

# Knowledge Graph Building

Qianxiang Hao, Haoran Xing

CS4624 Multimedia, Hypertext and Information Access

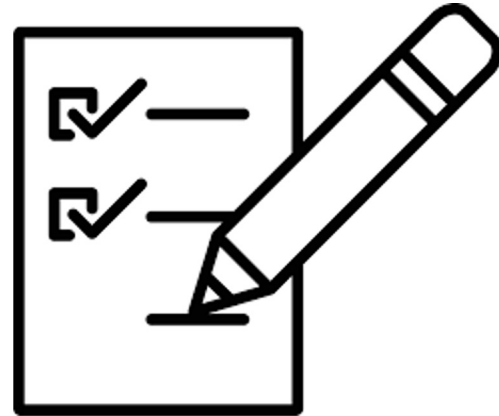
Instructor: Professor Edward A. Fox

Virginia Tech, Blacksburg VA 24061

May 9, 2024

# Outline

- Project Overview
- Problem & Motivation
- Design
- Implementation
- Testing
- Progress Updates
- Acknowledgements
- References



# Project Overview

Our team aims to grow the Virtuoso database by adding enhanced Electronic Theses and Dissertations (ETDs), building on the initial 200 ETD records. This expansion will allow for more thorough testing and analysis of the Knowledge Graph database.



# Problem & Motivation

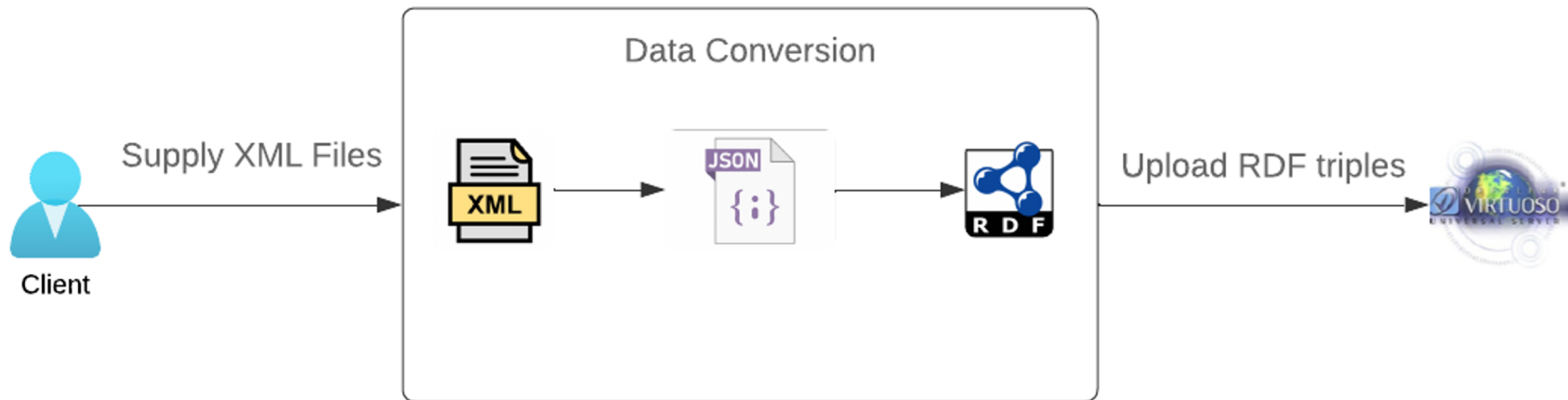
**Problem:** There are a lot of online school papers (ETDs), yet the current Knowledge Graph database houses only 200 ETDs. It is still too small, which limits testing and makes it tough for developers to use it fully.

**Motivation:** We aim to enhance the Knowledge Graph by adding 500,000 ETDs. The expansion of ETDs within the Knowledge Graph amplifies its effectiveness as a tool for academic research, innovation, and knowledge dissemination.



