

# Direct Mapping and Turtle Ontology for Management of Indonesian Movies Knowledge

*Ayu Novira Shinta Permatasari<sup>1</sup> and Herlina Jayadianti<sup>2\*</sup>*

<sup>1,2</sup> Informatics Department, Faculty of Engineering, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

**Abstract.** Web 2.0 or conventional web has developed into Web 3.0, known as semantic web. Semantic web technology requires ontology as the backbone in understanding a concept of knowledge. In the ontology computing process, the Resource Description Framework (RDF) is used as a framework to define web resources in triple form (subject-predicate-object) so that they can form metadata and describe the information contained on the web. The data used in this study is Indonesian movies data obtained from Kaggle in Comma Separated Values (.csv) format with a total of 242 lines of Indonesian movies data. The data processing is carried out by direct mapping using the help of DB2Triples to generate data from MySQL into RDF in turtle format (.ttl) file. The results shown that direct mapping can be used to map data from RDB to RDF semi-automatically. The data is mapped into the RDF according to the schema on the RDB without input from the user, so the results provided cannot be adjusted to the needs or desires of the user. Furthermore, the RDF generated in the turtle file format has formed classes and individuals automatically, but to be able to be used as a semantic web resource, RDF needs to be processed manually to form data properties and object properties, as well as assigning instance values.

**Keywords:** RDB to RDF, Direct Mapping, Turtle Ontology.

## 1 Introduction

Web 2.0 or conventional web has developed into Web 3.0 or semantic web. The main difference between the semantic web and the conventional web is that a computer or machine has the ability to define data in the form of metadata, then machine will study the data to become knowledge, so that the computer can understand the meaning of the data that is on a website.

Semantic web technology requires ontology as a backbone in understanding a concept of knowledge. Ontology is a process of assigning meaning, properties, and relations to data with the aim of representing a knowledge domain and forming a knowledge base [1]. In the ontology computing process, Resource Description Framework (RDF) is used as a framework to define web resources in triple form (subject-predicate-object) so that they can form metadata and can describe the information contained on the web [2].

Currently the majority of data on the web is stored in the Relational Database (RDB) not in RDF [3]. RDB is a type of database that uses a relational model, where the data stored is arranged in the form of relations or tables [4]. In order for the conventional web to migrate to the semantic web, the data that has been stored in RDB must be converted into RDF. Moving the data can be a

problem, because it is impossible to do it manually one by one to move data that is already large.

The World Wide Web Consortium (W3C) which is the standard organization for the World Wide Web in 2012 provided recommendations regarding mapping languages to map RDB to RDF, namely Direct Mapping and R2RML. Direct Mapping is a mapping language that can map data from RDB to RDF automatically according to the relational schema without requiring user input, while in R2RML the user can make adjustments between the relational schema and the desired target ontology [5].

Related research to Direct Mapping has been carried out by [6] stating that Direct Mapping can meet the need for a simple and automatic method in mapping RDB to RDF by utilizing query preservation where each query through a relational database can be translated into an equivalent query through the RDF graph.

The search for information provided by the conventional web regarding Indonesian movies comes from data stored in RDB with limited relations, so the information provided is less informative. If the conventional web wants to migrate to the semantic web, there is a layer difference between the two where data on the conventional web is stored in RDB while the semantic web uses RDF.

\* Corresponding author : [herlina.jayadianti@upnyk.ac.id](mailto:herlina.jayadianti@upnyk.ac.id)

The direct mapping research method will be applied to extract Indonesian movies data from RDB and map the data into RDF in turtle (.ttl) file format with the help of DB2Triples software. Furthermore, the resulting RDF will be processed in Protégé to be used as a resource in building a semantic web of Indonesian movies searches that provides searches related to Indonesian movies knowledge.

## 2 Research Methods

### 2.1 Related Research to Mapping RDB to RDF

Research [7] uses D2RQ as a declarative language that can describe the mapping between relational database schemas and the RDF-S/OWL ontology. D2RQ works by rewriting RDQL queries and Jena API calls to application-model SQL queries. This set of SQL query results is converted into a triple RDF to be passed to the higher layers of the Jena framework. The results obtained using this method are virtual RDF graphs that can only be read, but can be accessed simultaneously by the RDF application or can be published on the Web using the RDF Net API.

