

Prague University of Economics and Business

Faculty of Informatics and Statistics



SEMANTIC DATA VOCABULARY FOR ENTERTAINMENT AND MEDIA

Semestral project

Course: 4IZ441 - Graph data and knowledge

Lecturer: prof. Ing. Vojtěch Svátek, Dr.

Author: Hoang Anh Tran Nguyenová

Contents

Introduction1

1 Vocabulary Overview and Definitions 2

 1.1 Entities and Their Definitions 2

 1.2 Properties and Their Definitions..... 3

2 Implementation 4

 2.1 RDFS and Related Ontologies 4

 2.2 The Prefixed Reused Ontologies 4

 2.3 Media and Entertainment Vocabulary (media-entertainment-voc) 5

3 Sample Data in Turtle Format..... 9

4 SPARQL Queries.....15

Conclusion19

Sources 20

Introduction

The entertainment and media industry generates extensive data, including information about movies, TV shows, streaming platforms, and awards. However, much of this data is unstructured, leading to inefficiencies in integration and retrieval. This semestral project focuses on designing a semantic vocabulary tailored to this domain using RDF, RDFS, OWL, and schema.org standards. By providing structured descriptions of entities and their relationships, this vocabulary enhances data organization, interoperability, and accessibility, enabling advanced applications such as recommendation systems and semantic queries.

1 Vocabulary Overview and Definitions

The vocabulary overview offers a comprehensive breakdown of the core entities and properties that are critical for representing the entertainment and media domain. Each entity represents a key component of the industry, such as movies, TV shows, or streaming platforms, while properties define the relationships and attributes associated with these entities.

1.1 Entities and Their Definitions

1. MediaContent

- Definition: Represents a piece of media, such as a movie or TV show.
- Superclass: `schema:CreativeWork`.
- Attributes: `genre`, `releaseDate`, `rating`, `duration`.

2. Person

- Definition: Represents an individual involved in media content, such as an actor, director, or producer.
- Superclass: `schema:Person`.
- Relationships: Can `starredIn` or `directedBy`, or `producedBy` MediaContent.

3. StreamingPlatform

- Definition: Represents a digital platform where media content is hosted, such as Netflix, HBO, or Disney+.
- Superclass: `schema:Organization`.
- Relationships: MediaContent is `availableOn` StreamingPlatform.

4. Award

- Definition: Represents awards or accolades given to individuals or media content, such as the Oscars or Emmys.
- Attributes: Award name, date awarded.

5. Franchise

- Definition: Represents a collection of related media content, such as sequels, spin-offs, or remakes.
- Superclass: `schema:CreativeWork`.
- Relationships: Can `includesContent` MediaContent.

6. Review

- Definition: Represents feedback or critiques associated with media content.
- Superclass: `schema:Review`.
- Attributes: Reviewer name, rating, comment.

7. ProductionCompany

- Definition: Represents a company involved in the production of media content.
- Superclass: `schema:Organization`.
- Relationships: MediaContent is `producedBy` ProductionCompany.

1.2 Properties and Their Definitions

1. **starredIn**

- Definition: Links a **Person** to the **MediaContent** they have acted in.
- Domain: **Person**.
- Range: **MediaContent**.

2. **directedBy**

- Definition: Links a **MediaContent** to the **Person** who directed it.
- Domain: **MediaContent**.
- Range: **Person**.

3. **producedBy**

- Definition: Links a **MediaContent** to a **ProductionCompany** that produced it.
- Domain: **MediaContent**.
- Range: **ProductionCompany**.

4. **availableOn**

- Definition: Links **MediaContent** to a **StreamingPlatform** where it is hosted.
- Domain: **MediaContent**.
- Range: **StreamingPlatform**.

5. **wonAward**

- Definition: Links a **Person** or **MediaContent** to an **Award**.
- Domain: **Person** or **MediaContent**.
- Range: **Award**.

6. **releaseDate**

- Definition: Specifies the date a **MediaContent** was released.
- Domain: **MediaContent**.
- Range: **schema:Date**.

