

Issues and Challenges in the Extraction and Mapping of Linked Open Data Resources with Recommender Systems Datasets

Rosmamalmi Mat Nawi* 

Centre for Artificial Intelligence Technology,
Faculty of Information Science and Technology,
The National University of Malaysia, Selangor, Malaysia
E-mail: rosmamalmi@siswa.ukm.edu.my

Shahrul Azman Mohd Noah 

Centre for Artificial Intelligence Technology,
Faculty of Information Science and Technology,
The National University of Malaysia, Selangor, Malaysia
E-mail: shahrul@.ukm.edu.my

Lailatul Qadri Zakaria 

Centre for Artificial Intelligence Technology,
Faculty of Information Science and Technology,
The National University of Malaysia, Selangor, Malaysia
E-mail: lailatul.qadri@.ukm.edu.my


ABSTRACT

Recommender Systems have gained immense popularity due to their capability of dealing with a massive amount of information in various domains. They are considered information filtering systems that make predictions or recommendations to users based on their interests and preferences. The more recent technology, Linked Open Data (LOD), has been introduced, and a vast amount of Resource Description Framework data have been published in freely accessible datasets. These datasets are connected to form the so-called LOD cloud. The need for semantic data representation has been identified as one of the next challenges in Recommender Systems. In a LOD-enabled recommendation framework where domain awareness plays a key role, the semantic information provided in the LOD can be exploited. However, dealing with a big chunk of the data from the LOD cloud and its integration with any domain datasets remains a challenge due to various issues, such as resource constraints and broken links. This paper presents the challenges of interconnecting and extracting the DBpedia data with the MovieLens 1 Million dataset. This study demonstrates how LOD can be a vital yet rich source of content knowledge that helps recommender systems address the issues of data sparsity and insufficient content analysis. Based on the challenges, we proposed a few alternatives and solutions to some of the challenges.

Keywords: recommender system, linked open data, DBpedia

Received: March 4, 2021
Accepted: May 11, 2021

Revised: April 25, 2021
Published: June 30, 2021

***Corresponding Author:** Rosmamalmi Mat Nawi
 <https://orcid.org/0000-0003-1147-3092>
E-mail: rosmamalmi@siswa.ukm.edu.my



All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

1. INTRODUCTION

Recommendation Systems (RS) technology is currently a successful solution to addressing knowledge overload problem with the ever-increasing amount of online knowledge. Given its widespread use in many web applications, the importance of RS cannot be overstated. Due to the issue of sparseness, cold start, overspecialization, and even the nature of scalability in recommender systems, many researchers have done a lot of research to overcome these limitations by adapting diverse recent technologies. Among the applied technologies are big data (Al-Barznji & Atanassov, 2017; Hammou et al., 2019; Maillo et al., 2017), semantic (Ameen, 2019; Barros et al., 2020; Figueroa et al., 2019), and deep learning (Feng et al., 2019; Sankar et al., 2020; Zhang et al., 2019).

RS quality often degrades when insufficient information or few ratings are available, which is a commonly well-known challenge in RS. RS performance can be improved by enriching the user's profile with relevant features and envisioning the recommendations' improvements using more structured and semantically richer data about the item. This valuable information or data presented in some form of knowledge bases can be effectively utilized in RS to overcome some of its major issues, such as sparsity and cold-start. Linked Open Data (LOD) is one such semantic dataset seen to be potentially exploited by RS to improve the recommendations' quality. Nonetheless, the knowledge of LOD should not be limited only to exploiting relationships among items or to enrich data about items and users. It can also produce implicit knowledge about them and their relationships.

LOD is a relatively new topic with enormous potential in a variety of fields (Yochum et al., 2020), including RS. Linked data-based RS (referring to recommender systems applying LOD) still struggle to develop recommendations for end-users with acceptable accuracy. This is explained by the need for data from both user profiles and item descriptions (Figueroa et al., 2019). There is also very little support for users who consume information published as LOD, most of which focus on searching and visualizing the collected data (Oliveira et al., 2017). They can display related vocabulary and data in facets; however, they do not recommend relevant information to infer users' interests (Franz et al., 2010).

Datasets such as the Million Song, LastFM, and MovieLens can be further enriched by being automatically

mapped to the LOD identifiers. Then, additional features can be derived from publicly available knowledge bases like the DBpedia. For example, in the MovieLens dataset, we could get more granular features for movie representations using the LOD cloud. A large number of additional features can provide valuable information for various applications. It depends on how the data are provided for a particular item based on specific purposes. For example, if viewers like a movie with a specific director, they might also like other movies directed by the same director. In the movie domain, information about the movie's actors or directors is available via linked information. If the same actor starred in two movies, the two movies are linked. This will enable us in inferring new facts about movies, which can contribute toward more insightful details in the future. Such structured semantic data can be exploited to improve RS.

Previous research focusing on the mapping and extraction of MovieLens and IMDB can be found in the work of Peralta (2007), but the discussed challenges were mainly on the datasets and not related to linked data. Although works presented in Di Noia et al. (2012), Meymandpour and Davis (2015), and Kushwaha and Vyas (2014) describe approaches to integrate MovieLens and DBpedia, they did not describe and discuss the challenges encountered in extracting knowledge from the linked data (i.e., DBpedia) in terms of data quality and accuracy. Thus, this paper aims to discuss such challenges based on our experience and experiment in exploiting linked data to enhance the performance of recommender systems.

This paper demonstrates our study in exploring data about movies with two datasets: a public dataset of MovieLens 1 Million (ML1M), and DBpedia as the LOD dataset. Furthermore, we also present a general review of the LOD application to RS in existing works involving various elements. We then address the challenges of exploring linked data and the tasks for extracting and mapping for both used datasets. Based on the challenges, alternative solutions are proposed which may resolve some of the challenges.

To this end, we extract and integrate data provided by DBpedia and ML1M by applying the mapping set. The end of this study's contribution is that we provide a clean dataset (ML1M-DBpedia)¹ with additional information for the movie list from the ML1M, enhanced with the 'director' and 'starring' data. The primary aim is to tackle sparsity issues in the RS by enriching the dataset with spe-

¹The ML1M-DBpedia dataset is available for download at <https://data.mendeley.com/datasets/ypwk2536gh/1>

cific attributes. From these findings, it can allow practitioners in the field to analyze, evaluate, and compare their works with existing approaches.

The paper is organized as follows. Section 2 provides background information of the relevant topics of this paper which are the RS and linked data. Section 3 discusses prior works on the application of LOD in RS. Section 4 then deals with our experimental procedure and setting of mapping ML1M to DBpedia. We discuss the challenges of the linked and extraction process between the two datasets in Section 5 and finally, Section 6 presents conclusion of our work.

