Metadata Extraction

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

Progress Report, 15.06

Overview

- Extractor Update
- Backend Update
- Front End Update

Extractor Update

Serving the Extractor

- Monolithic Application that serves extractor components using an API
- Can send POST request with link to CSV file to add datasource to the extractor
- Extractor runs different extractor components to get the metadata concerns
- Metadata concerns are returned on appropriate endpoints

New Extractor Component: Timeframe Extractor

- Uses heuristics to infer the timeframe of the dataset
- E.g.: Can use first and last date in date column (if exists)

Backend Update

Backend Update: General Remarks

- Written in Spring Boot
- Uses embedded Fuseki server with TDB backend for storage

Backend Update: Admin Flow

- Admin can add links to CSV files with POST which are forwarded to the extractor
- Receives extracted concerns and casts them into the IDF IM
- Stores knowledge graph in Fuseki

Backend Update: End user Flow

- User uses UI to send search criteria with GET request
- Search criteria are cast into a SPARQL Query
- Query is forwarded to Fuseki
- Relevant knowledge graphs are sent back to the front end

Backend Update: Immediate Challenge

- Internally, using ORM (JPA) to communicate with Fuseki
- Integration still ongoing
- Roadmap: Want to get it done this week

Thank you

Questions?
Slide theme credit - Isaiah Mulang