7. **genre**

- Definition: Specifies the genre of a **MediaContent**, such as **Comedy** or **Drama**.
- Domain: **MediaContent**.
- Range: **schema:Text**.

8. **includesContent**

- Definition: Links a **Franchise** to individual **MediaContent** items within it.
- Domain: **Franchise**.
- Range: **MediaContent**.

9. **reviewedBy**

- Definition: Links a **Review** to the **MediaContent** it critiques.
- Domain: **Review**.
- Range: **MediaContent**.

10. **duration**

- Definition: Specifies the length of a **MediaContent** in minutes.
- Domain: **MediaContent**.
- Range: **xsd:integer**.

2 Implementation

2.1 RDFS and Related Ontologies

RDFS (RDF Schema)

RDFS provides a basic set of constructs to define classes and properties in RDF. It is used in this project to define the hierarchical relationships among entities like `MediaContent`, `Person`, and `StreamingPlatform`.

- Example: `rdfs:Class` is used to define `MediaContent` as a class.
- Example: `rdfs:subClassOf` is used to state that `MediaContent` is a subclass of `schema:CreativeWork`.

XML Schema (XSD)

XML Schema is used for data types in RDF, such as dates, strings, and integers. It provides a foundation for expressing literal values in RDF.

- Example: `xsd:date` is used to represent the release date of `MediaContent`.

Schema.org

Schema.org offers a broad vocabulary for structured data on the web, making it a crucial component for this project. It provides predefined entities such as `CreativeWork` and `Person`, which are extended in the vocabulary to include domain-specific concepts. By incorporating schema.org, the vocabulary aligns with global standards and becomes compatible with APIs and data sources like TMDb and IMDb.

- Example: `schema:CreativeWork` serves as the superclass for `MediaContent`.
- Example: `schema:Organization` is reused for `StreamingPlatform`.

2.2 The Prefixed Reused Ontologies

```
@prefix schema: <http://schema.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
```

2.3 Media and Entertainment Vocabulary (media-entertainment-voc)

Below is the complete RDF vocabulary along with its definitions, including the prefix.

```
@prefix em: <http://example.org/media-entertainment-voc#> .
```

Classes

MediaContent Class in Turtle

```
# MediaContent Class
em:MediaContent a owl:Class ;
  rdfs:label "Media Content" ;
  rdfs:comment "Represents movies or TV shows." ;
  rdfs:subClassOf schema:CreativeWork .
```

Person Class in Turtle

```
# Person Class
em:Person a owl:Class ;
  rdfs:label "Person" ;
  rdfs:comment "Represents an actor, director, or producer." ;
  rdfs:subClassOf schema:Person .
```

StreamingPlatform Class in Turtle

```
# StreamingPlatform Class
em:StreamingPlatform a owl:Class ;
  rdfs:label "Streaming Platform" ;
  rdfs:comment "Represents platforms hosting media content." ;
  rdfs:subClassOf schema:Organization .
```

Award Class in Turtle

```
# Award Class
em:Award a owl:Class ;
  rdfs:label "Award" ;
  rdfs:comment "Represents awards given to persons or media content." .
```

Franchise Class in Turtle

```
# Franchise Class
em:Franchise a owl:Class ;
  rdfs:label "Franchise" ;
  rdfs:comment "Represents a collection of related media content." ;
  rdfs:subClassOf schema:CreativeWork .
```

Review Class in Turtle

```
# Review Class
em:Review a owl:Class ;
  rdfs:label "Review" ;
  rdfs:comment "Represents feedback or critiques associated with media content." ;
  rdfs:subClassOf schema:Review .
```

ProductionCompany Class in Turtle

```
# ProductionCompany Class
em:ProductionCompany a owl:Class ;
  rdfs:label "Production Company" ;
  rdfs:comment "Represents a company involved in producing media content." ;
  rdfs:subClassOf schema:Organization .
```

