Geo Module

Masoumeh Seydi

DBIS Uni Konstanz

Aug. 2013

Overview

- Introduction

- Geo Module in BaseX

- Spatial Index

- Performance

- Conclusion

Motivation

Google Earth
650,675 followers on Google+

Widely used nowadays

Disaster Management

Geo Spatial Data

GML

WKB

WKT

KML

...

- Geographic properties
- Coordinates and topologies

GML, extensive XML based language:

- W3C XML Schema Language
- Developed by OGC

Why GML?

Compete ??!!!

`<?xml?>`

- Dynamic spatial data sources

- Schema changes

- Fancy semantic issues (data & schema)

- Integration of all geographic information
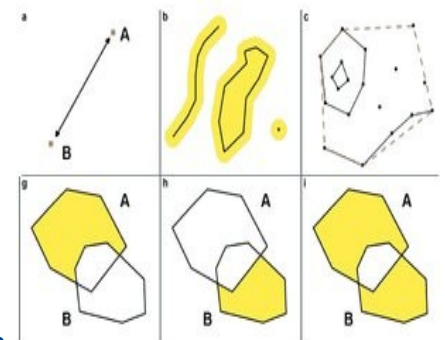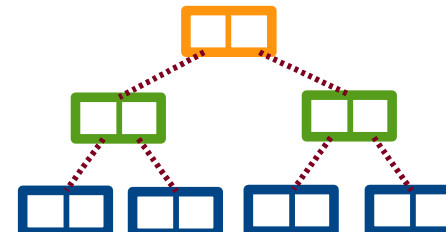
- No default coordinate system

EXpath Geo Module Specification

- Geospatial API function interface

- Functions conforming OGC

- Namespace conventions

- Error management
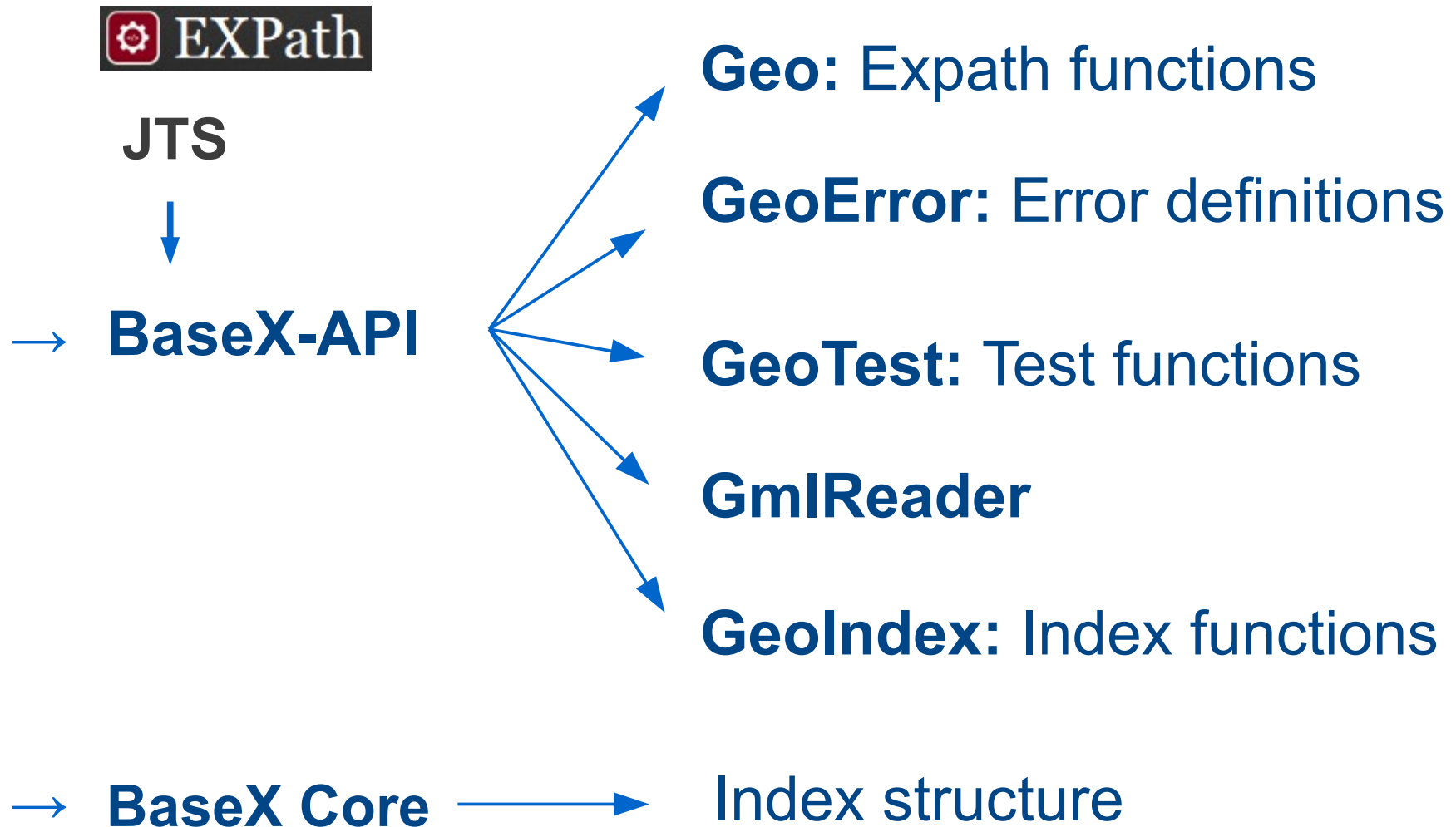
- Xquery or XSLT functions

JTS Topology Suite
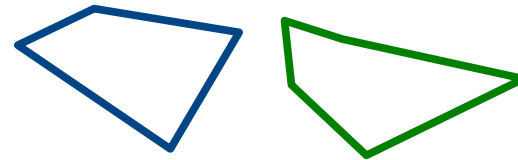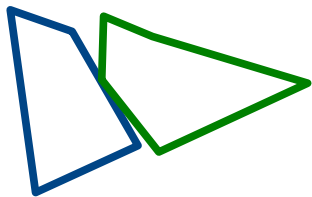
Geo Module in BaseX