Research [8] has succeeded in transferring data in the database into an ontology using DB2OWL by detecting certain cases for conceptual elements in the database and manage them so that it can convert database components into appropriate ontology components. However, this method can only be used on Oracle and MySQL databases which already provide a specific view of database metadata.

Research [9] has developed a method of constructing SQL sentences based on the creation of a mapping database schema. This method then adds the appropriate mapping records and generates a triple RDF from the RDB. With the RDB2OWL method, OWL ontologies can be generated automatically, but the resulting translation will increase the size of the mapping and reduce its readability.

Large mapping size will require large storage space as well. For this reason, [10] exposes a relational database as an RDF Schema by extracting the RDB using a reverse engineering method, then analyzing the metadata to build ontology entities using a set of rules that have been created. The advantage of this method is that it can generate dynamic RDF Virtual Graphs according to user needs, and can save storage space. However, the rules used in this method are still made manually, so they can only be applied by experts or experienced ones.

Research [11] has defined a simple, practical and intuitive interpretation of SQL database tables as RDF graphs by applying Direct Mapping method. This method takes care of all the essential features of SQL tables for breaking down non-RDF domain data representations into "live" domain agnostic representations, as well as leaving domain-specific manipulations to RDF tools. Another study related to this method [6] that Direct Mapping succeeded in generating OWL ontologies automatically and without user input. This method also fulfills two basic properties

consisting of information preservation and query preservation.

### 2.2 Data collection

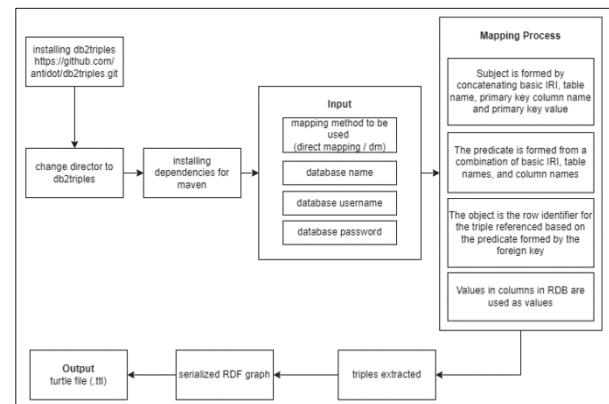
The data used in this study is Indonesian movies data obtained from Kaggle in Comma Separated Values (.csv) format with a total of 242 lines of Indonesian movies data. The data in the (.csv) file is processed into a MySQL relational database with the name `indonesian_movies` which consists of three tables, `ratings` with `id_rating` as the primary key (PK) containing data on audience ratings, `genres` with `id_genre` as PK containing data movie classification, and `movies` contains data about movies with `id_movie` as PK, `id_genre` and `id_rating` as foreign key, also contains synopsis, actors, directors, `user_ratings`, votes, and runtime.

### 2.3 Data processing

The data processing is carried out by direct mapping using the help of DB2Triples to generate data from MySQL into RDF in turtle format (.ttl) file. The RDF obtained from the direct mapping process will produce classes and instances, then ontology processing is carried out in protégé for creating object properties, data properties, and assigning values to instances.

### 2.4 Direct Mapping Process

Direct Mapping (DM) is an automatic transformation method used to generate RDF from RDB [12]. RDF graph or direct graph is described by using schema and data, primary key, and foreign key in relational database as input. The DM process in this study was carried out with the help of DB2Triples with the stages that can be seen in Figure 1.



**Fig. 1.** Direct Mapping Process

### 2.5 Ontology Processing in Protégé

The resulting RDF in the DM process is then processed in the protégé application. The turtle file will automatically form classes and instances, then form object properties, data properties, and assign values to instances.

### 2.5.1 Class definition

Classes on the Indonesian movies knowledge ontology consist of genres, ratings, and movies classes according to the results of direct mapping. Then to adjust knowledge, a people class was formed manually with two sub classes consisting of actors to manage knowledge about actors and directors to manage knowledge about directors. Ontology is also given an additional class to add knowledge, namely the awards class which contains knowledge about Indonesian movie awards with two sub classes, namely Festival Film Bandung and Festival Film Indonesia.