2. LINKED OPEN DATA ON RECOMMENDER SYSTEM

The application of Semantic Web (SW) technologies seems to offer promising solutions to the long-standing issues of recommender systems such as data sparsity and cold-start problems (Saat et al., 2018). This is due to the fact that SW contains structured knowledge which can be retrieved and exploited. Defining information using SW concepts will enable data to be retrieved and integrated in more meaningful and effective ways (Hidayat et al., 2012) by incorporating data from a wide range of sources.

Using different knowledge representation languages such as Resource Description Framework (RDF), W3C Web Ontology Language, and Uniform Resource Identifier (URI) protocols has allowed data to be interoperable. As a consequence, the data can be shared and reused across a wide range of platforms, applications, and communities. Yadav et al. (2020) declare that the aim is to link data from various isolated applications, highlight the importance of publicly bringing out data for other applications to use, and connect the data using standard schema.

LOD denotes freely accessible semantic data on the Web. Combining Linked Data and Open Data yields the concept of LOD. LOD's aim is to make data semantically linked and machine-processable (Alshammari & Nasraoui, 2019). The DBpedia is one of LOD's significant contributions (Auer et al., 2007), which has been recognized as one of the valuable sources for building RS. DBpedia data can be easily and readily extracted using the SPARQL query language and the defined properties mentioned in its ontology. This valuable information presented in a linked knowledge base can be effectively utilized in RS.

The existence of interconnected LOD datasets allows

federated queries to be performed between different datasets (Cappiello et al., 2016). Thus, users can integrate data obtained from other information sources by leveraging the links between the datasets. Furthermore, with the LOD cloud containing relevant information to extend the feature set and increase granularity (Kuchař, 2015), it could produce new and unique data. So, the newly gained (or inferred) knowledge from LOD is seen as a potential to address the issues of serendipity in recommendations, as discussed by Saat et al. (2018) and Deshmukh et al. (2018).

2.1. DBpedia

DBpedia is a dataset containing Wikipedia data and a project to make it accessible on the Web in a standardized manner. It helps one to query and integrate Wikipedia data with other web sources. Around 4.6 million resources are defined in the English version of DBpedia, with the majority of them (about 4.2 million) classified using the DBpedia ontology. There are about 60 properties for the *Film*² class in DBpedia.

A rich set of semantic data (referring to LOD) is accessible through SPARQL endpoints. DBpedia alone has billions of triples covering a wide range of topics (e.g., people, places, films, and books). Despite the fact that different endpoints may employ different vocabularies and ontologies, they all share a standard platform to access and retrieve semantic data, which is the SPARQL query language. Based on Ferré (2017), composing SPARQL queries is a time-consuming and error-prone task and is primarily inaccessible to most prospective users of semantic data. However, being a W3C standard that is widely utilized, SPARQL's expressivity and scalability for huge RDF stores are two of its advantages. This is likely to have contributed to SPARQL engines that are highly optimized (e.g., Virtuoso, Jena TDB).

There are over 3 billion triples in the DBpedia instance and it is one of the primary interlinking hubs in the LOD (Lehmann et al., 2015). As the data are envisioned for machine-to-machine utilization, however, an acceptable presentation makes them adversely relevant for end-users (Raynaud et al., 2018). Details about users or items are also freely accessible in the LOD cloud in RDF format. This information can be easily retrieved using a SPARQL query endpoint by providing only two types of information: the URI of the required resource and the name of its attribute.

²Available at <http://mappings.dbpedia.org/server/ontology/classes/Film>

2.1. Recommender Systems

Recommender Systems (RS) delineates a decision-making approach for users on complex information platforms. It can effectively recommend the relevant information to end-users (Zarzour et al., 2018). It is used in our daily lives for various reasons, including the recommendation of products, movies, music, news, and books. These recommendations are based on the user's historical behaviour and other users' behaviour having a similar taste.

The fundamental concept of the RS is based on the premise that if users share the same habits when shopping in online shops, choosing similar artists, reading the same news, or watching the same movies, they are assumed to have similar behaviours in the future. In other words, the RS predicted users' expectations for products and suggested objects so that other users will possibly like them. Currently, recommendation algorithms are widely used in many applications including Amazon, Spotify, Facebook, and TripAdvisor.

The RS can analyze items in two ways: explore items in terms of how users like them or how similar users like the same items, or explore items in terms of how users popularise items, irrespective of the user's similarities. Based on the two fundamental principles of the RS, it can be classified by the main three well-known techniques: Content-based filtering (CB) (Albatayneh et al., 2018; Lops et al., 2019), Collaborative filtering (CF) (Franz et al., 2010; Osman et al., 2020), and Hybrid-based filtering (HB) (Aggarwal, 2016; Jannach et al., 2010). HB technique combines one or more approaches to compensate for the weaknesses of the single methods. Another technique is Knowledge-based (KB), which is inferred between user requirements and items' features described in a certain knowledge base (Dell'Aglio et al., 2010; Ameen, 2019).

A large amount of research has focused on movie recommendation (Jamil et al., 2020; Vilakone et al., 2018), music recommendation (Nguyen et al., 2015; Vall et al., 2019), health recommendation (Nilashi et al., 2020; Stratiği et al., 2020), digital library recommendation (Beel et al., 2016), e-learning (Albatayneh et al., 2018; Pereira et al., 2018), e-commerce recommendation (Chu et al., 2020; Petrova et al., 2019), tourism recommendation (Delic et al., 2018; Yochum et al., 2020), and many other areas.

Recently, Group Recommender Systems (GRS) have become an effective tool for consulting and recommending items according to a group of like-minded users' choices. As the name suggests, a GRS is a system that provides recommendations to a group of users. GRS has been designed, instead of providing one or more items to indi-

viduals, to concurrently recommend them to a group with similar interests to satisfy each of them (Nozari & Koochi, 2020). Furthermore, domains such as movies, restaurants, and tourism may typically deal with a group of users. Some works that deal with recommendations to a group of users are Roy et al. (2018), Nawi et al. (2020), and Wang et al. (2020).

2.1.1. Issues in Recommender Systems

Scalability, high computation, diversity (Yadav et al., 2020), and gray sheep (Erion & Maurizio, 2017) are among the concerns when designing an efficient recommendation system. However, above all, two core issues have piqued researchers' interest in the RS area, which are the *cold-start* and *data sparsity* problems. The former occurs during registration of a new user or adding up new resources or items in the system. Apparently, there will be no details in the system regarding the user's interest or preferences for any particular item. In contrast, the latter arises because users usually rate a small portion of the available items. Since there are few ratings, it is uncertain that two users or objects would have similar ratings. As a result, scores are predicted based on a small number of neighbours, which may degrade the recommendation quality since there is insufficient data, or no ratings at all are available (Tomeo et al., 2015). Some works that tackle these issues are by means of item correlation in social networks (Cao et al., 2019; Hong et al., 2015), manipulating user profile demography (Bañas et al., 2015; Xu et al., 2018), enhanced filtering and prediction technique (Mohamed et al., 2019; Ortega et al., 2016), and exploiting external knowledge such as the DBpedia (Meymandpour & Davis, 2015; Di Noia et al., 2016).