EXPath

JTS

↓

→ **BaseX-API**

Geo: Expath functions

GeoError: Error definitions

GeoTest: Test functions

GmlReader

GeoIndex: Index functions

→ **BaseX Core** → Index structure

Geo Module Functionality

<gml:Polygon> …. </gml:Polygon>

<gml:Polygon> …. </gml:Polygon>

Read

Geometric operation
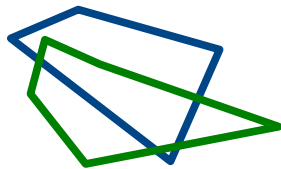
touches   **intersects**   **boundary**   **distance**   ....
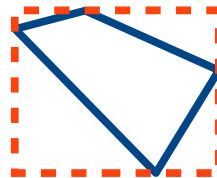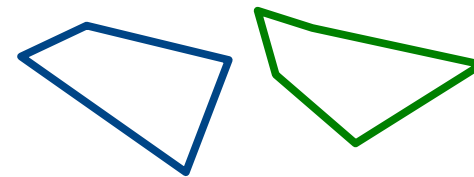
Outputs
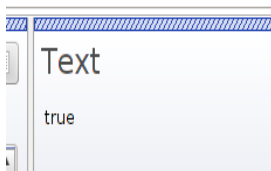
Text
true

Text
<gml:LineString xmlns:gml="ht
  <gml:coordinates>1.0,1.0 55.
coordinates>
</gml:LineString>

Text
10.950184911000926

9

Geo Module in Practice

let $x := <gml:Polygon>

                          <gml:outerBoundaryIs>… </gml:outerBoundaryIs>

             <gml:Polygon>

for $p in //gml:Polygon

return if (intersects($x, $p)) then $p else ()

**Unnecessary readings,
big time complexity**

Geo Module in Practice

let $x := <gml:Polygon>

   <gml:outerBoundaryIs>… </gml:outerBoundaryIs>

  <gml:Polygon>

for $p in DB//gml:Polygon

return if (intersects($x, $p)) then $p else ()

