



Metadata Extraction

Members: Abhishek Nadgeri, Salmaan Tariq, Omar Ejje, Lei Wang, Amit Mudgal, David Abdelsalam

Update Presentation, 06.07.2021

Overview

- Frontend Update
- Extractor Update
- Backend Update
- Infomodel
- Documentation

Frontend Update

[Go to Upload Page](#)

Dataset Quality: ▾

From Date:

To Date:

London traffic dataset




A dataset describing the traffic in london

Car pollution dataset

A dataset comparing the environmental impact of different cars

Beirut economic data

A dataset describing the economy in Lebanon

[Go to Upload Page](#)Dataset Quality: All From Date: To Date: 

sampletitle

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source....

keyword1keyword2keyword3keyword4keyword5keyword6keyword7keyword8

Timeframe: 07-05-2021 to 07-05-2021

EXCELLENT[Show More Details](#)

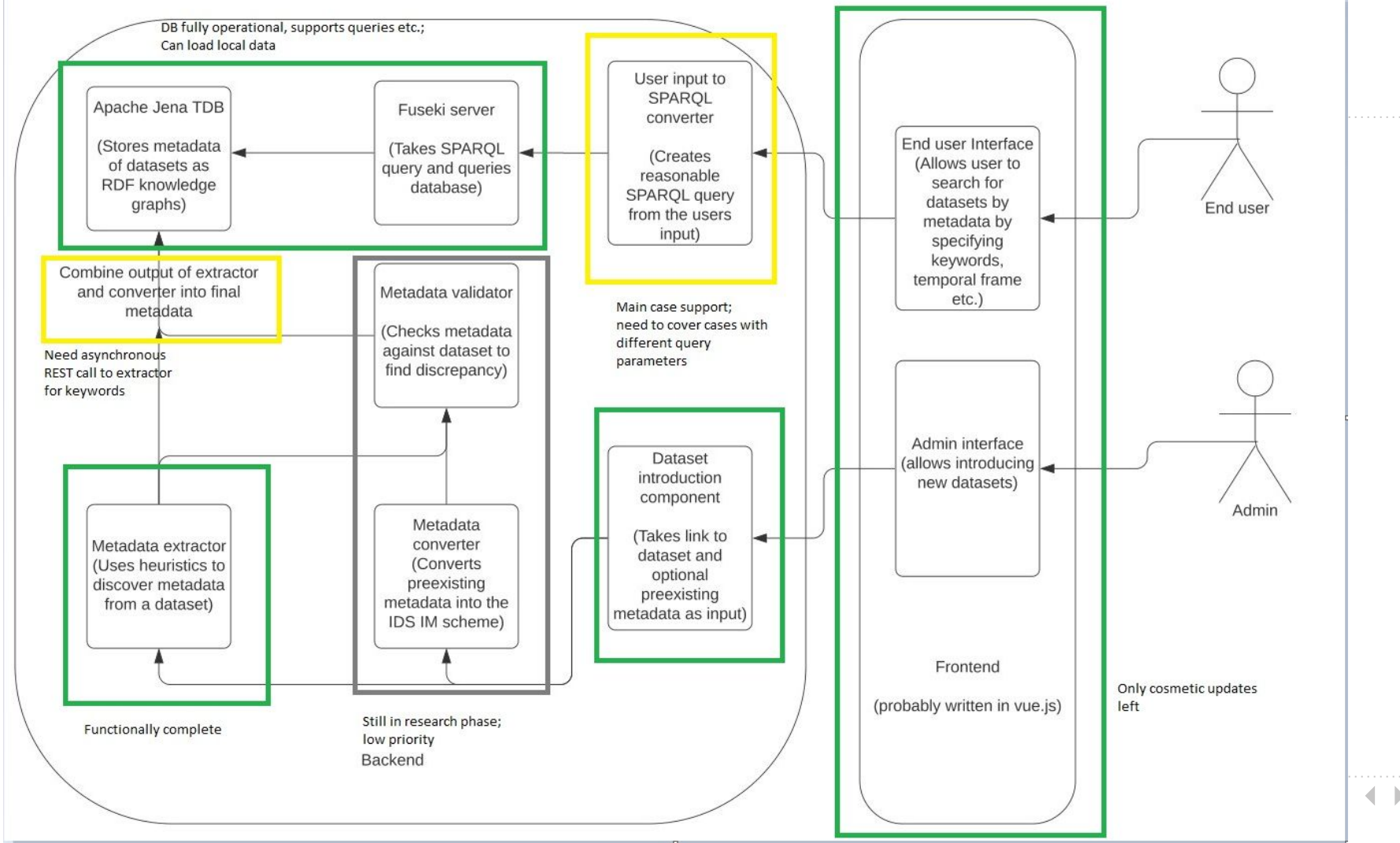
sampletitle

blablalbla

keyword1keyword2

Timeframe: 07-05-2021 to 07-05-2021

GOOD[Show More Details](#)



Extractor Update

Extractor Update

- Extract date from the data and convert its type to xsd:dateTimeStamp format too conform with the format of every other component in Infomodel.
- Extract other metadata:
 - File name
 - File size
 - Creation Date
 - File Type
- Extractor integration done

Backend Update

Backend Update

- Extractor integration done for original concerns
 - Infomodel class mostly done
 - Admin workflow almost complete
-
- Now working on finishing Infomodel class
 - This week: Query generator and adding new metadata concerns
 - Goal: Full demo next week

Infomodel

Infomodel

- Updated Dataquality output (RDF).
- Dataquality and Data in CSV file using different resources
- Dataquality using resource
- Data in csv file using textresource.

Question:

- Representation and Resource library in IDS infomodel don't have DQV output, now still can't output in DQV format. Next step plan feed into Jena model add DQV format, possible?
- The APIs in Representation and Resource, seems just keywords and title can store Dataquality data.

Documentation

Documentation (Salmaan)

- Setting up wiki pages



Thank you

Questions?

Slide theme credit - Isaiah Mulang