Properties

Person Class Properties in Turtle

```
# Person Class Properties and links to MediaContent
em:starredIn a owl:ObjectProperty ;
  rdfs:label "Starred In" ;
  rdfs:domain em:Person ;
  rdfs:range em:MediaContent .

# Person Class and MediaContent Class Properties and links to Award
em:wonAward a owl:ObjectProperty ;
  rdfs:label "Won Award" ;
  rdfs:domain [ a owl:Class ; owl:unionOf (em:Person em:MediaContent) ] ;
  rdfs:range em:Award .
```


MediaContent Class Properties in Turtle

```
# Person Class and MediaContent Class Properties and links to Award
em:wonAward a owl:ObjectProperty ;
  rdfs:label "Won Award" ;
  rdfs:domain [ a owl:Class ; owl:unionOf (em:Person em:MediaContent) ] ;
  rdfs:range em:Award .

# MediaContent Class Properties and links to Person
em:directedBy a owl:ObjectProperty ;
  rdfs:label "Directed By" ;
  rdfs:domain em:MediaContent ;
  rdfs:range em:Person ;
  owl:inverseOf em:hasDirector .

# MediaContent Class Properties and links to Person as an inverse of directedBy
em:hasDirector a owl:ObjectProperty ;
  rdfs:label "Has Director" ;
  rdfs:domain em:MediaContent ;
  rdfs:range em:Person .

# MediaContent Class Properties and links to ProductionCompany
em:producedBy a owl:ObjectProperty ;
  rdfs:label "Produced By" ;
  rdfs:domain em:MediaContent ;
  rdfs:range em:ProductionCompany .

# MediaContent Class Properties and links to StreamingPlatform
em:availableOn a owl:ObjectProperty ;
  rdfs:label "Available On" ;
  rdfs:domain em:MediaContent ;
  rdfs:range em:StreamingPlatform .

# MediaContent Class Properties and specifies the release date
em:releaseDate a owl:DatatypeProperty ;
  rdfs:label "Release Date" ;
  rdfs:domain em:MediaContent ;
  rdfs:range schema:Date .

# MediaContent Class Properties and specifies the genre
em:genre a owl:DatatypeProperty ;
  rdfs:label "Genre" ;
  rdfs:domain em:MediaContent ;
  rdfs:range schema:Text .

# MediaContent Class Properties and specifies the duration in minutes
em:duration a owl:DatatypeProperty ;
  rdfs:label "Duration" ;
  rdfs:domain em:MediaContent ;
  rdfs:range xsd:integer .
```

Franchise Class Properties in Turtle

```
# Franchise Class Properties and links to MediaContent
em:includesContent a owl:ObjectProperty ;
  rdfs:label "Includes Content" ;
  rdfs:domain em:Franchise ;
  rdfs:range em:MediaContent .
```

Review Class Properties in Turtle

```
# Review Class Properties and links to MediaContent
em:reviewedBy a owl:ObjectProperty ;
  rdfs:label "Reviewed By" ;
  rdfs:domain em:Review ;
  rdfs:range em:MediaContent .
```

3 Sample Data in Turtle Format

This section provides sample data for the entertainment and media vocabulary in Turtle format. It shows how the main classes (MediaContent, Person, StreamingPlatform, ProductionCompany, Award, Franchise, and Review) are connected. Each class is represented with its key properties and links to other entities.

Summary of Classes in the Sample Data:

- **MediaContent:** Represents movies and TV shows such as Inception, Avatar, and Frozen.
- **Person:** Represents individuals involved in media production, such as Leonardo DiCaprio, Christopher Nolan, and James Cameron.
- **StreamingPlatform:** Represents platforms hosting media content, including Netflix, Hulu, and Disney+.
- **ProductionCompany:** Represents companies like Lightstorm Entertainment and Syncopy that produce media content.
- **Award:** Represents accolades given to individuals or media content, such as the Oscar 2010 for Best Actor awarded to Leonardo DiCaprio.
- **Franchise:** Represents collections of related media, such as the Avatar Franchise, which includes Avatar and Avatar 2.
- **Review:** Represents reviews and feedback for media content, such as critic and user reviews.

Here's how the vocabulary works with real-world examples. Movies are represented as MediaContent, people like directors as Person, and platforms like Netflix and Disney+ as StreamingPlatform. The sample data also links movies to awards, production companies, and franchises, showing how everything in the entertainment industry is interconnected.