Unnecessary readings,
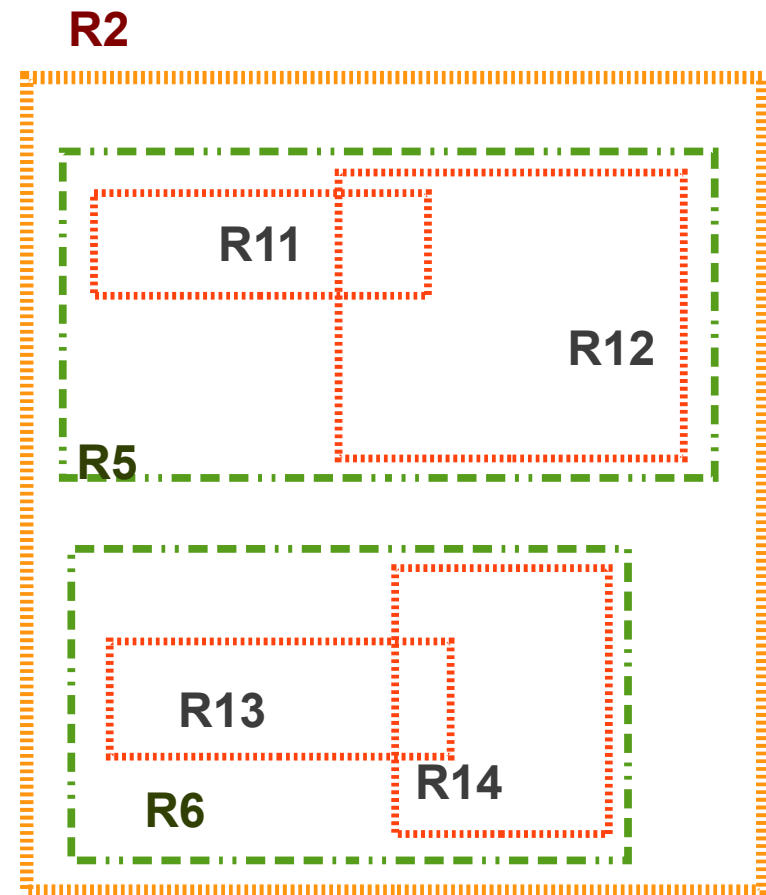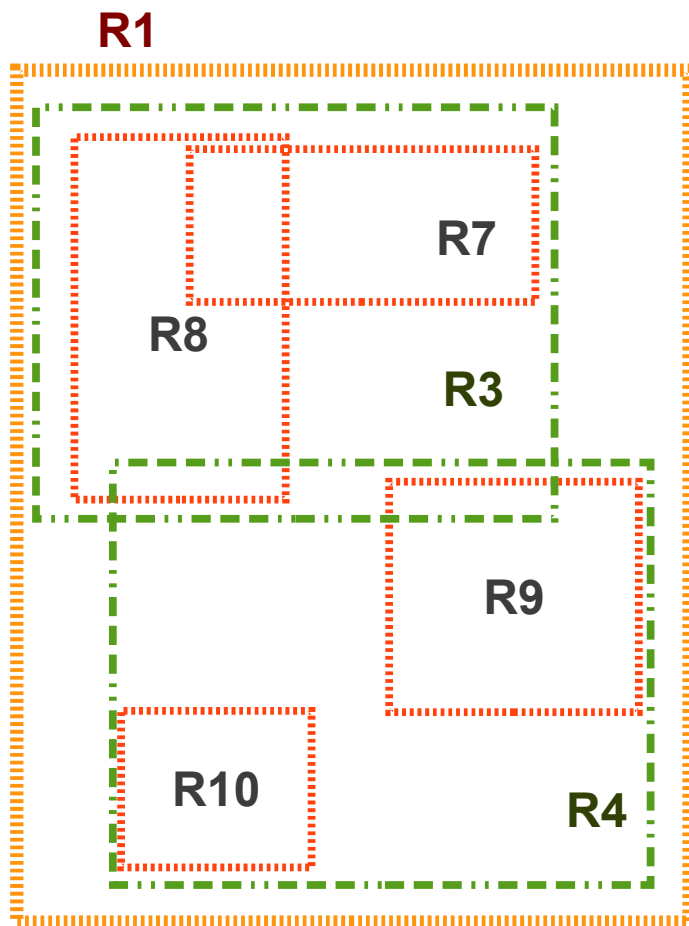**big time complexity** ⟶ **Indexing**
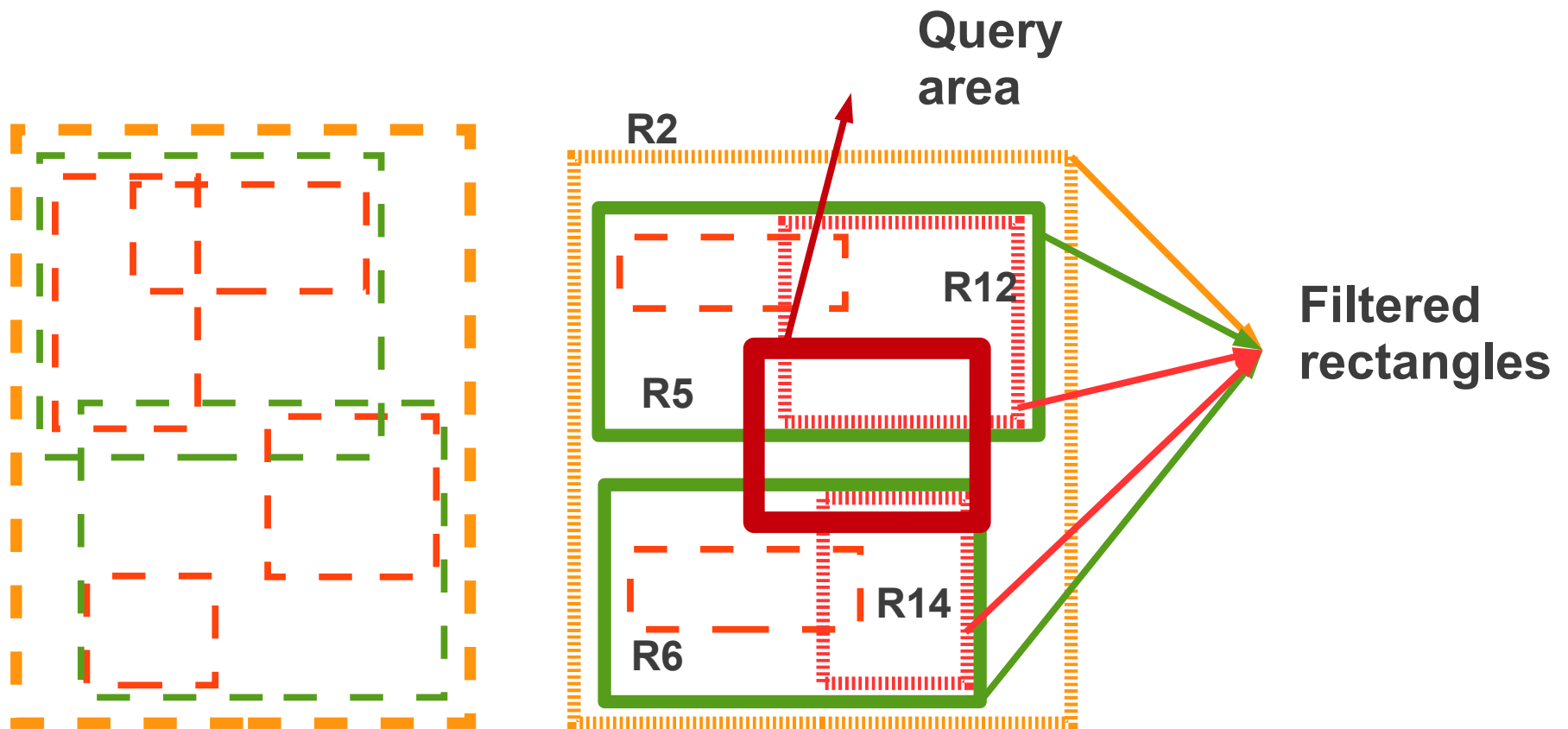
Spatial Index

- R-Tree, Antonin Guttman in 1984
    - Common indexing technique
    - Some disadvantages
- STR-Tree, Leutenegger et al. In 1997
    - → Improved performance of R-Tree
- JTS STRtree:
    - Serializable
    - Two step filtering

R-Tree

Intersection queries

R-Tree (Cont'd)

Query area

R2

R12

R5

R14

R6

Filtered rectangles

R12, R14 → **chosen objects**

14

Indexing

Once:

a.gml

....

......

read

Build a
STRtree

serialize

a.file

Pre
values

Query time:

a.file

Pre
values

**Read to
memory**

Filtering

**Geometric
operation**

15
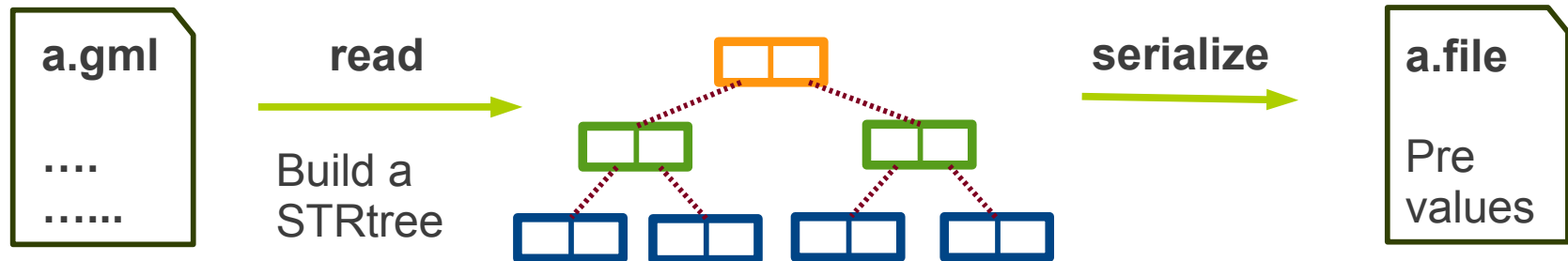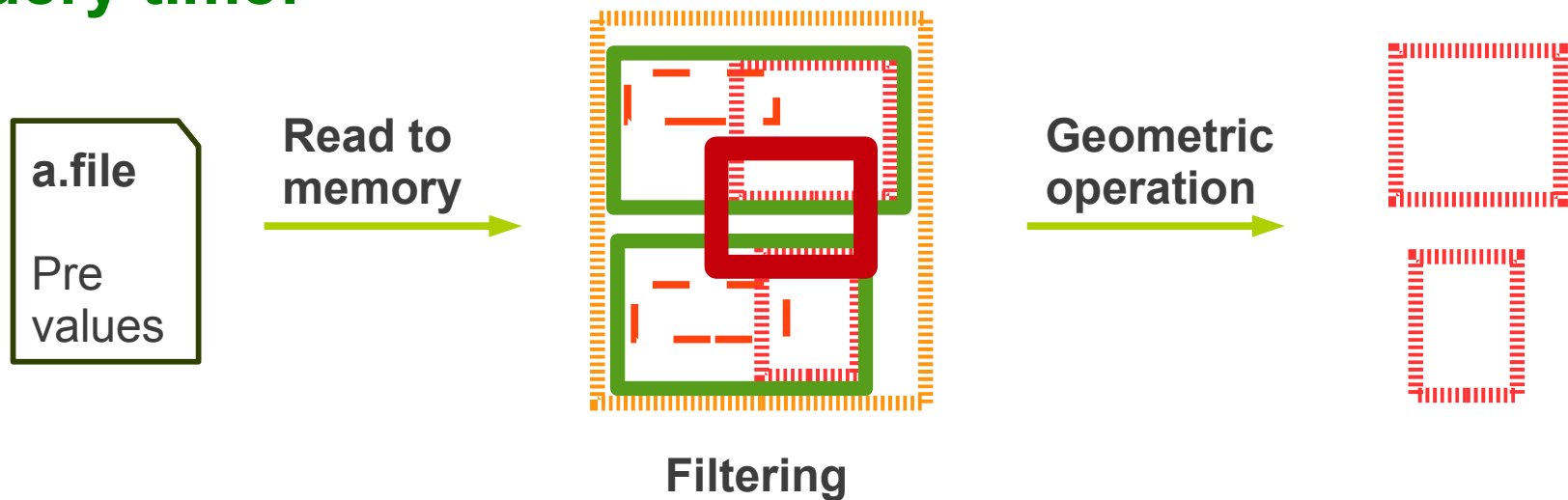
Indexing: GeoIndex Module

- Index functions in a separated module

let $a:= <gml:Polygon>... </gml:Polygon>

return **geo-index**:**intersects("DB", $a)**

- **New function implementations, using index structure**
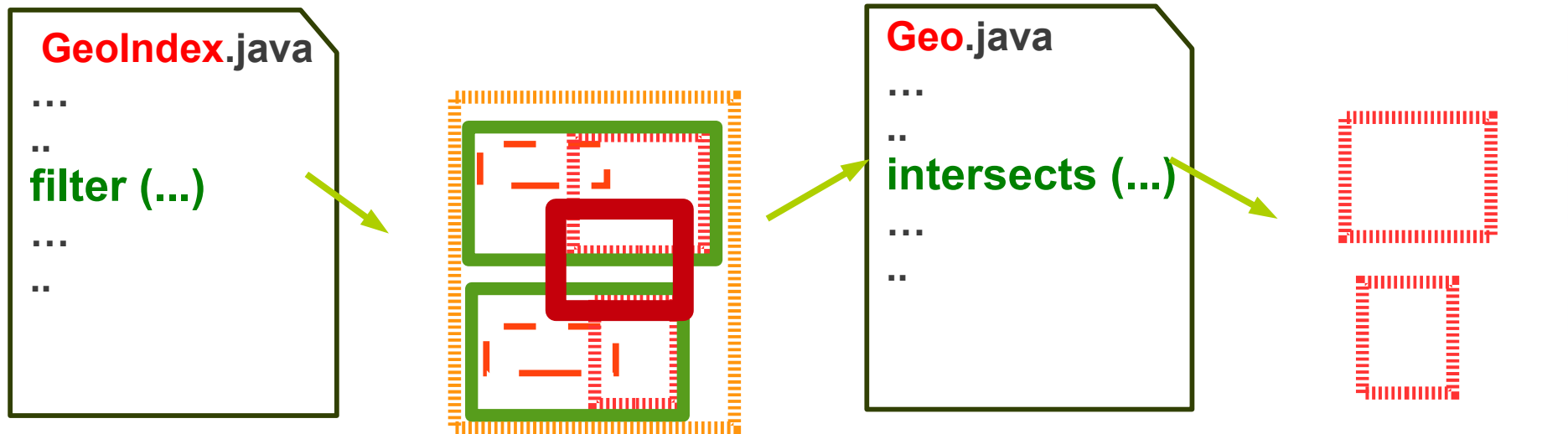- **Contains filtering**
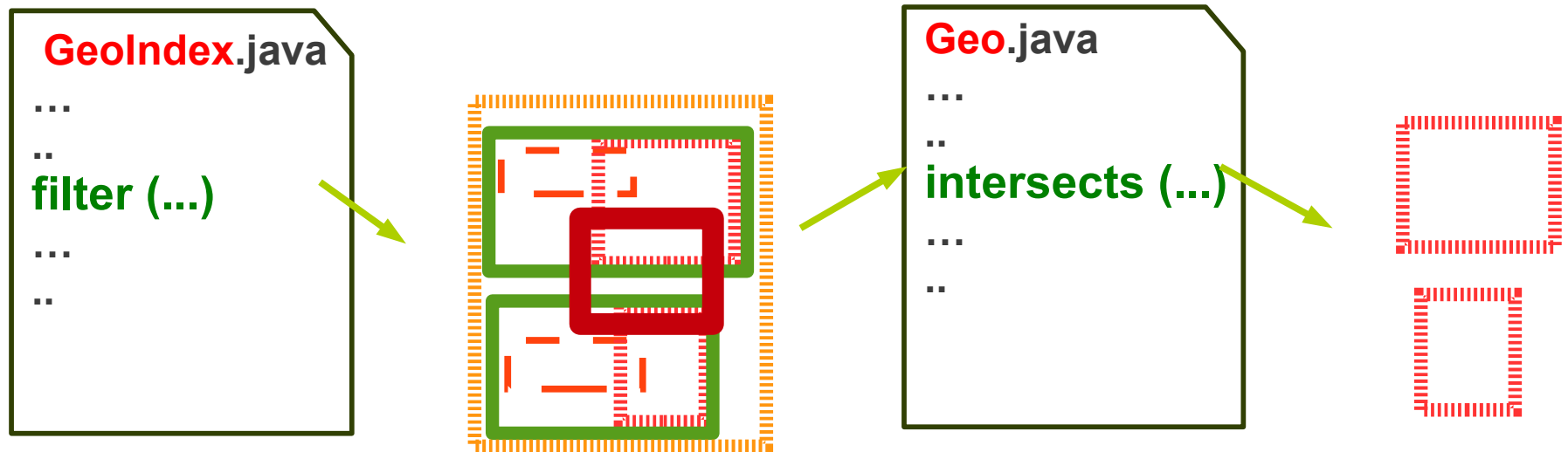
Implicit Indexing