3. PRIOR WORK

Several works that address recommendations in numerous domains have been suggested in the literature, even though very few methods take advantage of the LOD initiative to provide efficient recommendations. LOD was often used to mitigate the cold-start and data sparsity issues associated with collaborative recommendation (Vall et al., 2019; Yadav et al., 2020). This section presents the prevalent research work of semantic-based approaches to the RS. As in Table 1, we categorize the previous research on RS applying semantic technology based on both elements of RS and LOD.

LOD-based recommendation research has its emergence in the ground of ontology-based RS pioneered by

Table 1. Application of LOD in Recommender Systems

| Study | Domain / Dataset | RS elements | LOD elements |
|--------------------------------|---|---|---|
| Yadav et al. (2020) | Movie / ML1M + Yahoo! Webscope R4 + DBpedia + FOAF | - New User Cold Start - HB - User profile (social network) | - Item similarity-based ontology - FOAF |
| Figuerola et al. (2019) | DBpedia + iMDB | - Novelty & accuracy - Heterogeneous information network | - ALLied framework - Simple Knowledge Organization System (SKOS) |
| Zhao et al. (2019) | Multidomain / Freebase + ML20M + Amazon + LFM-1b | - KB - Heterogeneous information network | - Linkage analysis: popular & recency |
| Alshammari and Nasraoui (2019) | Multidomain / ML100k + Book crossing + DBpedia | - CF model-based, matrix factorization - Explanation of recommendation | - Semantic knowledge graph - Semantic measure distance: LDS |
| Iana et al. (2019) | Conference / SciGraph + WikiCFP | - CB - Recommendation-based (author, keyword, abstract) - Neural network approach | - Computer Science Ontology (CSO) |
| Nilashi et al. (2018) | Movie / iMDB + ML1M + Yahoo! Webscope R4 | - CF model-based (SVD) - Isu scalability and sparseness | - Film ontology (http://www.movieontology.org/) - Semantic similarity between item |
| Pereira et al. (2018) | Education / Open University | - User profile and interest in education (social network) | - Semantic representation (ontology of FOAF and SIOC) - Semantic search |
| Srinivasan and Mani (2018) | Movie / ML20M | - topN recommendation - Diversity of recommendation | - Movie based graph (LOD) |
| Palumbo et al. (2017) | Movie / ML1M | - HB - topN recommendation - Heterogenous information network | - Knowledge graph |
| Vagliano et al. (2017) | Multidomain / iMDB + LibraryThing + Amazon + DBpedia + Wikidata + Yago + ML1M | - User review - Novelty, diversity recommendation - topN recommendation | - Semantic annotation, AIDA |
| Musto et al. (2017) | Multidomain / ML1M + DBBook + LastFM | - CF - Graph-based recommendation | - Integrate knowledge from LOD cloud - Feature selection (semantic-aware recommendation) |
| Piao and Breslin (2016) | Music / DBpedia | - Similarity measure - Vector Space Model | - Semantic measure distance: LDS |
| Oliveira et al. (2017) | Movie / DBpedia + LinkedMDB | - Degree of social relation (social network) | - Ontology - Semantic measure distance: LDS |
| Di Noia et al. (2016) | Multidomain / ML1M + LibraryThing + LastFM + DBpedia | - CF with topN recommendation - Feature selection, Random Forest model | - Semantic path-based |
| Meymandpour and Davis (2015) | Film / ML100k + ML1M + DBpedia + LinkedMDB + Freebase + Yago | - HB - Matrix factorization (SVD++) | - Item semantic analysis |
| Rowe (2014) | Film / MovieTweatings | - SVD++ algorithm | - Semantic aware |
| Ko et al. (2014) | Mobile TV / MobileIPTV + DBpedia + LinkedMDB + Freebase | - CF - Limited content of the item and user data | - Semantic cluster based on LOD category |

Table 1. Continued

| Study | Domain / Dataset | RS elements | LOD elements |
|---------------------------|---|---|--|
| Ostuni et al. (2013) | Multidomain / ML1M + LastFM + DBpedia | - HB (CF + CB) - topN recommendation - Random Forests, Gradient Boosted Regression Trees (GBRT) | - Graph semantic - Semantic Vector Space Model (sVSM) |
| Yang et al. (2013) | Movie / ML100k + DBTropes | - CF - Implicit data - Slope One algorithm | - Implicit item relation LOD-based - Semantic measure distance: LDS |
| Mirizzi et al. (2012) | Movie / ML100k + DBpedia + LinkedMDB + Freebase | - CB - Cold start issue - Facebook application | - Semantic similarity between item - sVSM |
| Bostandjiev et al. (2012) | Music / LastFM + DBpedia + Facebook API + Twitter API | - HB - Explaining aspect (interaction and visualisation) | - Semantic Web API, social Web |

LOD, Linked Open Data; RS, Recommendation Systems; ML1M, MovieLens 1 Million; FOAF, Friend of a Friend; SIOC, Semantically-Interlinked Online Communities; HB, hybrid-based filtering; KB, knowledge-based; CF, collaborative filtering; CB, content-based filtering; LDS, Linked Data Semantic Distance; AIDA, referring to online tool for entity detection and disambiguation.

Middleton et al. (2004). Several research efforts investigate RS based on linked data, and the wealth of data provided by the LOD cloud can be seen in Table 1. In much of the latest literature, authors have exploited DBpedia mainly to extract valuable information. As limited content or information is the core issue in the RS, LOD has significantly played a significant role, and many researchers are taking advantage of using it.

Bostandjiev et al. (2012) developed an application named “*TasteWeights*,” which is a kind of RS in which users’ preferences for music genres are extracted from Facebook. DBpedia is exploited to discover all of the music played by new artists in the same genre that the active user liked, and then recommends the same to other users. Another DBpedia-based similarity measure has been proposed by Meymandpour and Davis (2015). The authors adopted Partitioned Information Content, a similarity measure inspired by Information Theory. They adapted to LOD’s scenario to determine the similarity of two resources defined by LOD properties. The authors have utilized DBpedia extensively to define or modify various similarity measures based on LOD properties (Piao & Breslin, 2016).