MediaContent Class

```
### Media Content
em:Inception a em:MediaContent ;
  schema:name "Inception" ;
  schema:datePublished "2010-07-16"^^xsd:date ;
  schema:genre "Science Fiction" ;
  em:directedBy em:Nolan ;
  em:availableOn em:Netflix .
```

```

em:Avatar a em:MediaContent ;
  schema:name "Avatar" ;
  schema:datePublished "2009-12-18"^^xsd:date ;
  schema:genre "Adventure" ;
  em:directedBy em:Cameron ;
  em:availableOn em:DisneyPlus .

em:Frozen a em:MediaContent ;
  schema:name "Frozen" ;
  schema:datePublished "2013-11-27"^^xsd:date ;
  schema:genre "Animation" ;
  em:directedBy em:Buck ;
  em:availableOn em:DisneyPlus .

em:JurassicPark a em:MediaContent ;
  schema:name "Jurassic Park" ;
  schema:datePublished "1993-06-11"^^xsd:date ;
  schema:genre "Adventure" ;
  em:directedBy em:Spielberg ;
  em:availableOn em:PrimeVideo .

em:BlackPanther a em:MediaContent ;
  schema:name "Black Panther" ;
  schema:datePublished "2018-02-16"^^xsd:date ;
  schema:genre "Action" ;
  em:directedBy em:Coogler ;
  em:availableOn em:DisneyPlus .

em:Moana a em:MediaContent ;
  schema:name "Moana" ;
  schema:datePublished "2016-11-23"^^xsd:date ;
  schema:genre "Animation" ;
  em:directedBy em:Clements ;
  em:availableOn em:DisneyPlus .

em:StarWars a em:MediaContent ;
  schema:name "Star Wars: A New Hope" ;
  schema:datePublished "1977-05-25"^^xsd:date ;
  schema:genre "Science Fiction" ;
  em:directedBy em:Lucas ;
  em:availableOn em:DisneyPlus .

em:Coco a em:MediaContent ;
  schema:name "Coco" ;
  schema:datePublished "2017-11-22"^^xsd:date ;
  schema:genre "Animation" ;
  em:directedBy em:Unkrich ;
  em:availableOn em:DisneyPlus .

```

Person Class

```
### Persons
em:Nolan a em:Person ;
    schema:name "Christopher Nolan" ;
    em:starredIn em:Inception, em:Interstellar, em:Tenet .

em:DiCaprio a em:Person ;
    schema:name "Leonardo DiCaprio" ;
    em:starredIn em:Inception .

em:Cameron a em:Person ;
    schema:name "James Cameron" ;
    em:directedBy em:Avatar .

em:Buck a em:Person ;
    schema:name "Chris Buck" ;
    em:directedBy em:Frozen .

em:Spielberg a em:Person ;
    schema:name "Steven Spielberg" ;
    em:directedBy em:JurassicPark .

em:Coogler a em:Person ;
    schema:name "Ryan Coogler" ;
    em:directedBy em:BlackPanther .

em:Lucas a em:Person ;
    schema:name "George Lucas" ;
    em:directedBy em:StarWars .

em:Clements a em:Person ;
    schema:name "Ron Clements" ;
    em:directedBy em:Moana .

em:Unkrich a em:Person ;
    schema:name "Lee Unkrich" ;
    em:directedBy em:Coco .
```

StreamingPlatform Class

```
### Streaming Platforms
em:Netflix a em:StreamingPlatform ;
    schema:name "Netflix" .

em:DisneyPlus a em:StreamingPlatform ;
    schema:name "Disney+" .

em:PrimeVideo a em:StreamingPlatform ;
    schema:name "Prime Video" .
```