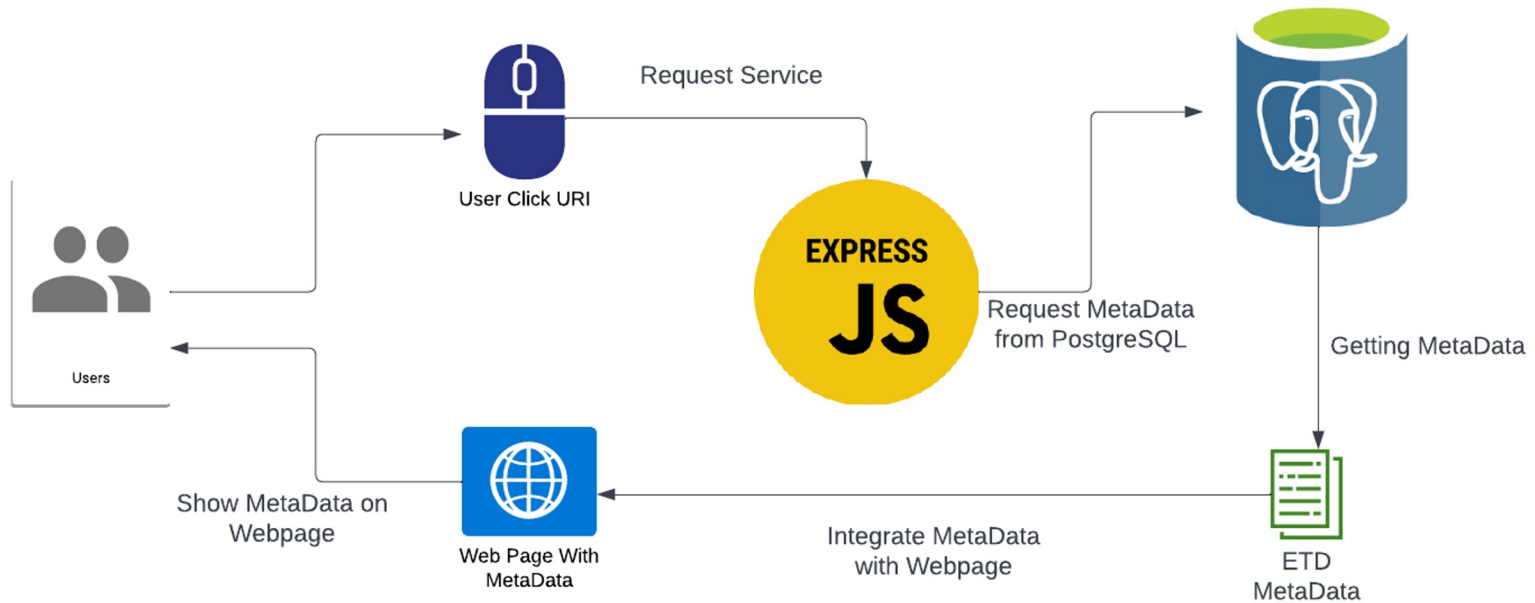
# Design

## Data Conversion & Upload



# Design

## URI Resolution



## URI Resolution Example

**predicate**

<http://etdkb.endeavour.cs.vt.edu/v1/predicate/abstractOf>



ETD Knowledge Graph

Predicate Value: abstractOf  
Description: Indicates the abstract of an ETD

# Implementation

## Data Conversion



# Implementation

## Data Conversion

```
<title bbox="[329.34, 515.9, 1361.62, 665.47]" conf="0.77" pg_no="68" obj_id="138802060648">PART III: EXPERIMENTAL PROTOCOLS AND METHODS</title>
```