More recently, Zhao et al. (2019) present the first public linked KB dataset for RS, named “KB4Rec v1.0,” which has linked three commonly used RS datasets with the popular KB Freebase. ALLied, developed by Figueroa et al. (2019), refers to a framework for selecting, evaluating, and creating algorithms to recommend resources from Linked Data (LD) belonging to different application domains. Diversity-Ensured Semantic-aware Item REcommenda-

tion (DESIRE), presented by Srinivasan and Mani (2018), deals with a consistent and dependable knowledge source in order to significantly improve quality and provide a diverse topN recommendation list.

The LD can be analyzed through the perspective of its graph structure. The study by Musto et al. (2017), Srinivasan and Mani (2018), Iana et al. (2019), and Ostuni et al. (2013) are among works focusing on graph structure linked-based in RS. LD in graph-based algorithms uses this structure to compute relevance scores for items expressed as nodes in a graph. Semantic exploration with feature selection by Musto et al. (2017) generates semantic-aware recommendations, and the authors studied the impact of the LOD’s knowledge on a graph-based recommendation algorithm’s overall performance. The results presented in Musto et al. (2017) affirm that knowledge from the LOD cloud can have a significant impact on the recommendation algorithm. In Rowe (2014), the authors developed a semantic-aware extension of the SVD++ model, named “SemanticSVD++.” It incorporates semantic item categories into the model. Meanwhile, Ko et al. (2014) apply semantics for clustering television program category and content.

Path-based algorithms compute similarities for producing recommendations using knowledge about semantic paths within a graph structure. Di Noia et al. (2016) introduced “SPRank,” which uses LOD to combine machine learning with learning the best path to consider relying on learning to rank for topN recommendation, while Ostuni et al. (2013) demonstrate how to use LOD sources and a

hybrid filtering approach to generate topN recommendations from implicit feedback. Iana et al. (2019) have been recently exploiting SciGraph to provide users with recommendations of conferences to submit their publications to and utilize it for information on past conferences and publications. They also apply WikiCfP for details on upcoming seminars.

Another exciting direction concerning the exploitation of LOD for RS is explored in social networks, as in Mirizzi et al. (2012), Oliveira et al. (2017), and Pereira et al. (2018), by implementing the semantic representation of ontology concept. In the educational domain, Pereira et al. (2018) present an infrastructure able to extract users' profiles and educational context from Facebook, and recommend educational resources. Their study proposes the techniques of information extraction and SW technologies for extraction, enrichment, and definitions of user-profiles, and interests are represented using Friend of a Friend (FOAF). The Semantically-Interlinked Online Communities ontology was used in their study to represent all groups that users join. "Entity2rec," introduced by Palumbo et al. (2017), is an approach to learning user-item relatedness from knowledge graphs for topN item recommendation. They use a knowledge graph via node2vec, encompassing collaborative information from user feedback and item information from LOD.

MORE, introduced in Mirizzi et al. (2012), is a movie recommendation application that works in tandem with Facebook. The remarkable aspect about MORE is that besides analyzing the content, it also considers the user's Facebook profile to overcome the cold-start problem in RS. Oliveira et al. (2016) consume the LOD in RS, adopting the social network and linking to more than one LOD dataset. Their work focuses on identifying related resources from different types, for instance: Books, people, soundtracks, or some other resources considered similar by the algorithm to the user's movie search can be possibly recommended.

Nilashi et al. (2018) tackle scalability and sparseness in RS by proposing a framework implementing film ontology and semantics for item similarity. Meanwhile, Vagliano et al. (2017) combine semantic annotation of user feedback with additional data from the LOD cloud. Yadav et al. (2020) solve the issue of pure new user cold-start in RS by building user profiles based on LOD, and social network-based and collaborative features. The following features are used to calculate user similarity: analyzing the social network, collaborative features from Dense User-Item Matrix, and using the information extracted from

the LOD cloud, such as FOAF.

Studies by Meymandpour and Davis (2015), Musto et al. (2014), and Oliveira et al. (2017) use LOD to obtain similarities of items in a CB approach. However, our study focuses on a CF approach with the integration of LOD technologies to overcome sparseness issues. Nilashi et al. (2018) somehow also applied a CF approach in their work. However, they use ontology and dimensionality reduction to tackle sparseness and scalability. Apart from DBpedia, Heitmann and Hayes (2010) considered different sources such as the SmartRadio and Myspace as alternatives to overcome the challenge of the data acquisition problem, as in CF approach. In a greater perspective, we see LOD as an important and influential technological medium applied in RS for various aspects, such as enriching the user and item information.

4. EXPERIMENTAL SETUP

Some of the significant attributes that involve DBpedia in the movie domain are '*dbo:director*,' '*dbo:editor*,' '*dbo:starring*,' '*dbo:producers*' and '*dbo:director*.' We apply two attributes in DBpedia, that the most commonly used in Wikipedia Web pages that portray a movie: '*dbo:starring*' and '*dbo:director*.' These two attributes are relevant in this study because they play an essential role in the selection of movies to watch.

A study by Gmerek (2015) and Carrillat et al. (2018) prove that the starring attribute has a strong correlation with box office movies. According to Rawal and Saavedra (2017), audiences have some set of expectations of a movie produced through its trailers due to the known actors and directors working for the movies. Furthermore, Kim (2013) found that top actors and directors are strongly associated with the success of movies. These arguments support that the director and starring attributes influence a film's commercial success. The same intuition applies when users decide what movies to watch, which is based on the directors and actors. Based on these arguments and findings, it justifies the choice of these two attributes in this study.

Linked data can provide rich semantic representations of the recommended items since they include exciting features (Vagliano et al., 2017). Thus, recommender systems that use linked data should be able to enrich the item or user profiles with LOD sources' information (Sansone et al., 2019), particularly from the DBpedia. For example, movies represented in DBpedia contain essential information such as starring actors and directors. As illustrated in Fig. 1, additional information about the actor who starred

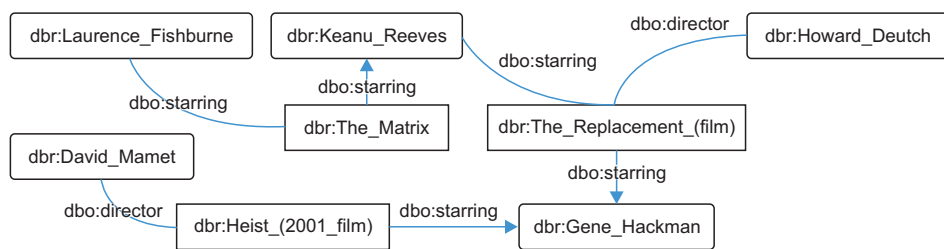


Fig. 1. Link of data relation between attribute through item.