let $a:= <gml:Polygon> … </gml:Polygon>

for $x in **geo-index:filter("DB", $a)** → **Filtering func.**

return if (**geo:intersects($a, $x)**) then $x else () → **Original geo function**

17

Implicit Indexing

GeoIndex.java
…
..
filter (...)
…
..

Geo.java
…
..
intersects (...)
…
..

let $a:= <gml:Polygon> … </gml:Polygon>

for $x in **geo-index:filter("DB", $a)** → **Filtering func.**

return if (**geo:intersects($a, $x)**) then $x else () → **Original geo function**

Worse performance → Map (cashing)

Map (Cashing)

- Prevents creation of the same geometry in each function call

    for $x in geo-index:filter("DB", $polygon)

    return if (geo:**intersects($polygon, $x)**) then $x

    else ()   **Every time $polygon is
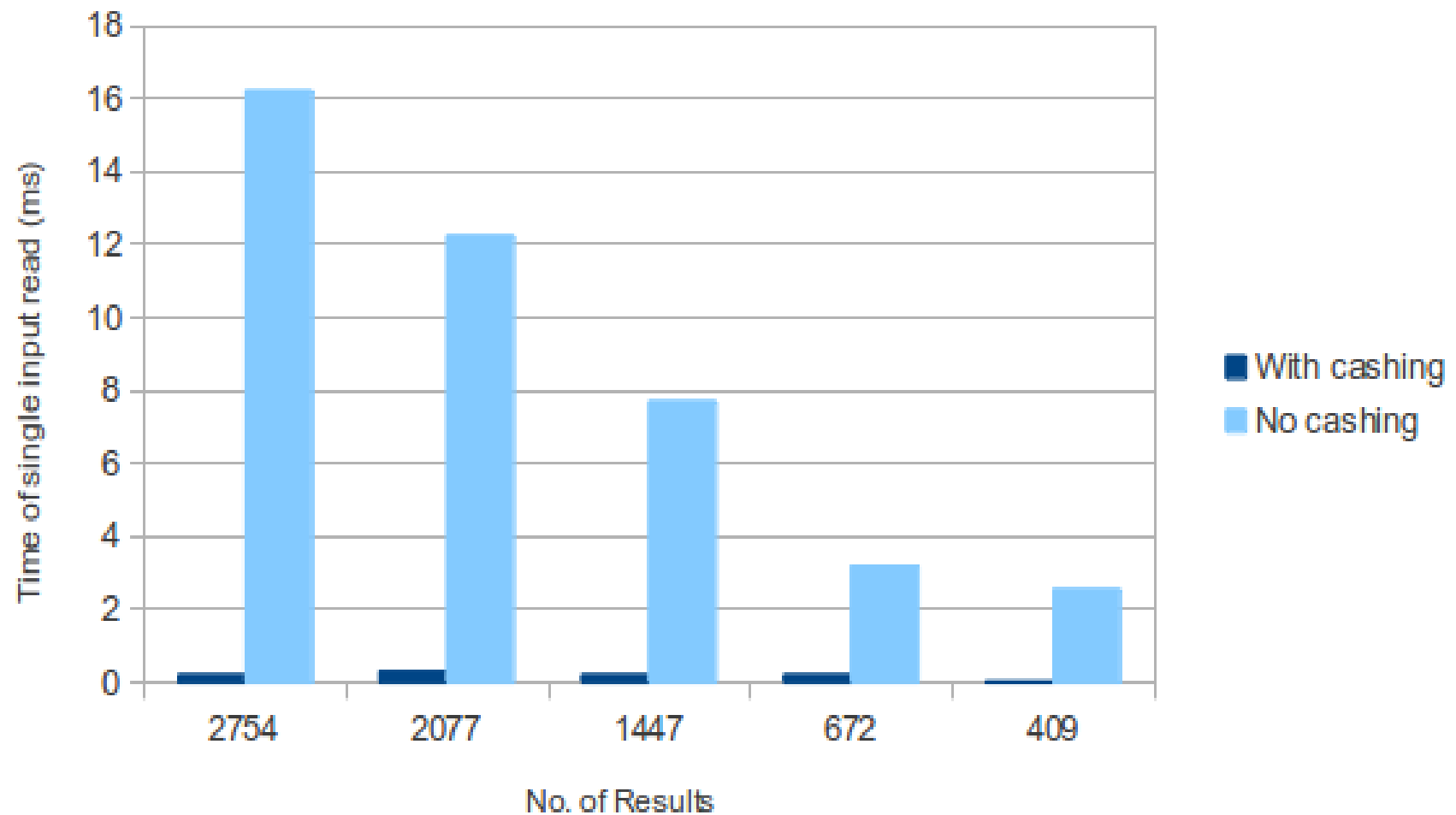   created**