Ontology is also added to one class to increase knowledge, namely the awards class which contains knowledge about Indonesian movie awards with two sub classes such as Festival Film Bandung and Festival Film Indonesia.

### 2.5.2 Properties Definition

There are two types of properties consisting of object properties and data properties. Object property is a relation between classes that will act as a predicate. The object properties needed in this research are genre\_of, has\_genre, rating\_of, has\_rate, directed, directed\_by, played\_by, played\_in, nominated\_for and nominated\_by.

The data property is an attribute of each class. Class genres has data property id\_genre and genre, class ratings have data property id\_rating, rating\_code, and rating\_details, class movies have data property id\_movie, title, year\_released, synopsis, user\_rating, votes, languages, and runtime, class actors have data property actor\_name, class directors have data property director\_name, and class awards has data property award\_category.

### 2.5.3 Value of instances

Giving the value of instances is done by filling in the property value of each instance according to their needs.

## 2.6 System planning

The system will be built by providing Indonesian movies search services. Users can directly use the services provided by entering search keywords. When the user enters a search keyword, the system will perform a search, if the entered keyword is incorrect then the process returns to the keyword input stage, but if the keyword entered is correct then the search results will be displayed and the process is complete.

## 3 Result And Discussion

### 3.1 RDB to RDF Mapping

Mapping data from RDB to RDF with a direct mapping model produces an RDF graph in a turtle file which is written in the form of triples consisting of subjects,

predicates, and objects based on the schema and data in RDB. The resulting RDF can be seen in Program Code 1.

---

#### Program Code 1: RDF Graph

```
<http://foo.example/DB/movies/id_movie=m0001> a
<http://foo.example/DB/movies>;
<http://foo.example/DB/movies#actors> "Adipati Dolken, Mawar Eva de Jongh, Sarah Sechan, Jourdy Pranata, Tubagus Ali, Ravil Prasetya, Nagra Kautsar Pakusadewo, Canti Tachril, Thalia Basir, Sp Lili, Anna Tairas";
<http://foo.example/DB/movies#directors> "Rako Prijanto";
<http://foo.example/DB/movies#id_genre> "g01";
<http://foo.example/DB/movies#id_movie>
"m0001";
<http://foo.example/DB/movies#languages>
"Indonesian";
<http://foo.example/DB/movies#ref_id_rating>
<http://foo.example/DB/ratings/id_rating=r02>;
<http://foo.example/DB/movies#runtime> 100;
<http://foo.example/DB/movies#synopsis> "Ayudia (Mawar De Jongh) is not satisfied enough in enjoying the moment as a wife, but she suddenly becomes pregnant.";
<http://foo.example/DB/movies#title>
"#FriendButMarried 2";
<http://foo.example/DB/movies#users_rating>
6.5E0;
<http://foo.example/DB/movies#votes> 120;
<http://foo.example/DB/movies#year_released>
"2020"^^
<http://www.w3.org/2001/XMLSchema#date> .
```

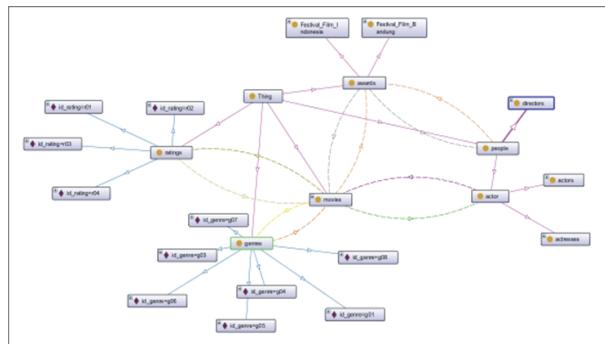
---

Based on Program Code 1, the subject is <http://foo.example/DB/movies/ id\_movie=m0001> which is formed by combining the basic IRI, table names (movies), primary key column names (id\_movie) and primary key value (m0001). The predicate is represented by each column that is formed from the concatenation of the base IRI, table name, and column name such as <http://foo.example/ DB/movies#title>, and the values in the columns in the RDB are the values in the RDF literals. Each foreign key returns a triple with a predicate consisting of the foreign key column name, the referenced table, and the referenced column name, so the object of these triples is the row identifier for the referenced triple like <http://foo.example/DB/movies#ref\_id\_genre> <http://foo.example/DB/genres/id\_genre=g01>.