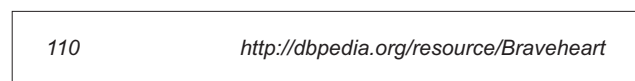


Fig. 2. Example of mapping set.

in the movie can be explored through the LOD (e.g., the relation *'dbo:starring'* existing between Keanu Reeves and *The Matrix*). This data can then be analyzed and manipulated to enhance recommendation quality based on several aspects, such as semantic analysis of items (Meymandpour & Davis, 2015; Musto et al., 2017), the implicit relationship between items (Ostuni et al., 2013; Yang et al., 2013), search queries (Wenige & Ruhland, 2018), and knowledge graph for item recommendation (Anelli et al., 2020; Palumbo et al., 2017; Piao & Breslin, 2018).

4.1. Mapping MovieLens to DBpedia

To obtain the specific information of resources from the LOD, the URI of such resources are necessary. For that reason, mapping needs to be done in order to identify the required item to the corresponding object in the linked database. It is the only entry point to the LOD. We experimented with two datasets in this study, the ML1M and the DBpedia datasets.

MovieLens is a movie recommendation project created by the University of Minnesota's Department of Computer Science and Engineering. It is a traditional CF system that gathers user movie preferences. Three datasets are available on the MovieLens website. The first one consists of 100,000 ratings (ML100k). The second has around a million ratings (ML1M) for 3,883 movies by 6,040 users, and another one contains 20 million ratings (ML20M).

The mapping set by Meymandpour & Davis (2015)³ is applied, which provides 3,303 movies' URI. Some other mappings are publicly available, such as those of Kushwaha and Vyas (2014),⁴ and Di Noia et al. (2016).⁵ Compared to the different mapping sets, Meymandpour and Davis

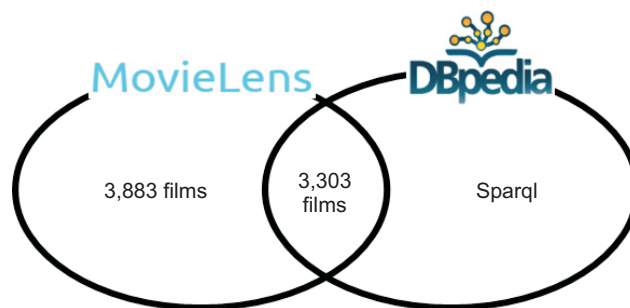


Fig. 3. Mapping MovieLens 1 Million to DBpedia.

(2015) provide more stable and consistent mappings. The example mapping for the *'Braveheart'* film is shown in Fig. 2, and note that the number 110 is referring to the 'movie_id' of the *'Braveheart'* film in the ML1M dataset.

The URI serves as an identifier for that data related to the associated movie. It can be used to extract additional information. ML1M items have been assigned to the corresponding DBpedia entities, and as mentioned earlier we leverage publicly available mappings. It is derived by running a SPARQL query based on the item's descriptive properties (e.g., the name of the movie). Since not every item in the ML1M data has a corresponding DBpedia entity, out of the 3,883 movies from ML1M, 580 URIs are missing from the mappings. We can say that it happened to reduce 580 films, as illustrated in Fig. 3, thus indicating that we obtained 964,534 ratings from 6,040 users on 3,303 items after this mapping.

Fragment mapping is a function that maps fragments to a set of endpoints, u , and refers to DBpedia in this study. A fragment f correlates to a fragment description, $fd(f)=\langle u, tp \rangle$, while a triple pattern (tp) , $tp;triples(f)$ correspond to the RDF triples of f that fulfill $fd(f)$. The dataset's tp can be accessed using $fd(f).u$. The sample of fragment description can be referred to in Fig. 4. However, the mapping for a film fragment is missing due to some retrieval

³DBpedia-MovieLens_Benchmark_Datasets_100K_1M

⁴https://www.researchgate.net/publication/297369577_mapping-movie-lens-dbpedia

⁵<https://github.com/sisinflab/LODrecsys-datasets/tree/master/MovieLens1M>

of film unavailability attributes in DBpedia. For example, the fragment can retrieve the data from the 'dbo:director' attribute but no output for 'dbo:starring.' Some of the explanations are discovered through the issue addressed in Section 5.

4.2. SPARQL Code

LOD features are then extracted from the DBpedia SPARQL endpoint in this study. The following URI pattern is used to access each resource in DBpedia:

`http://dbpedia.org/resource/<name_of_resource>`

```
triple(f):      {dbr:The_Replacements_(film)
                 dbo:starring dbr:Gene_Hackman
                 dbr:Heist_(2001_film)
                 dbo:director dbr:David_Mamet
                 dbr:The_Matrix
                 dbo:starring dbr:Laurence_Fishburne, ...}
```

Fig. 4. Fragment description.

```
""""SELECT DISTINCT ?label ?abstract ?director ?starring WHERE
{
  <http://dbpedia.org/resource/Braveheart>
  rdfs:label ?label ;
  dbo:abstract ?abstract ;
  dbo:director ?director ;
  dbo:starring ?starring
  FILTER ( lang(?label) = "en" )
  FILTER ( lang(?abstract) = "en" )
}""""
```

Fig. 5. SPARQL code for data extracting from DBpedia.

SPARQL code applies in extracting the data as can be referred to in Fig. 5. DBpedia's two properties are considered as mentioned before: 'dbo:director' and 'dbo:starring.' Based on this method, we can get the 'Braveheart' film director and starring actors. Moreover, we can get the related films directed by the same director based on the SPARQL code (Fig. 6), referring to 'dbo:director={mel_gibson}' matched to the URI `http://dbpedia.org/resource/Mel_Gibson`.

Fig. 7 illustrates this interconnection of the movie from ML1M to the DBpedia based on the director's attribute. Besides this, the related film with the same director could be figured out. A list of movies directed by 'dbo:director={Mel_Gibson}' other than 'Braveheart,' includes 'Apocalypto,' 'Hacksaw Ridge,' 'The Man Without a Face' and 'The Passion of the Christ.'

5. CHALLENGES IN EXTRACTING PROCESS

Although the LOD cloud has indeed been augmented with a large amount of information related to various domains, they still contain inconsistencies, inaccuracy, and missing data problems (Kushwaha & Vyas, 2014). One

```
""""SELECT DISTINCT ?person ?name ?about
WHERE {
  <http://dbpedia.org/resource/Braveheart> dbo:director ?directBy.
  ?person dbo:director ?directBy.
  ?person foaf:name ?name.
  ?person dbo:abstract ?about.
  FILTER (lang(?about) = "en")
} ORDER BY ?name""""
```

Fig. 6. SPARQL code for films with same directors.