XML to JSON

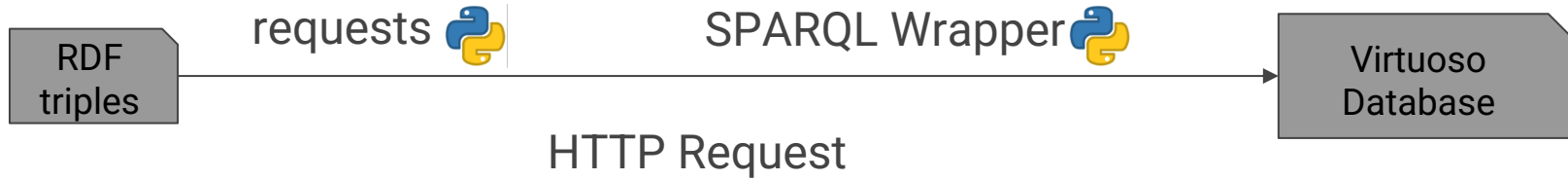
```
"title": {  
  "@bbox": "[329.34, 515.9, 1361.62, 665.47]",  
  "@conf": "0.77",  
  "@pg_no": "68",  
  "@obj_id": "138802060648",  
  "#text": "PART III: EXPERIMENTAL PROTOCOLS AND METHODS"  
},
```

JSON to RDF triple

```
<http://etdkb.endeavour.cs.vt.edu/v1/objects/164506145953> ns1:chapterOf <http://etdkb.endeavour.cs.vt.edu/v1/objects/138802060648> ;
```

# Implementation

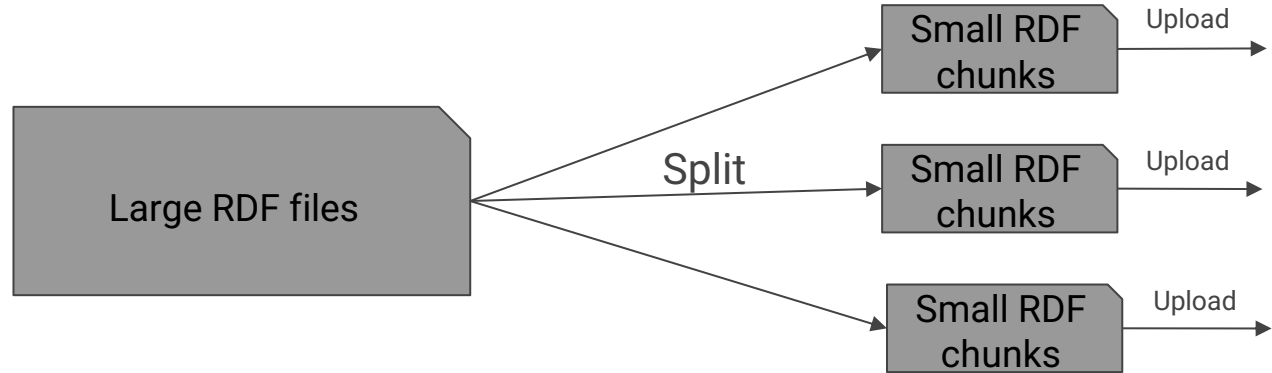
## Data Uploading



# Implementation

## Data Uploading

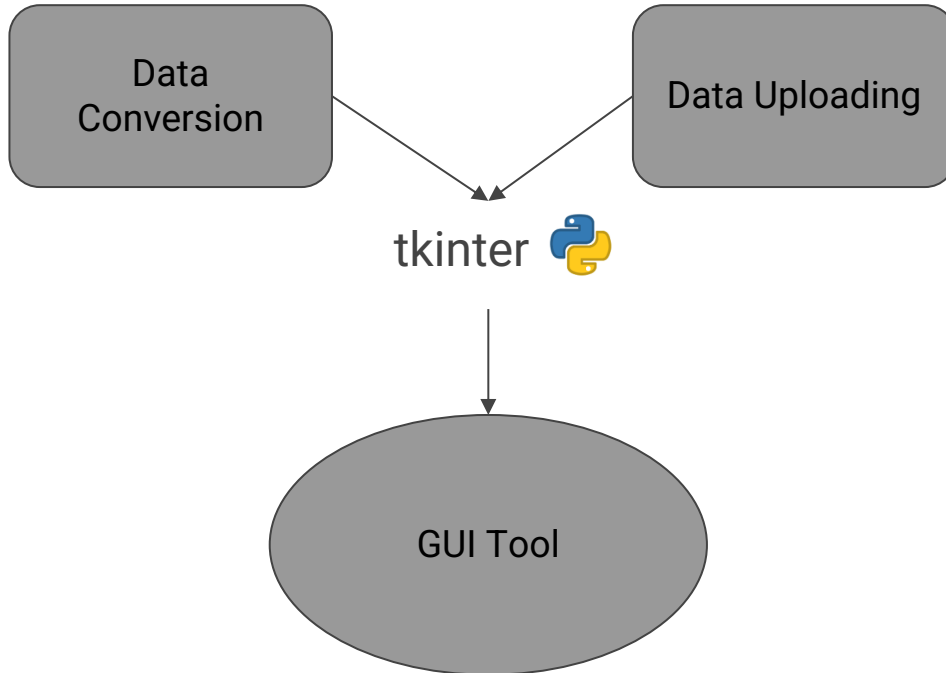
Chunking RDF Data:



- Avoid SPARQL endpoint upload limit
- Upload process more manageable and reliable

# Implementation

## GUI Application Development

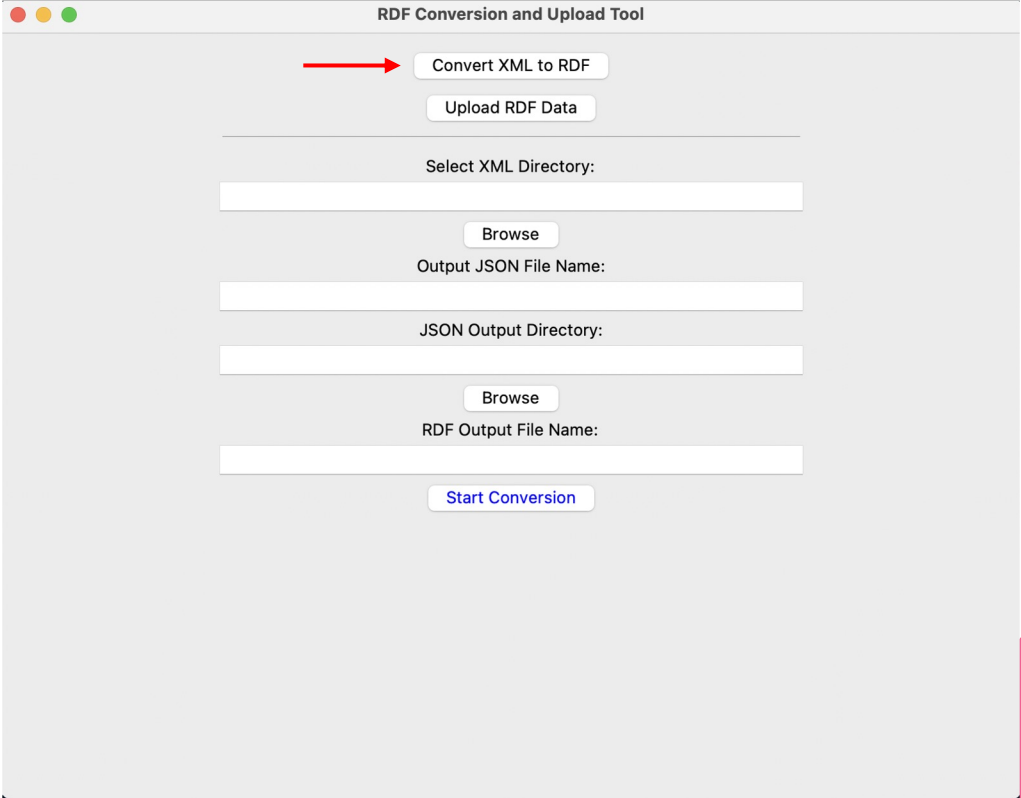


