Project: Movie Linked Data Publication

1. Group members

Name	Roll No.
Mannan Baidi	19CS10043
Shashwat Shukla	19CS10056
Ajay Ram Meena	19IE3AI18

2. Project Introduction

Multiple platforms provide access to movie-related data, but this information is frequently unstructured and unconnected, making it difficult to search and analyze effectively. In this project, we intend to develop a structured ontology of movie data capable of linking information from the IMDb movie database. The ontology will provide a standard vocabulary and framework for integrating movie-related information, thereby making it simpler to search, discover, and analyze the data.

In order to accomplish this, we have extracted data from IMDb, mapped it to the ontology, and linked it in order to facilitate the effective search, discovery, and analysis of movie-related information. A website has also developed that uses linked data to allow users to search for movies based on various criteria and analyze movie ratings, demonstrating the ontology's effectiveness in facilitating the search and discovery of films.

3. Problem statement and objectives
Problem

There is a vast amount of movie-related data available on multiple platforms; however, these data sources are frequently unstructured and unconnected, making it difficult to search and analyze the information effectively. This project aims to develop an ontology of movie data that can link information from IMDb into a structured format in order to address this issue. The ontology will provide a standard vocabulary and framework for representing and integrating movie-related information, thereby facilitating data search, discovery, and analysis.

Objectives

- Develop an ontology that specifies a standard vocabulary and data representation structure for movie-related information.

- Extract the data from [IMDB movie database](),map it to the ontology.
- Using the ontology, connect the extracted data to facilitate the effective search, discovery, and analysis of movie-related information.
- Construct a website that uses linked data to allow users to search for movies based on a variety of criteria and analyze movie ratings. Demonstrate how the ontology can facilitate the effective search and discovery of films.

4. Methodology
   a. Data Collection
      i. The first part of this project was the data collection. We collected our data from IMDb by downloading all the tsv files from IMDb which were listed on this [link]() . Note : as this data is very large so we took a subset of movies and filtered them out.
      ii. Then we proceeded to store these data in the json format via a python file name script.py. As IMDb treats movies and person different so we made two different json files for the same through our script.
   b. Ontology Generation
      i. We then decided with defining our ontology having various classes and datafields. In total we have 12 classes, 17 object properties and 15 Data properties in our ontology. This ontology was created with the help of **protege**
      ii. As the data is very large we have to make a script to auto generate the rdf representation of each individual according to our specific format.
         1. That is we first inserted some of the individuals with the help of protege and generated the rdf representation file.
         2. Then following the format of the same we just copy pasted the format and generated the rdf file for our whole data using the predefined rdf format.
         3. This rdf file was then converted to turtle file using protege.
   c. Endpoint creation :
      i. Sparql endpoints were generated with the help of Apache Jena Fuseki server.
      ii. We simply started a fuseki server in our local pc and then generated the http endpoint. That endpoint was then passed to our frontend flask application for display of results.
   d. Frontend :
      i. Our frontend basically has two functionalities
         1. It searches the movie on the basis of the given keyword depending upon Movie Title, if you search movie title then it will give all the information related to movie. If you search person then the same query is ran for that also.

2. You can also execute the sparql query from the query tab directly.

5. Related Links
   a. <u>Github Repository Link</u>
      Video presentation link is available in the github repository readme.

6. References

Bollacker, K.D., Evans, C., Paritosh, P.K., Sturge, T., & Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. SIGMOD Conference.

Avancha, S., Kallurkar, S., & Kamdar, T. Design of Ontology for The Internet Movie Database (IMDb).

Kim, H.L., & Yang, J. (2017). Constructing a Knowledge Base for Entertainment by Interlinking Multiple Data Sources. EAI Endorsed Trans. Ind. Networks Intell. Syst., 4, e1.

Raimond, Yves et al. "The Music Ontology." International Society for Music Information Retrieval Conference (2007).