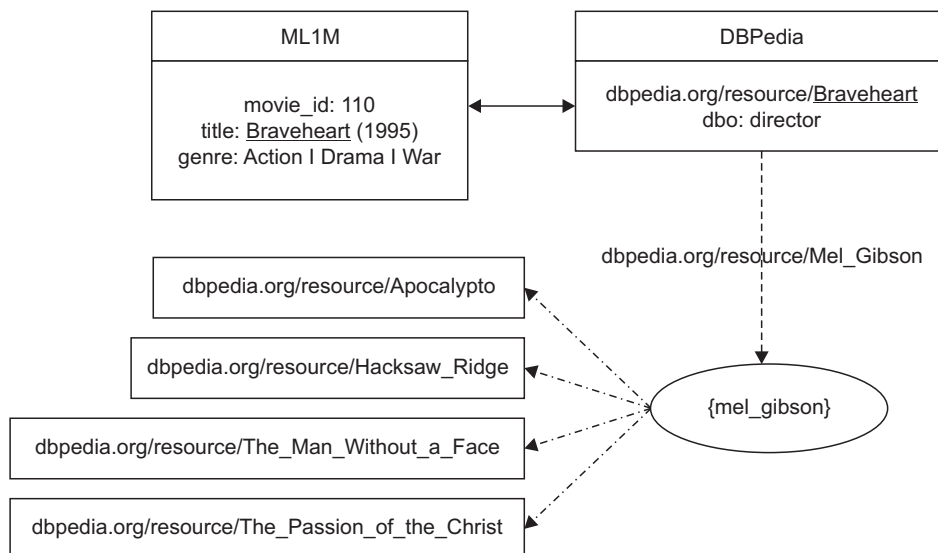


Fig. 7. Data relation for 'Braveheart' film. ML1M, MovieLens 1 Million.

of the challenging issues in the LOD is data quality. Data quality concerns issues such as inaccuracy, incompleteness, and inconsistency, which entail significant restrictions on the data's optimum utilization. Zaveri et al. (2014) also revealed that data derived from semi-structured or even structured sources frequently contains discrepancies and misrepresentations, as well as incomplete information. Despite that the mapping set is stable and consistent compared to others, we faced some challenges in extracting DBpedia's data process. The discussion is classified into seven issues, as follows.

5.1. Incomplete URI

As shown in Fig. 3, the mapping set provides 3,303 URI for the films out of 3,883, as discussed in Section 4.1. There are some situations where, even though the mapping provides the URI, hardly any result is displayed. Some examples are from the links (films) of 'Champagne ([http://dbpedia.org/resource/Champagne_\(film\)](http://dbpedia.org/resource/Champagne_(film)))' and 'Downhill ([http://dbpedia.org/resource/Downhill_\(film\)](http://dbpedia.org/resource/Downhill_(film)))'. Most probably, it does not come with the appropriate URI link. Based on the SPARQL code of getting films with the same directors (Fig. 6), it resulted that both movies come from the same director of 'Albert Hitchcock'. We also figure out, based on the same SPARQL code, that the supposed URI that should display the output is ([http://dbpedia.org/resource/Champagne_\(1928_film\)](http://dbpedia.org/resource/Champagne_(1928_film))) for the 'Champagne' film and ([http://dbpedia.org/resource/Downhill_\(1927_film\)](http://dbpedia.org/resource/Downhill_(1927_film))) refers to the 'Downhill' film.

5.2. Contradiction in the Output Fragment

There exists a contradiction of the output fragment between the attributes and the abstracts of the movies. Some of the output provides one or two actors only. However, referring to the abstract, there are more actors compared to the given output. For example, the film 'Raining Stones (http://dbpedia.org/resource/Raining_Stones)' produced starring results with only one actor, 'Bruce Jonas', while in abstract details, three actors are displayed: 'Bruce Jonas', 'Ricky Tomlinson', and 'Tom Hickey'. Another case in point is the film 'The Abominable Snowman ([http://dbpedia.org/resource/The_Abominable_Snowman_\(film\)](http://dbpedia.org/resource/The_Abominable_Snowman_(film)))'.

5.3. Mappings of Different/Wrong Movies

Incorrect mappings were also found in the mapped dataset. For example, the mapping set for the movie_id of 97, 'Hate (Haine, La) (1995)' is mapped to a different movie from ML1M. It is mapped to the URI of ([http://dbpedia.org/resource/Whatever_Will_Be,_Will_Be_\(1995_film\)](http://dbpedia.org/resource/Whatever_Will_Be,_Will_Be_(1995_film))),

that refers to a 1995 Hong Kong drama film directed by 'Jacob Cheung', which starred 'Aaron Kwok' and 'Kelly Chen'. In contrast, 'Mathieu Kassovitz' wrote, co-edited, and directed 'Hate (Haine, La) (1995)', a French black and white drama film. These two films differ considerably based on the dissimilar given abstracts.

5.4. Actors with Different Names

Another challenge is that some actors have different names in some films, even if it refers to the identical actor. For example, some movies display the data of 'Jada Pinkett' for the starring information, and some others showed 'Jada Pinkett-Smith'. Other actors who have similar situations are 'Joanne Whalley/Joanne Whalley-Kilmer' and 'Tisha Campbell/Tisha Campbell-Martin'.

5.5. Brotherhood Film Direction

Some movies use the word 'Brothers' for directors to indicate that the movies are being directed by two or more siblings. 'Hughes Brothers (Albert and Allen Hughes)', 'Farrelly Brother (Peter and Bobby Farrelly)', and 'Coen Brothers (Joel and Ethan Coen)' are some of the example of movies that use the term 'Brothers' for their directed films. The issue is imminent if it even happened when only one of the brothers directs another film and applies the individual name. For example, 'Dumb and Dumber (1994)' is directed by 'Peter Farrelly' from the 'Farrelly Brothers'. In contrast, 'Bobby Farrelly' is not involved in the directing. The term 'Brothers' will be used if both of them direct the same movie. This issue will be affected by the number of the films being directed by a particular director. As a solution, the co-directed films are divided such that the same movies are linked to both individual directors.

5.6. Confusing Output

Confusion exists when the mappings between entities do not make sense. For example, the 'movie_id = 1547' which is mapped to the film 'Shiloh ([http://dbpedia.org/resource/Shiloh_\(film\)](http://dbpedia.org/resource/Shiloh_(film)))' exhibits a confusing director. The director and abstract information's data are dissimilar: 'Chip Rosenblum' and 'Dale Rosenblum'. The director's output shows 'Chip Rosenblum', whereas the abstract information reveals 'Dale Rosenblum'. This scenario requires additional searching information to get the actual data and the result shows 'Dale "Chip" Rosenblum' as the commercial name. For that issue, we take the 'Dale Rosenblum' as the director.

Another issue is where the actor list's output does not

properly reflect the film’s main cast. It can be clearly associated as some famous actor recognized for portraying a particular character in some films, especially those with sequels. In order to achieve more tangible assurance we also should adhere to the abstract. For instance, the movie ‘Mission Impossible II (http://dbpedia.org/resource/Mission:_Impossible_II)’ has four main casts, but does not list the actor ‘Tom Cruise’ in the starring output. It is advantageous to attempt retrieving the films’ full four main casts, since it will increase the possibility of similarity of a particular item with known actors.