- Standardize Workflow
- Simplifying Process

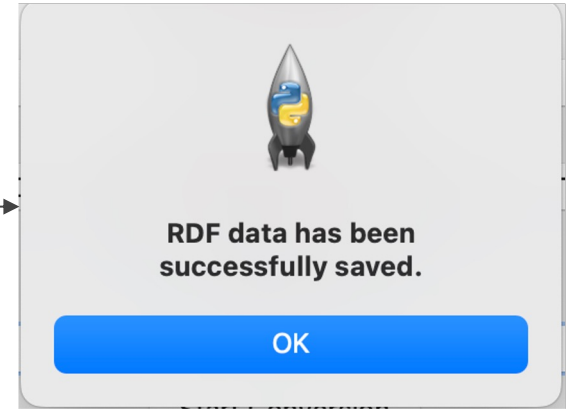
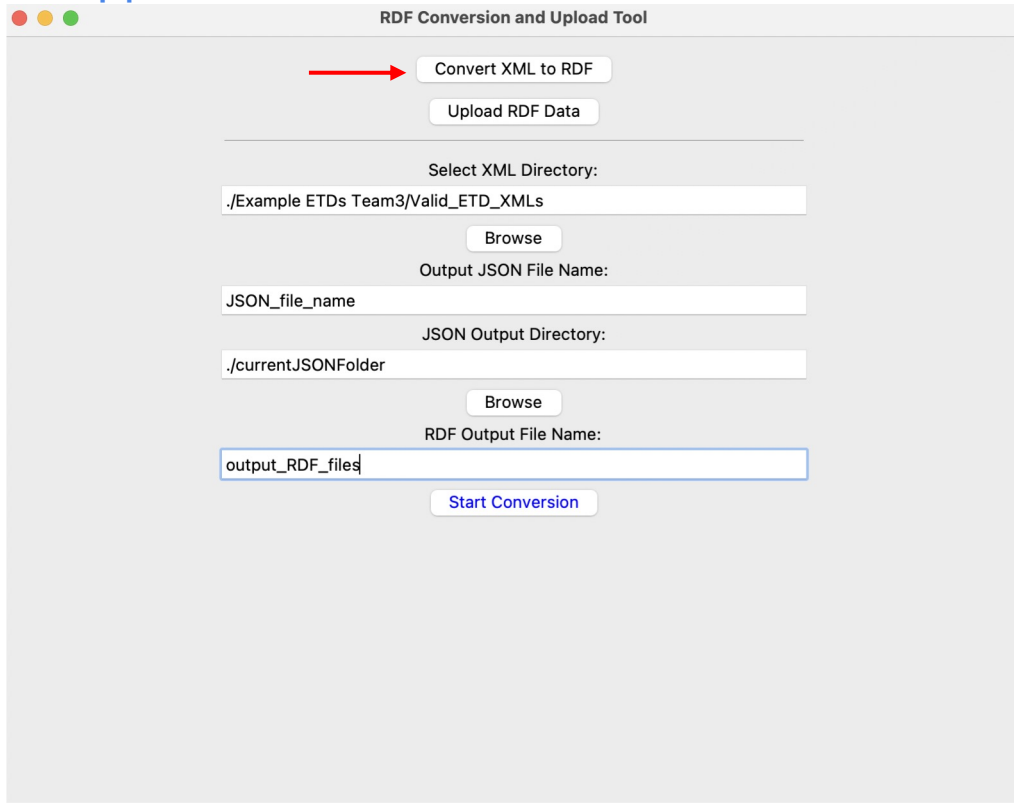
## GUI Application Screenshots



