernández M. F., Florescu D., Robie J., and Siméon J. *XQuery 1.0: An XML Query Language*. <http://www.w3.org/TR/xquery/>, January 2007. W3C Recommendation.
- Manola F. and Miller E. *RDF Primer*. <http://www.w3.org/TR/rdf-primer/>, February 2004. W3C Recommendation.
- Prud'hommeaux E. and Seaborne A. *SPARQL Query Language for RDF*. <http://www.w3.org/TR/rdf-sparql-query/>, January 2008. W3C Recommendation.