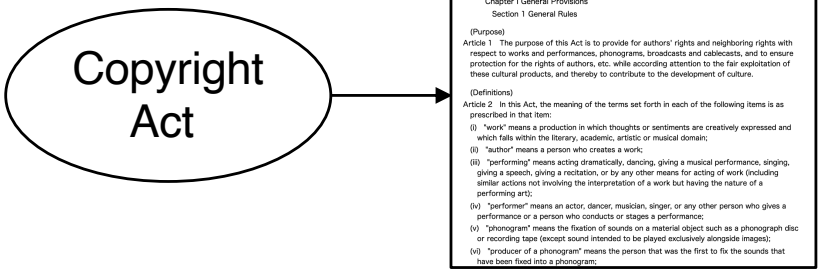
SPARQL with XQuery-based Filtering

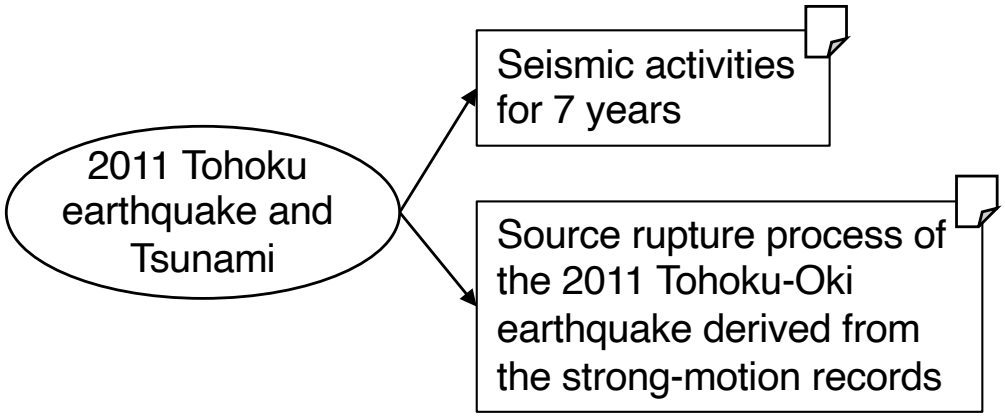
Takahiro Komamizu

Nagoya University

LOD with XML

- LOD (or RDF) is good at representing meta data.
- XML is good at representing document structure.
  - Data-centric: easily translated into RDF
  - Document-centric: redundant in RDF
- Entities in LOD can be linked with related documents.
  - Statute entities with statute texts
  - Event enetities with report documents





Interleaving Processing of LOD and XML

- Possible requirements
  - Entities related with specified documents
    - e.g., “acts which a specified Diet member agreed”
  - Documents related with entities
    - e.g., “damage reports about disaster events in 2015”
- Goal: efficient processing LOD and documents
  1. Search entities with conditions on documents
     - Evaluation of conditions on XML documents is required.
  2. Search parts of documents with entities
     - Extraction of parts of documents and combine with entities is required.

XQueryFilter

metadata in LOD

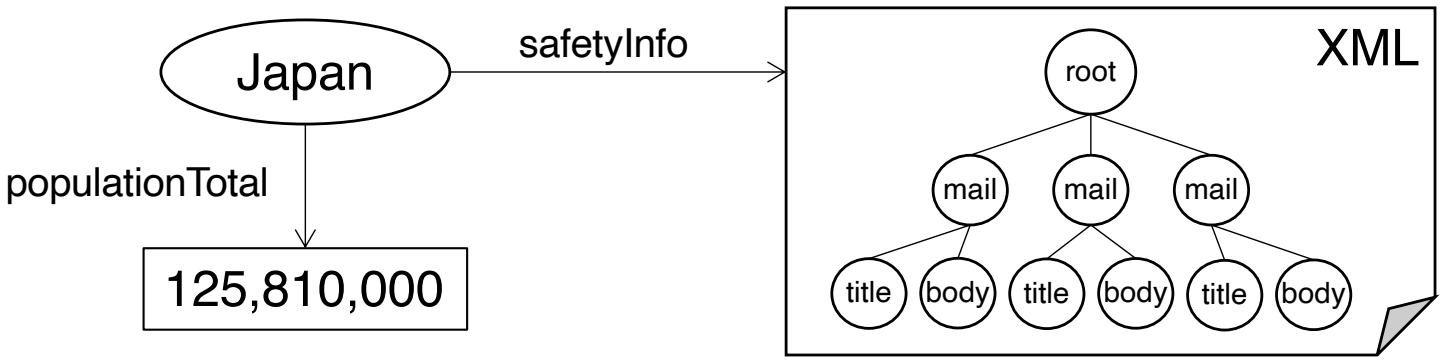
Find countries with more than 10M people and with safety information emails about coronavirus by the national agency after March 2020.