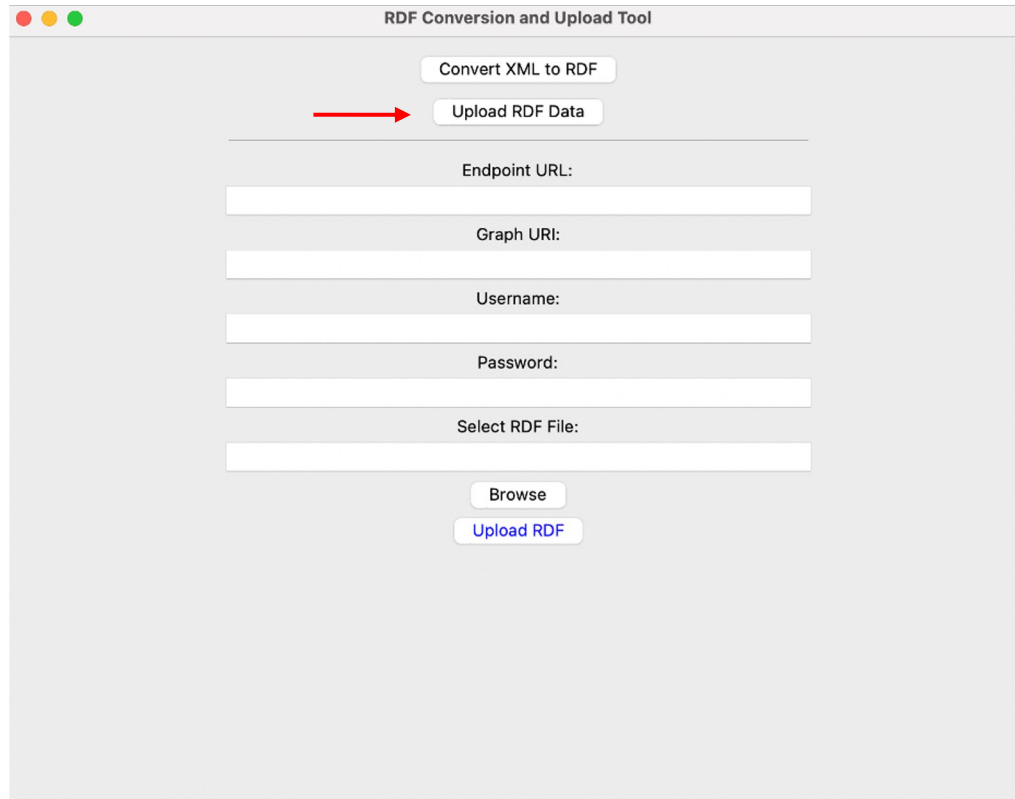
# GUI Application Screenshots



## GUI Application Screenshots



## GUI Application Screenshots



RDF Conversion and Upload Tool

Convert XML to RDF

Upload RDF Data

Endpoint URL:

Graph URI:

Username:

Password:

Select RDF File:

Browse

Upload RDF

## GUI Application Screenshots

RDF Conversion and Upload Tool

Convert XML to RDF

Upload RDF Data

Endpoint URL:

Graph URI:

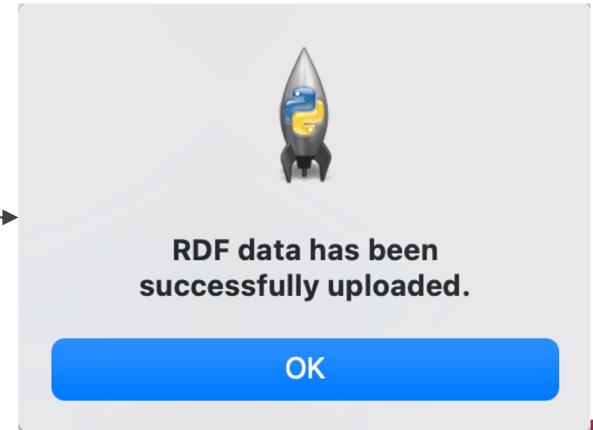

Username:

Password:

Select RDF File:

Browse

Upload RDF



# Testing

- Counting triples
- Query objects in SPARQL interface
- Comparing data manually

# Testing

Query Text

```
SELECT *  
WHERE {  
  ?subject ?predicate ?object  
}
```

Results Format

HTML

Execute Query

Reset

Showing all data in a Graph



# Testing

Query Text

```
SELECT ?subject ?predicate ?object
WHERE{
  ?subject ?predicate ?object
  FILTER regex(str(?predicate), "abstractOf")
  FILTER regex(str(?subject), "6524634")
}
```