Award Class

```
### Awards
em:Oscar2020 a em:Award ;
    schema:name "Oscar 2020" ;
    em:wonAward em:Inception .

em:GoldenGlobe2021 a em:Award ;
    schema:name "Golden Globe 2021" ;
    em:wonAward em:Avatar .

em:BAFTA2022 a em:Award ;
    schema:name "BAFTA 2022" ;
    em:wonAward em:BlackPanther .

em:AcademyAward1994 a em:Award ;
    schema:name "Academy Award 1994" ;
    em:wonAward em:JurassicPark .

em:GoldenGlobe2017 a em:Award ;
    schema:name "Golden Globe 2017" ;
    em:wonAward em:Moana .

em:Oscar2010 a em:Award ;
    schema:name "Oscar 2010" ;
    em:wonAward em:DiCaprio .
```

Franchise Class

```
### Franchises
em:AvatarFranchise a em:Franchise ;
    schema:name "Avatar Franchise" ;
    em:includesContent em:Avatar, em:Avatar2 .

em:JurassicFranchise a em:Franchise ;
    schema:name "Jurassic Franchise" ;
    em:includesContent em:JurassicPark, em:JurassicWorld .

em:StarWarsFranchise a em:Franchise ;
    schema:name "Star Wars Franchise" ;
    em:includesContent em:StarWars .

em:FrozenFranchise a em:Franchise ;
    schema:name "Frozen Franchise" ;
    em:includesContent em:Frozen .
```

ProductionCompany Class

```
### Production Companies
em:Lightstorm a em:ProductionCompany ;
    schema:name "Lightstorm Entertainment" ;
    em:producedBy em:Avatar, em:Avatar2 .

em:Amblin a em:ProductionCompany ;
    schema:name "Amblin Entertainment" ;
    em:producedBy em:JurassicPark .

em:MarvelStudios a em:ProductionCompany ;
    schema:name "Marvel Studios" ;
    em:producedBy em:BlackPanther .

em:Pixar a em:ProductionCompany ;
    schema:name "Pixar Animation Studios" ;
    em:producedBy em:Coco .

em:Lucasfilm a em:ProductionCompany ;
    schema:name "Lucasfilm Ltd." ;
    em:producedBy em:StarWars .
```

Review Class

```
### Reviews
em:Review1 a em:Review ;
    schema:name "Top Critic Review" ;
    em:reviewedBy em:Avatar ;
    schema:reviewBody "An amazing visual experience!" .

em:Review2 a em:Review ;
    schema:name "User Review" ;
    em:reviewedBy em:Inception ;
    schema:reviewBody "A mind-bending masterpiece." .

em:Review3 a em:Review ;
    schema:name "Critic Review" ;
    em:reviewedBy em:JurassicPark ;
    schema:reviewBody "A thrilling adventure with groundbreaking effects." .

em:Review4 a em:Review ;
    schema:name "Audience Review" ;
    em:reviewedBy em:Moana ;
    schema:reviewBody "Heartwarming and beautifully animated." .

em:Review5 a em:Review ;
    schema:name "Expert Review" ;
    em:reviewedBy em:Coco ;
    schema:reviewBody "An emotional journey with stunning visuals." .
```


4 SPARQL Queries

This section includes SPARQL queries that show how the entertainment and media vocabulary can be used in practice. These queries demonstrate how to retrieve information from the RDF dataset, such as details about movies, awards, production companies, and streaming platforms.

1. Movies and their directors

Lists all movies and their directors, showing how `MediaContent` is linked to `Person` through `em:directedBy`.

```
SELECT ?movieName ?directorName
WHERE {
  ?movie a em:MediaContent ;
         schema:name ?movieName ;
         em:directedBy ?director .
  ?director schema:name ?directorName .
}
```

2. Streaming platforms hosting movies

Finds all movies and the platforms where they are available, using `em:availableOn` to connect `MediaContent` and `StreamingPlatform`.

```
SELECT ?movieName ?platformName
WHERE {
  ?movie a em:MediaContent ;
         schema:name ?movieName ;
         em:availableOn ?platform .
  ?platform schema:name ?platformName .
}
```

3. Movies by genre

Retrieves movies of a specific genre, like "Science Fiction," using `schema:genre`.

```
SELECT ?movieName
WHERE {
  ?movie a em:MediaContent ;
         schema:name ?movieName ;
         schema:genre "Science Fiction" .
}
```