There are also cases where the output displayed for specific attributes does not coincide with the abstract information. An example is ‘The Eye of Tammy Faye’ (http://dbpedia.org/resource/The_Eyes_of_Tammy_Faye).’ According to the abstract detail, there is a film director named ‘Fenton Bailey’. The results for the director attribute, on the other hand, show two directors, ‘Randy Barbato’ and ‘Fenton Bailey’. In situations like this, we prioritise the output attribute display.

Other than that, actors with very similar names may also be disconcerting. For cases of nearly identical name ambiguity with different people, data number errors may occur. An example is a name between ‘Elaine Stritch’ and ‘Elaine Strich’. Other examples are ‘Tom Skerritt’ and ‘Tom Skeritt’, and ‘Joseph Cotton’ and ‘Joseph Cotten’.

5.7. Special Characters in URI

Some URIs are accompanied by special character symbols representing a text, which could justify why output does not display. Examples can be seen from the URIs of:

(http://dbpedia.org/resource/Who’s_Harry_Crumb%3F)
and
(http://dbpedia.org/resource/Cleo%3%90_from_5_to_7).

The former refers to the film ‘Who’s Harry Crumb?’ from the ‘movie_id=3387’ in ML1M, while the origin film’s name for the latter URI is ‘Cleo From 5 to 7’. One of the reasons is that the film’s name is not in English or uses non-English symbols and characters. This mapping problem prevents data from being produced. To obtain the

actual movie title, we would then have to refer back to the ML1M data set through the mapping set’s movie id.

6. DISCUSSION

All of the above-mentioned challenges may impact the data analysis relating to the numbers of directors and actors, and thus will directly affect the process of determining the uniformity and consistency of the data. Therefore, pre-processing involving cleaning and filtering to remove unnecessary data as necessary.

We should also note that the cardinality of attributes may vary. For example, a film may have only one director but ten actors. It is expected that the ‘*dbo:starring*’ attribute values will dominate the encoded vector of a movie as compared to the ‘*dbo:director*.’ However, based on the extracting process some films also provide more than one director as the output, especially for films belonging to the animation genre. As such, we retain the minimum one value when data exists and increase it to a maximum of 4 using the symbol ‘|’ as its separator. The data can be referred to Table 2.

As a result of the aforementioned challenges, a clean additional information of the movie data as illustrated in Table 2 is being provided, namely ML1M-DBpedia. It thus contributes in resolving some of the issues discussed previously. Eventually, the data can then be exploited to overcome the issues of sparsity in RS. With sparse ratings, it is unlikely that two users or objects would have similar ratings. As a result, scores are predicted for a limited number of neighbours. Since sparsity has an effect on identifying item similarities, it has an impact on achieving higher predictive accuracy and recommendation relevance.

For example, the data with ratings given by users in the ML1M can then be analyzed and manipulated to enrich it. For example, the data relating to the relation of users and items for the two films as can be seen in Fig. 8, if further analysed, will result in the recommendation of the film ‘Braveheart’ since the user likes the film ‘The Man without a Face,’ which was directed by the same director. Meanwhile, Table 3 shows the data for the item related to ML1M and DBpedia based on the ‘*dbo:director*’ attri-

Table 2. Additional data for ‘*dbo:director*’ and ‘*dbo:starring*’

| movie_id | dbo:director | dbo:starring |
|----------|---------------|--|
| 1 | John Lasseter | Tom Hanks Tim Allen |
| 2 | Joe Johnston | Robin Williams Kirsten Dunst Bradley Pierce David Alan Grier |
| 3 | Howard Deutch | Jack Lemmon Walter Matthau Ann-Margret Sophia Loren |

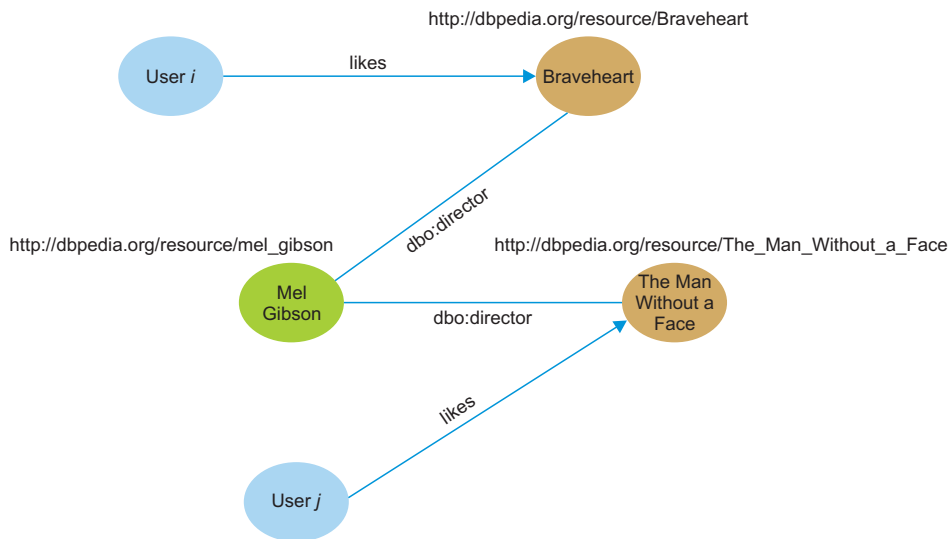


Fig. 8. Data relation of user and item based on 'dbo:director' attribute.

Table 3. Data of users and ratings based on the relation for the 'dbo:director' attributes

| DBpedia | Film | ML1M | |
|--|--|--------------|----------------|
| | | User (U) | Rating (R) |
| dbo: director {mel_gibson} | Braveheart (movie_id:110) (dbpedia.org/resource/Braveheart) | 2 | 5 |
| | | 7 | 5 |
| | | 8 | 5 |
| | | 10 | 4 |
| | | 11 | 3 |
| The Man Without A Face (movie_id:491) (dbpedia.org/resource/The_Man_Without_a_Face) | | 26 | 3 |
| | | 123 | 3 |
| | | 133 | 4 |
| | | 148 | 3 |
| | | 175 | 4 |

Each film display for the five initial ratings only.
ML1M, MovieLens 1 Million.

bute to generate a recommendation. It shows the results of additional information relating to 'dbo:director' and 'dbo:starring' extracted from DBpedia and mapped to the *movie_ID* in the ML1M dataset. Given that the recommended items are linked to the LOD dataset, its information may be expanded to ascertain the items considered similar to those used in the past. This rich information can then be exploited to tackle the sparsity issue, by providing new predicted ratings based on the particular attribute. It would collaboratively be helpful for the system to provide more relevant and high-quality recommendations as data sparsity is decreased.