Results Format

HTML

Execute Query

Reset

Querying predicate contain “abstractOf”, subject and object with certain Object ID

# Testing

SPARQL | HTML5 table

subject	predicate	object
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/6524634">http://etdtkb.endeavour.cs.vt.edu/v1/objects/6524634</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/6524625">http://etdtkb.endeavour.cs.vt.edu/v1/objects/6524625</a>

Querying results contain “abstractOf”, RDF with certain Object ID

# Progress Update

Uploaded 60 ETDs under graph “<http://localhost:8890/v6/>”

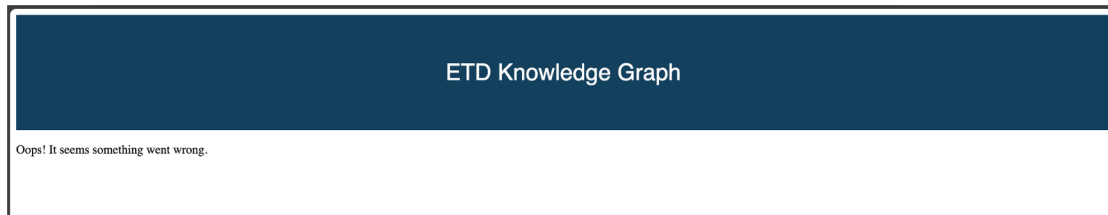
SPARQL | HTML5 table

subject	predicate	object
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/131947637896">http://etdtkb.endeavour.cs.vt.edu/v1/objects/131947637896</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/title">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/title</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/etd/131947637896/meta/tit">http://etdtkb.endeavour.cs.vt.edu/v1/etd/131947637896/meta/tit</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/157651723168">http://etdtkb.endeavour.cs.vt.edu/v1/objects/157651723168</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/title">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/title</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/etd/157651723168/meta/tit">http://etdtkb.endeavour.cs.vt.edu/v1/etd/157651723168/meta/tit</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/143942877705">http://etdtkb.endeavour.cs.vt.edu/v1/objects/143942877705</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/131947637896">http://etdtkb.endeavour.cs.vt.edu/v1/objects/131947637896</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/65117016109">http://etdtkb.endeavour.cs.vt.edu/v1/objects/65117016109</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/32558508054">http://etdtkb.endeavour.cs.vt.edu/v1/objects/32558508054</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/18849662552">http://etdtkb.endeavour.cs.vt.edu/v1/objects/18849662552</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/157651723168">http://etdtkb.endeavour.cs.vt.edu/v1/objects/157651723168</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/80539467301">http://etdtkb.endeavour.cs.vt.edu/v1/objects/80539467301</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/abstractOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/37699325119">http://etdtkb.endeavour.cs.vt.edu/v1/objects/37699325119</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/etd/131947637896/meta/author">http://etdtkb.endeavour.cs.vt.edu/v1/etd/131947637896/meta/author</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/131947637896">http://etdtkb.endeavour.cs.vt.edu/v1/objects/131947637896</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/etd/157651723168/meta/author">http://etdtkb.endeavour.cs.vt.edu/v1/etd/157651723168/meta/author</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/157651723168">http://etdtkb.endeavour.cs.vt.edu/v1/objects/157651723168</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/etd/32558508054/meta/author">http://etdtkb.endeavour.cs.vt.edu/v1/etd/32558508054/meta/author</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/32558508054">http://etdtkb.endeavour.cs.vt.edu/v1/objects/32558508054</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/etd/37699325119/meta/author">http://etdtkb.endeavour.cs.vt.edu/v1/etd/37699325119/meta/author</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/authorOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/37699325119">http://etdtkb.endeavour.cs.vt.edu/v1/objects/37699325119</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/109670763959">http://etdtkb.endeavour.cs.vt.edu/v1/objects/109670763959</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/135374849263">http://etdtkb.endeavour.cs.vt.edu/v1/objects/135374849263</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/116525186698">http://etdtkb.endeavour.cs.vt.edu/v1/objects/116525186698</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/34272113734">http://etdtkb.endeavour.cs.vt.edu/v1/objects/34272113734</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/118238792376">http://etdtkb.endeavour.cs.vt.edu/v1/objects/118238792376</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/159365328855">http://etdtkb.endeavour.cs.vt.edu/v1/objects/159365328855</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/118238792412">http://etdtkb.endeavour.cs.vt.edu/v1/objects/118238792412</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/73685044547">http://etdtkb.endeavour.cs.vt.edu/v1/objects/73685044547</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/73685044547">http://etdtkb.endeavour.cs.vt.edu/v1/objects/73685044547</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/68544227485">http://etdtkb.endeavour.cs.vt.edu/v1/objects/68544227485</a>
<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/138802060611">http://etdtkb.endeavour.cs.vt.edu/v1/objects/138802060611</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf">http://etdtkb.endeavour.cs.vt.edu/v1/predicate/captionOf</a>	<a href="http://etdtkb.endeavour.cs.vt.edu/v1/objects/65117016089">http://etdtkb.endeavour.cs.vt.edu/v1/objects/65117016089</a>