- Used in Geo functions

    → Shorter query time

19

Map (Cont'd)

Query Examples

- Data
  - Netherlands Map in GML 2.0
  - From University of Twente, Dept of Geoinformation Processing
- Visualization
  - Automatic conversion from GML to KML
  - Google Earth

Query Examples (Cont'd)

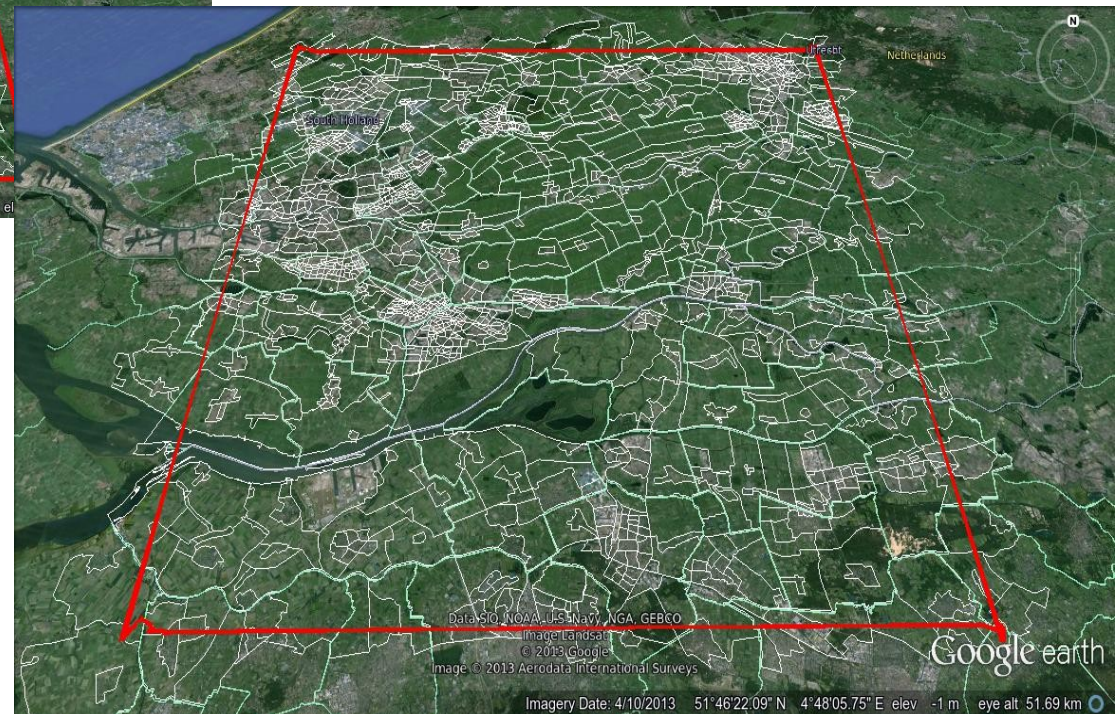**A set of geometries and a query geometry**
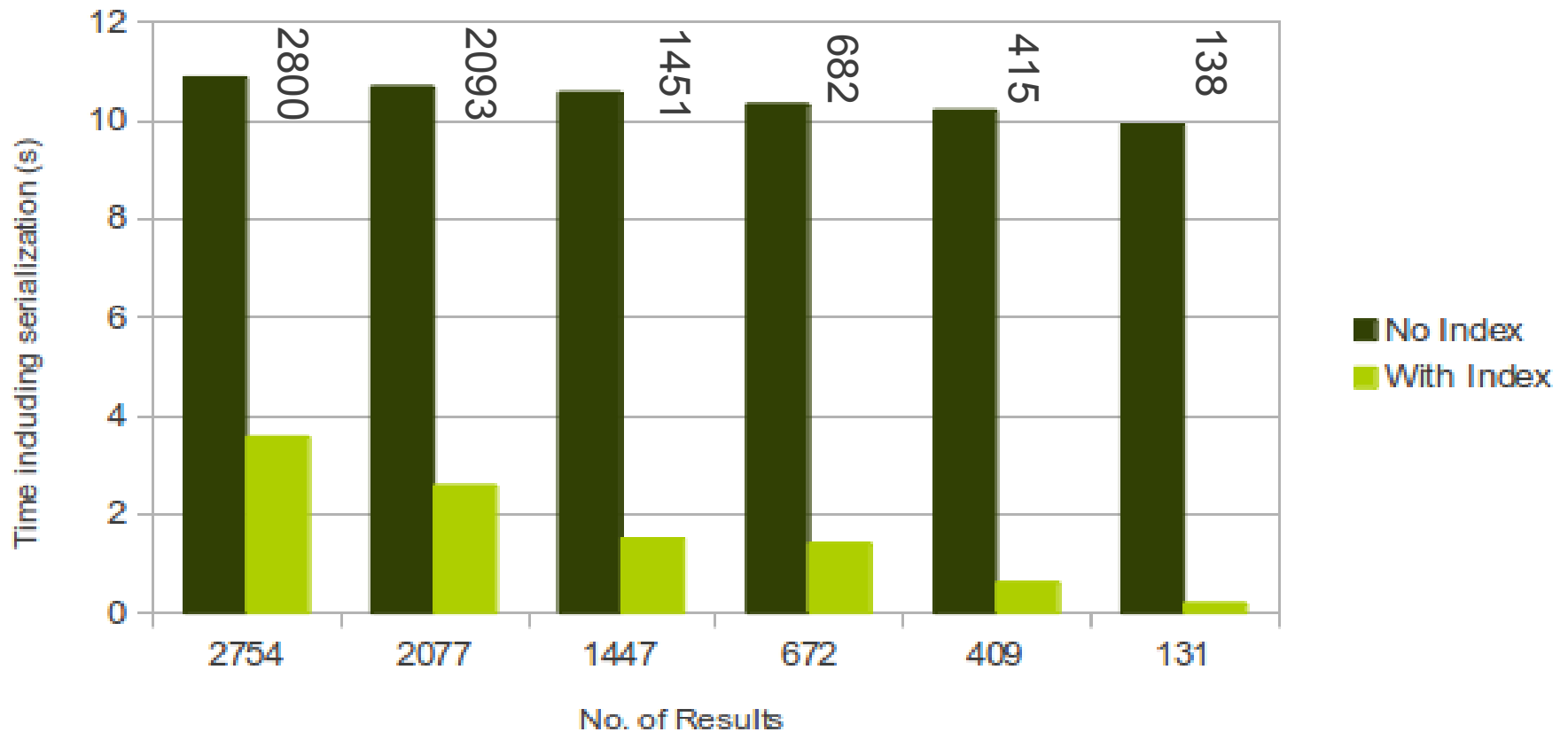
overlaps