The RDF graph generated by direct mapping is compiled based on the schema and data on the RDB automatically that the results provided are adjusted to the database, so that adjustments to the needs are done manually at the ontology processing stage in the protégé.

### 3.2 Ontology Processing in Protégé

The Indonesian Movies ontology is processed according to the properties that have been designed and assigned values according to the needs of each instance. The processed ontology results are depicted in an ontograph as can be seen in Figure 2.



**Fig. 2.** Indonesian Movies Ontograph

Fig. 2 shows that the ontology has loaded each class as designed, there are also relationships between classes according to object properties which are represented by dotted arrows and instances are indicated by arrows in class ratings and genres.

After the ontology processing, SPARQL query testing is performed to ensure that the ontology processing contains the appropriate knowledge. The first test was carried out to display knowledge about movies with a movie title filter. SPARQL query for the first test can be seen in Program Code 2.

**Program Code 2:** SPARQL Query Test with Movie Title

```
SELECT ?title ?actor ?director ?genre ?rating
?award
WHERE{ ?movies :title ?title.
?movies :played_by ?actors.
?actors :actor_name ?actor.
?movies :directed_by ?directors.
?directors :director_name
?director.
?movies :has_genre ?genres.
?genres :genre ?genre.
?movies :has_rating ?ratings.
?ratings :rating_code ?rating.
?movies :nominated_for ?awards.
?awards :award_category ?award.
FILTER(regex(str(?title),
"dilan", "i")) }
```

"SELECT" is a class declaration that will be displayed on knowledge results, "WHERE {}" is a term or condition, while "FILTER" is used to sort knowledge results, where in this test "? title" is used to sort according to the movie title and "dilan" to define the title that want to be displayed. The test results can be seen in Figure 3.

judul	pemeran	surakara	genre	rating	penghargaan
"Dilan 1991"	"Jerome Kurnia"	"Pidi Baiq"	"Drama"	"13+."	"Belum ada penghargaan yang drafit"
"Dilan 1991"	"Aditya Zara"	"Pidi Baiq"	"Drama"	"13+."	"Belum ada penghargaan yang drafit"
"Dilan 1991"	"Iqbaal Dhafikhi Ramadhan"	"Pidi Baiq"	"Drama"	"13+."	"Belum ada penghargaan yang drafit"
"Dilan 1990"	"Iqbaal Dhafikhi Ramadhan"	"Pidi Baiq"	"Drama"	"13+."	"Best Original Song"
"Dilan 1990"	"Iqbaal Dhafikhi Ramadhan"	"Pidi Baiq"	"Drama"	"13+."	"Best Actor"
"Dilan 1990"	"Aditya Zara"	"Pidi Baiq"	"Drama"	"13+."	"Best Original Song"
"Dilan 1990"	"Aditya Zara"	"Pidi Baiq"	"Drama"	"13+."	"Best Actor"
"Dilan 1990"	"Brandan Salin"	"Pidi Baiq"	"Drama"	"13+."	"Best Original Song"
"Dilan 1990"	"Brandan Salin"	"Pidi Baiq"	"Drama"	"13+."	"Best Actor"

**Fig 3.** The SPARQL query used in this test can be seen in Program Code 4.

**Program Code 4:** SPARQL Test Query with Cast Name

```
SELECT ?actor ?title ?award
WHERE{ ?actors :actor_name ?actor.
?actors :played_in ?movies.
?movies :title ?title.
?actors :nominated_for
?awards.
?awards :award_category
?award.
FILTER(regex(str(?actor),
"iqbaal", "i")) }
```