4. Awards won by movies

Lists awards won by movies, connecting `Award` and `MediaContent` through `em:wonAward`.

```
SELECT ?movieName ?awardName
WHERE {
  ?award a em:Award ;
         schema:name ?awardName ;
         em:wonAward ?movie .
  ?movie schema:name ?movieName .
}
```

5. Reviews for a movie

Retrieves reviews for a specific movie, such as Avatar, showing the link between `Review` and `MediaContent` using `em:reviewedBy`.

```
SELECT ?reviewName ?reviewBody
WHERE {
  ?review a em:Review ;
         schema:name ?reviewName ;
         schema:reviewBody ?reviewBody ;
         em:reviewedBy em:Avatar .
}
```

6. Movies on a platform

Lists all movies available on a specific platform, like Disney+, using `em:availableOn`.

```
SELECT ?movieName
WHERE {
  ?movie a em:MediaContent ;
         schema:name ?movieName ;
         em:availableOn em:DisneyPlus .
}
```

7. Movie directors

Retrieves all directors of movies by linking `Person` and `MediaContent` with `em:directedBy`.

```
SELECT ?directorName
WHERE {
  ?movie a em:MediaContent ;
         em:directedBy ?director .
  ?director schema:name ?directorName .
}
```

8. Franchises and their content

Lists franchises and the movies or TV shows they include, using `em:includesContent`.

```
SELECT ?franchiseName ?contentName
WHERE {
  ?franchise a em:Franchise ;
             schema:name ?franchiseName ;
             em:includesContent ?content .
  ?content schema:name ?contentName .
}
```

9. Awards for a person

Shows awards won by a specific person, like Leonardo DiCaprio, using `em:wonAward`.

```
SELECT ?awardName
WHERE {
  ?award a em:Award ;
         schema:name ?awardName ;
         em:wonAward ?person .
  ?person schema:name "Leonardo DiCaprio" .
}
```

10. Movies released after a date

Finds movies released after a specific date by filtering `schema:datePublished`.

```
SELECT ?movieName ?releaseDate
WHERE {
  ?movie a em:MediaContent ;
         schema:name ?movieName ;
         schema:datePublished ?releaseDate .
  FILTER(?releaseDate > "2010-01-01"^^xsd:date)
}
```

11. Movies per genre

Counts how many movies belong to each genre, grouping results by `schema:genre`.

```
SELECT ?genre (COUNT(?movie) AS ?movieCount)
WHERE {
  ?movie a em:MediaContent ;
         schema:genre ?genre .
}
GROUP BY ?genre
```

12. Production companies and their movies

Lists production companies and the movies they've produced, using `em:producedBy`.

```
SELECT ?companyName ?movieName
WHERE {
  ?company a em:ProductionCompany ;
    schema:name ?companyName ;
    em:producedBy ?movie .
  ?movie schema:name ?movieName .
}
```

Conclusion

This project showcases the design and use of a custom semantic vocabulary tailored for the entertainment and media industry. By following RDF/OWL standards and incorporating established ontologies like schema.org, it provides a solid framework for organizing and querying entertainment-related data. The sample data and SPARQL queries demonstrate its practical value, showing how semantic technologies can enhance data management and make systems more interoperable in the media world.

Creating this vocabulary was important because existing general-purpose ontologies, like schema.org, don't fully capture the unique relationships and details of the entertainment industry. Media content involves complex connections—actors starring in films, franchises linking multiple works, streaming platforms hosting content, and awards recognizing achievements. These complexities require a specialized structure to represent the data accurately.

Working on this vocabulary has been a great learning experience for me. It's fascinating to see how semantic technologies can simplify complex data and make it easier to work with. This project has not only helped me understand the potential of these tools but also how they can make a real difference in industries like entertainment.

Sources

Ontology for Media Resources 1.0, [no date]. Online. Available from: <https://www.w3.org/TR/mediaont-10/> [Accessed 12 November 2024].

Schema.org - Schema.org, [no date]. Online. Available from: <https://schema.org/> [Accessed 12 November 2024].