Query Examples (cont'd)

contains

intersects

Index Efficiency

- Better performance, not satisfying ...

GML Reader

Serializing the XML node

⤷ String
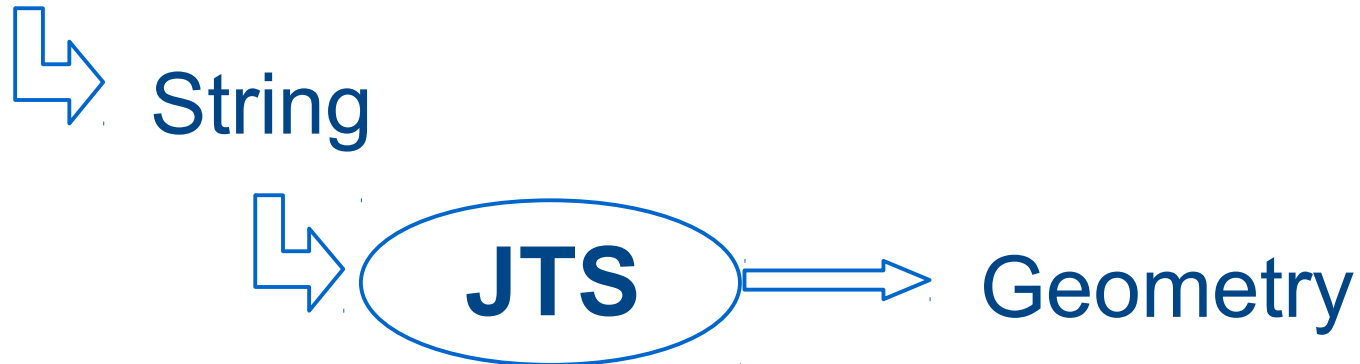
⤷ (**JTS**) ⟹ Geometry

GML Reader

Serializing the XML node

↳ String   **Parse**
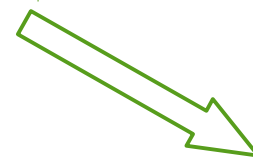
↳ (**JTS**) ⟹ Geometry

BaseX node ⟹ **Geometry**

GML Reader: JTS vs Custom
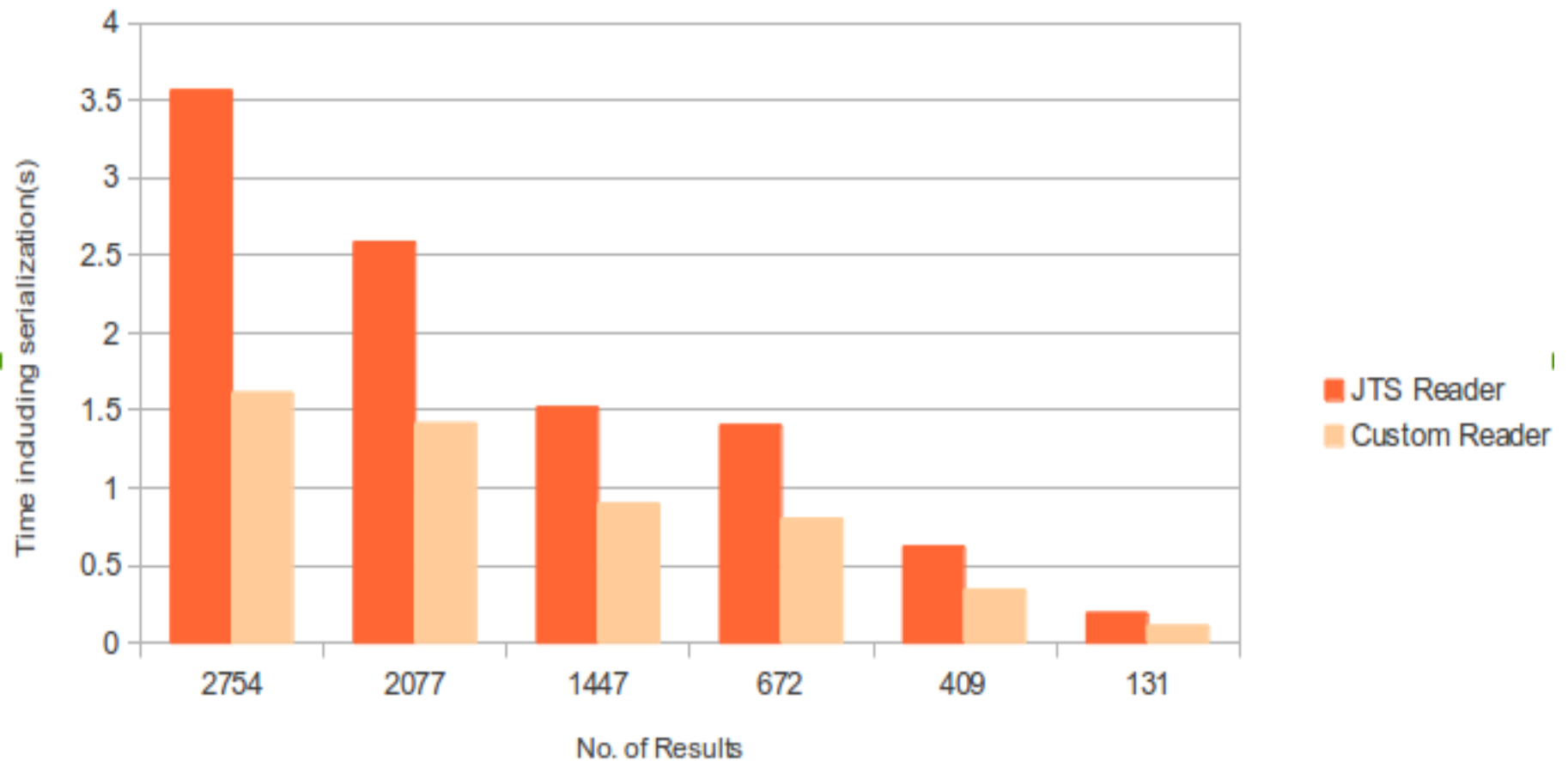
Detailed Timing