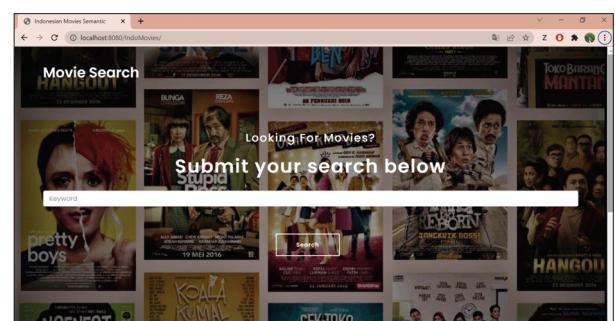
Program code 4 is used to display knowledge based on the movie's cast using the "? actor" filter, then "iqbaal" which is the name of the actor is used to sort the results displayed. The results of the second test can be seen in Figure 4.

peneman	judul	penghargaan
"Iqbaal Dhafikhi Ramadhan"	"This Earth of Mankind"	"Best Actor"
"Iqbaal Dhafikhi Ramadhan"	"The Earth of Mankind"	"Peneman Utama Pria Terpuji"
"Iqbaal Dhafikhi Ramadhan"	"Friendship Award"	"Best Actor"
"Iqbaal Dhafikhi Ramadhan"	"Dilan 1991"	"Peneman Utama Pria Terpuji"
"Iqbaal Dhafikhi Ramadhan"	"Dilan 1991"	"Best Actor"
"Iqbaal Dhafikhi Ramadhan"	"Dilan 1990"	"Peneman Utama Pria Terpuji"
"Iqbaal Dhafikhi Ramadhan"	"Dilan 1990"	"Best Actor"
"Iqbaal Dhafikhi Ramadhan"	"Dilan 1990"	"Peneman Utama Pria Terpuji"

**Fig. 4.** SPARQL Query Results Showing Knowledge of Movies with the Actor Name Dilan

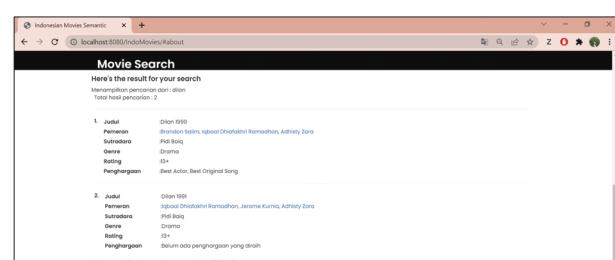
### 3.3 Semantic Web Interface

The search page is a page for users to search based on keywords as well as the first page that will appear when the Indonesian movies search website is accessed. The search page can be seen in Figure 5.



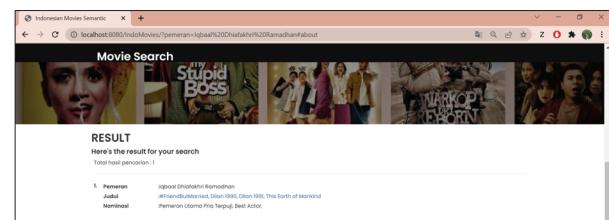
**Fig. 5.** The Search Page

Based on Fig. 5, there is a form to enter search keywords and a search button to run the search process. The results obtained based on the keywords entered will be displayed on the search results page as shown in Figure 6.



**Fig. 6.** Search Results Page

Fig. 6 shows the search results when the keyword "dilan" is entered. The results displayed are knowledge about movies with the titles "Dilan 1990" and "Dilan 1991" such as knowledge about actors, directors, genres, ratings, and awards that have been achieved. Knowledge of the cast can be displayed further by selecting the name of the actor that wanted to know. Figure 7 shows the knowledge about the movies that were played and the awards won by the actors when the name of the actor "Iqbaal Dhiafakhri Ramadhan" was chosen.



**Fig. 7. Search Results Page About the Cast**

### 3.4 Results Test

The tests carried out in this study are used to test the results displayed by the system through the interface that has been built. The results test can be seen in Table 1.