archived emails in XML

Query

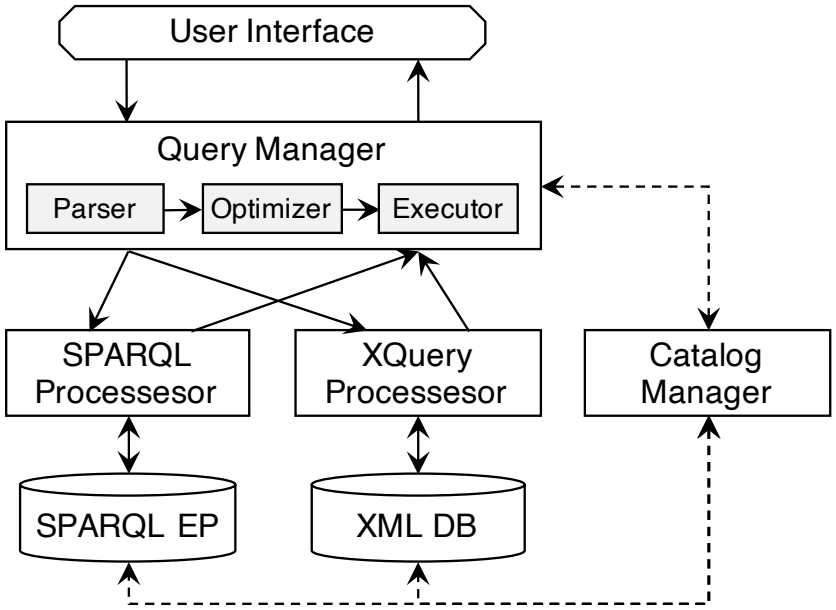
```
SELECT ?s
WHERE{ ?s a :Country; :safetyInfo ?doc; :populationTotal ?pop.
       FILTER ( ?pop > 10,000,000 ) . }
       XQueryFILTER (
          LET $x := doc(?doc)//mail[leaveDate > xs:date('2020-03-01')]
          RETURN contains($x, 'coronavirus')
       ). }
```

Boolean condition



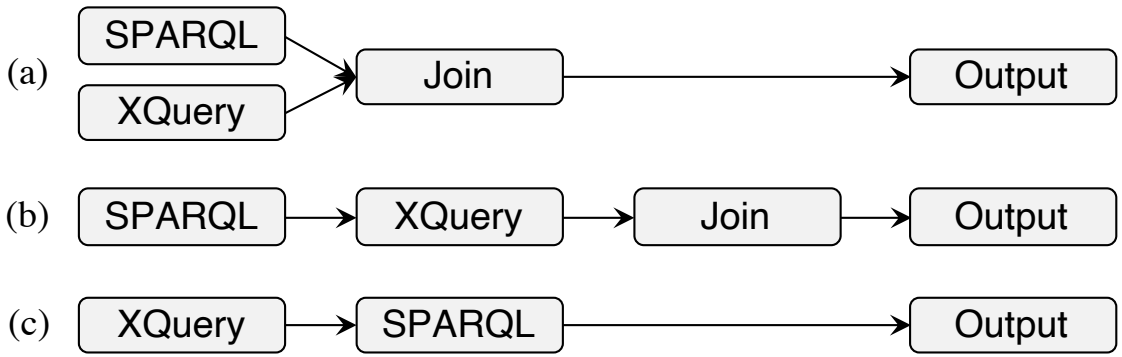
Query Processing

System Architecture

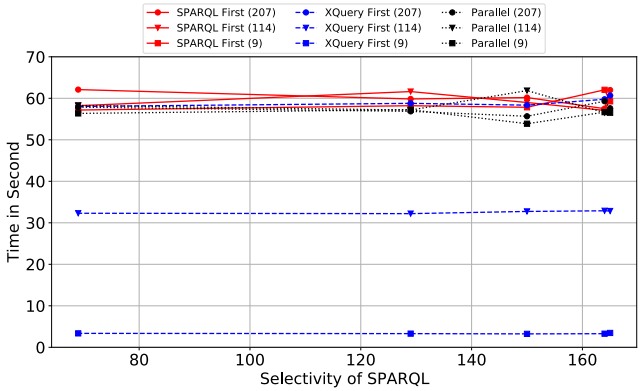


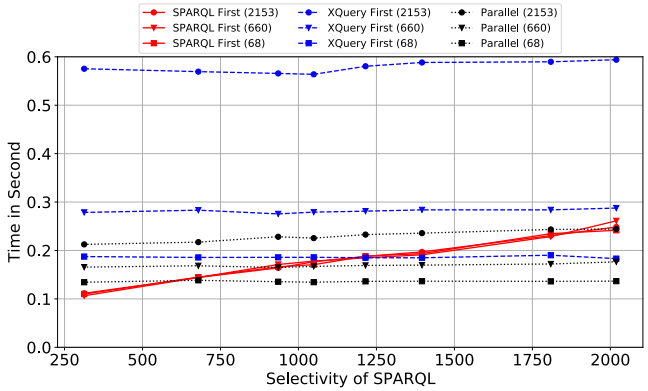
1. **Parser**: Decompose into SPARQL and XQuery.
2. **Optimizer**: Decide the execution order of the individual queries.
   - **Catalog**: performance statistics of underlying databases
3. **Executor**: Execute these queries and composed results.

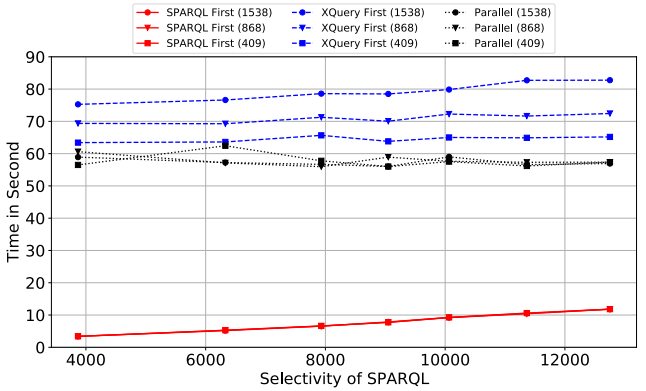
Execution orders



Result: Best orders differ in different DB settings







DB Performance	
SPARQL EP	**low**
XML DB	**high**

XQuery first approach is better choice when the selectivity of XQuery is high.

DB Performance	
SPARQL EP	**comparable**
XML DB	**comparable**

When DBs perform comparably, Parallel and SPARQL first approaches can be a choice depending on the selectivities.

DB Performance	
SPARQL EP	**high**
XML DB	**low**

SPARQL first approach is better choice when the performance of XML DB is low.