let $a:= <gml:Polygon> … </gml:Polygon>

return (

geo-index:**filter("DB", $a)** ⟶ **1451 Objects filtered**
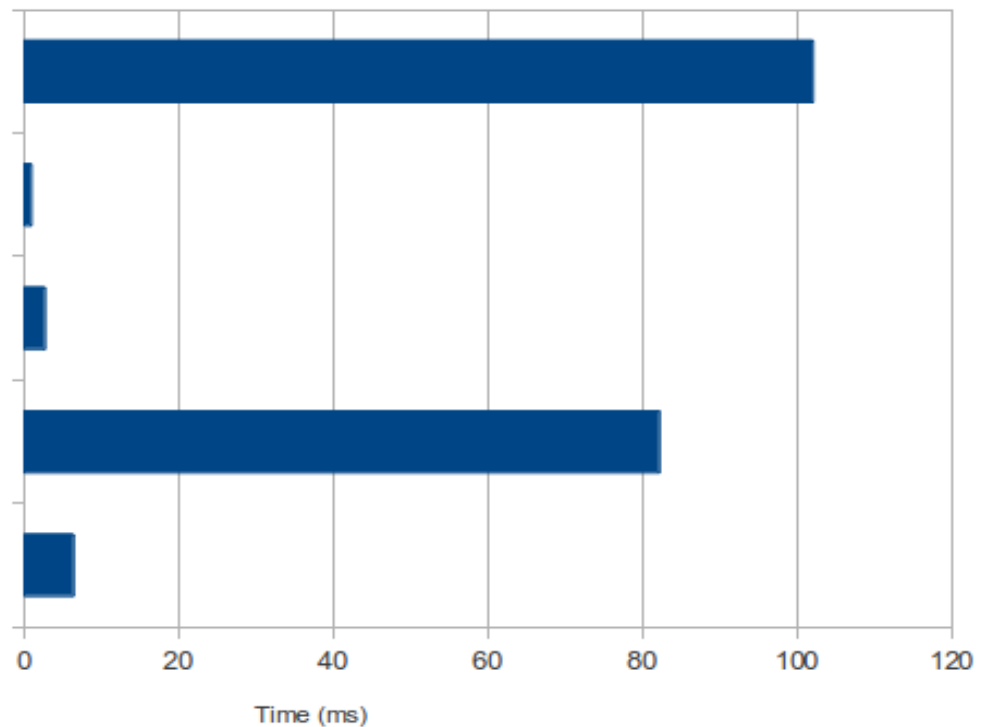
[geo:**intersects(▪ , $a**)]

)

Total time
(1447 results)

Filtering

intersects

Reading all geometries

Reading input geometry

Time (ms)

Detailed Timing (Cont'd)

rank	self	accum	count	method
1	6.95%	6.95%	94	org.basex.util.**Token.split**
2	5.33%	12.28%	72	org.expath.ns.**GmlReader.createPolygon**
3	4.59%	16.86%	62	org.basex.util.**Token.split**
4	4.22%	21.08%	57	org.basex.query.func. **JavaModuleFunc.eval**
5	3.55%	24.63%	48	org.expath.ns.**Geo.geo**
6	3.18%	27.81%	43	org.basex.util.**Token.split**
7	3.03%	30.84%	41	org.basex.util.**Token.split**
8	2.96%	33.80%	40	org.basex.util.**Token.split**
9	2.74%	36.54%	37	org.expath.ns.**Geo.geo**

- ….

Conclusion

- Geo spatial functions in BaseX

- Ability to do Geo-Query

- Indexed geometric operations

- Real-world test cases

Future Works

- Profiling

- Performance tuning

- Geo spatial visualization

- Supporting other common encoding (GML 3, OSM, KML, …)

- Conversion between encodings

- Path finding (traffic segments) in a map

References

1. http://expath.org/spec/geo

2. Leutenegger S., Edgington J.M., Lopez M.A., STR: A Simple and Efficient Algorithm for R-Tree Packing, NASA Contractor Report 201661, ICASE Report No. 97-14, February 1997

3. "Secrets of the JTS Topology Suite - M. Davis". May 2013

4. S. Steiniger and A.J.S. Hunter "The 2012 Free and Open Source GIS Software Map – A Guide to facilitate Research, Development and Adoption"

5. http://lin-ear-th-inking.blogspot.de/2007/06/history-of-jts-and-geos.html

6. http://en.wikipedia.org/wiki/R-tree

7. Y. Manolopoulos; A. Nanopoulos; Y. Theodoridis (2006). R-Trees: Theory and Applications. Springer. ISBN 978-1-85233-977-7. October 2011

8. http://www.digitalpreservation.gov/formats/fdd/fdd000296.shtml

9. http://blog.dc.esri.com/2006/11/20/geodata-format-wars-gml-vs-kml-vs/

Suggestions for the related Master thesis topics are welcomed!

Thank you!