**Table 1. Test Result**

No	Test Case	Procedure	Expected results	Results shown	Status
1	Test by entering the keyword "dilan"	1. Open the search page. 2. Enter the test keyword. 3. Press the search button.	System will display knowledge about Indonesian movies whose titles contain the word "dilan" on the search results page.	System displays information about movies with the titles "Dilan 1990" and "Dilan 1991".	Show knowledge
2	Test with cast selection based on movies information displayed in test case 1	1. Perform procedure test case 1. 2. Press the name of the actor that want to search based on the search results in test case 1.	System will display knowledge about the actor being searched for on the search results page.	System displays information about the cast according to the name of the actor being searched for.	Show knowledge
3	Test with the selection of movies titles based on the cast information displayed in test case 2	1. Perform procedure test case 2. 2. Pressing the title of the movie that want to search based on the search results in test case 2.	System will display knowledge about the movies with the title you are looking for on the search results page.	System displays information about the movie according to the title of the movie being searched for.	Show knowledge

## 4 Conclusion

Based on the results of research that has been done regarding mapping data from RDB to RDF using direct mapping and turtle ontology to serve as semantic web resources, it can be concluded that direct mapping can be used to map data from RDB to RDF semi-automatically. The data is mapped into the RDF according to the schema on the RDB without input from the user, so the results provided cannot be adjusted to the needs or desires of the user. Furthermore, the RDF generated in the turtle file format has formed classes and individuals automatically, but to be able to be used as a semantic web resource, RDF needs to be processed manually to form data properties and object properties, as well as assigning instance values.

## References

- [1.] D. Fensel, J. A. Hendler, H. Lieberman, and W. Wahlster: *Spinning the Semantic Web: bringing the World Wide Web to its full potential*: The MIT Press (2003).
- [2.] S. Auer, L. Feigenbaum, D. Miranker, A. Fogarolli, and J. Sequeda: Use Cases and Requirements for Mapping Relational Databases to RDF: *Database*: no. June 2010, pp. 1–20, (2010).
- [3.] M. Hert, G. Reif, and H. C. Gall: A Comparison of RDB-to-RDF Mapping Languages Categories and Subject Descriptors: *7th Int. Conf. Semant. Syst. (I-Semantics '11)*: no. September, pp. 25–32, (2011).
- [4.] O. Y. Saygili: Relational Database Management System (RDBMS): *Introd. to Priv. Cloud using Oracle Exadata Oracle Database*: pp. 5–8, (2020).
- [5.] L. E. T. Neto, V. M. P. Vidal, M. A. Casanova, and J. M. Monteiro: R2RML by assertion: A semi-automatic tool for generating customised R2RML mappings: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*: vol. **7955** LNCS, pp. 248–252, (2013).
- [6.] J. F. Sequeda, M. Arenas, and D. P. Miranker: On directly mapping relational databases to RDF and OWL: *WWW'12 - Proc. 21st Annu. Conf. World Wide Web*, pp. 649–658, (2012).

- [7.] C. Bizer, F. U. Berlin, W. Semantik, and D. R. Q. Jena: D2RQ – Treating Non-RDF Databases as Abstract Virtual RDF Graphs D2RQ Example Mapping Language, pp. 2–3, (2003).
- [8.] N. Cullot, R. Ghawi, and K. Yétonnon: DB2OWL: A tool for automatic database-to-ontology mapping: *SEBD 2007 - Proc. 15th Ital. Symp. Adv. Database Syst.*, pp. 491–494, (2007).
- [9.] G. Bumans and K. Čerans: RDB2OWL: A practical approach for transforming RDB data into RDF/OWL: *ACM Int. Conf. Proceeding Ser.*, pp. 1–3, (2010).
- [10.] H. Y. Ling and S. F. Zhou: Translating relational databases into RDF: *2010 2nd Conf. Environ. Sci. Inf. Appl. Technol. ESIAT 2010*, vol. **3**, pp. 464–467, (2010).
- [11.] M. Arenas, A. Bertails, E. Prud'hommeaux, and J. F. (Eds. . Sequeda: “A Direct Mapping of Relational Data to RDF: *W3C Work. Draft 20 Sept. 2011*: no. September (2012).
- [12.] F. Michel, J. Montagnat, and C. Faron-Zucker: A survey of RDB to RDF translation approaches and tools: *Informatique, Signaux Et Systèmes*, p. 23, (2014).