# Progress Update

## Difficulties

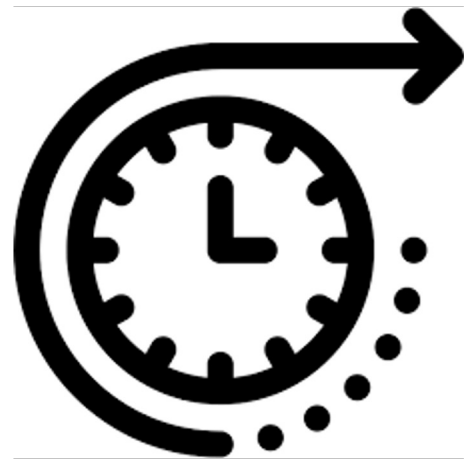
- URI Resolution
  - Object ID
  - PostgreSQL Database
  
- XML file shortage



Error message for URI Webpage

# Future Works

- PostgreSQL access issue
- Object ID consistency
- Expanding 332 or more ETDs for Knowledge Graph



# Acknowledgements

We want to thank **Dr. Edward A. Fox** for his valuable guidance, **Satvik Chekuri** for his technical support and insights, and **Amr Aboelnaga** and **Bipasha Banerjee** for providing the essential XML files for our project.

THANK  
YOU



# References

- [1] Group, PostgreSQL Global Development. PostgreSQL, 28 Apr. 2024, [www.postgresql.org/](http://www.postgresql.org/). Accessed 1 May 2024.
- [2] Fundamentals, Ontotext. “What Is SPARQL?” Ontotext, Ontotext AD, 18 Mar. 2021, [www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/](http://www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/). Accessed 7 May 2024.
- [3] Li, Yike, et al. “Construction of satellite spectrum knowledge graph: Ontology design and knowledge application.” 2023 International Conference on Ubiquitous Communication (Ucom), 7 July 2023, <https://doi.org/10.1109/ucom59132.2023.10257575>.
- [4] Gaugler, Daniel, and Ashish Aggarwal. Information Storage and Retrieval - CS5604 Final Report Team 1 (Knowledge Graphs), in: Virginia Tech, Blacksburg, VA, 4 Dec. 2023, pp. 1–32, <https://hdl.handle.net/10919/118698>.
- [5] Aina, Adeyemi, et al. Team 5 - Infrastructure and DevOps Fall 2023, In CS 5604: Information Storage and Retrieval, Virginia Tech, Blacksburg, VA, 17 Jan. 2024, pp. 1–90, <https://hdl.handle.net/10919/118666>.
- [6] StrongLoop, IBM, and other expressjs.com contributors. “Node.js Web Application Framework.” Translated by StrongLoop/IBM, Express, StrongLoop, IBM, 2017, [expressjs.com/](http://expressjs.com/). Accessed 7 May 2024.