7. CONCLUSION

The usage of LOD data poses new challenges and issues in developing the next-generation RS and, more generally, complex web applications. Several studies on the LOD in RS have been steadily arising over the last few years in different aspects.

This paper has presented collaborative RS that leverage the knowledge encoded in the LOD resources. In particular, since the focus of this study was on the movie domain, DBpedia was exploited to obtain more information about movies, such as actors and directors. Furthermore, we en-

countered and highlighted some of the issues encountered throughout the mapping and extraction of data from the ML1M dataset and the DBpedia resource. Such issues can benefit future research in this area, particularly the cleaning and organising knowledge extracted from similar resources. Thus, resolving such issues may assist data providers in discovering and correcting errors by considering all findings given so that any violations can be fixed before further exploration. Although the solutions to these issues and challenges have been proposed, further work relating to evaluation in terms of the quality and accuracy of the mapped data is necessary and thus becomes the near future work of this study.

Other future work includes integrating to other LOD data sources such as Linked MDB and Freebase. The LOD initiative sets the criteria for cross-domain interoperability and has accumulated vast volumes of knowledge in recent years. It provides various ways by which RS performance can be enhanced by enriching the user's profile with related features.

Furthermore, the refinement of the LOD's role in the recommendation process to a group of users is worth exploring. The cold start problem that arises in individual recommender systems is also being addressed in a group recommender system that collectively recommends items to a group of individuals based on their preferences (Dara et al., 2020). Therefore, it is interesting to investigate the potential of using the wealth of relations embedded within the LOD resources so as to produce a more accurate and diversified recommendation to a group of users.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the sponsorship received to carry out this study from Tun Hussein Onn University of Malaysia and from the Malaysia Ministry of Higher Education.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

- Aggarwal, C. C. (2016). *Recommender systems: The textbook*. Springer.
- Al-Barznji, K., & Atanassov, A. (2017, October 4-6). Collaborative filtering techniques for generating recommendations on big data. In V. Sgurev, K. Boyanov, M. Hadjiski, K. Stoilova, K. Boshnakov, E. Nikolov, & V. Andreev (Eds.), *Proceedings of the International Conference Automatics and Informatics' 2017* (pp. 225-228). John Atanasoff Society of Automatics and Informatics.
- Albatayneh, N. A., Ghauth, K. I., & Chua, F.-F. (2018). Utilizing learners' negative ratings in semantic content-based recommender system for e-learning forum. *Journal of Educational Technology & Society*, 21(1), 112-125. <https://www.jstor.org/stable/26273873>.
- Alshammari, M., & Nasraoui, O. (2019, September 17-19). Augmented semantic explanations for collaborative filtering recommendations. In A. Fred, & J. Filipe (Eds.), *Proceedings of the IC3K 2019: 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 83-88). SciTePress.
- Ameen, A. (2019). Knowledge based recommendation system in semantic web - a survey. *International Journal of Computer Applications*, 182(43), 20-25. <https://doi.org/10.5120/ijca2019918538>.
- Anelli, V. W., Deldjoo, Y., Di Noia, T., Di Sciascio, E., & Merra, F. A. (2020, May 31-June 4). SAShA: Semantic-aware shilling attacks on recommender systems exploiting knowledge graphs. In A. Harth, S. Kirrane, A.-C. N. Ngomo, H. Paulheim, A. Rula, A. L. Gentile, P. Haase, & M. Cochez (Eds.), *Proceedings of the 17th International Conference, ESWC 2020* (vol. 12123, pp. 307-323). Springer. https://doi.org/10.1007/978-3-030-49461-2_18
- Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A nucleus for a web of open data. In K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, & P. Cudré-Mauroux (Eds.), *Proceedings of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007* (vol. 4825, pp. 722-735). Springer.
- Bañas, D., Havrilová, C., & Paralič, J. (2015, September 3-5). Combination of user profile information and collaborative filtering in recommendations. In A. Szakal (Ed.), *Proceedings of the 2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)* (pp. 359-363). IEEE.
- Barros, M., Moitinho, A., & Couto, F. M. (2020, April 14-17). Hybrid semantic recommender system for chemical compounds. In J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, & F. Martins (Eds.), *Proceedings of the 42nd European Conference on IR Research, ECIR 2020* (vol. 12036, pp. 94-101). Springer.
- Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research-paper recommender systems: A literature survey. *Internation*

- tional Journal on Digital Libraries*, 17(4), 305-338. <https://doi.org/10.1007/s00799-015-0156-0>.
- Bostandjiev, S., O'Donovan, J., & Höllerer, T. (2012, September 9-13). TasteWeights: A visual interactive hybrid recommender system. In P. Cunningham, N. Hurley, I. Guy, & S. S. Anand (Eds.), *Proceedings of the RecSys '12: 6th ACM Conference on Recommender Systems* (pp. 35-42). Association for Computing Machinery.
- Cao, D., He, X., Miao, L., Xiao, G., Chen, H., & Xu, J. (2019). Social-enhanced attentive group recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 33(3), 1195-1209. <https://doi.org/10.1109/TKDE.2019.2936475>.
- Cappiello, C., Di Noia, T., Marcu, B. A., & Matera, M. (2016, June 6-9). A quality model for linked data exploration. In A. Bozzon, P. Cudre-Maroux, & C. Pautasso (Eds.), *Proceedings of the 16th International Conference on Web Engineering, ICWE 2016* (vol. 9671, pp. 397-404). Springer.
- Carrillat, F. A., Legoux, R., & Hadida, A. L. (2018). Debates and assumptions about motion picture performance: A meta-analysis. *Journal of the Academy of Marketing Science*, 46(2), 273-299. <https://doi.org/10.1007/s11747-017-0561-6>.
- Chu, P.-M., Mao, Y.-S., Lee, S.-J., & Hou, C.-L. (2020). Leveraging user comments for recommendation in e-commerce. *Applied Sciences*, 10(7), 2540. <https://doi.org/10.3390/app10072540>.
- Dara, S., Chowdary, C. R., & Kumar, C. (2020). A survey on group recommender systems. *Journal of Intelligent Information Systems*, 54(2), 271-295. <https://doi.org/10.1007/s10844-018-0542-3>.
- Delic, A., Neidhardt, J., Nguyen, T. N., & Ricci, F. (2018). An observational user study for group recommender systems in the tourism domain. *Information Technology & Tourism*, 19(1-4), 87-116. <https://doi.org/10.1007/s40558-018-0106-y>.
- Dell'Aglio, D., Celino, I., & Cerizza, D. (2010, November 8). Anatomy of a semantic web-enabled knowledge-based recommender system. In A. Bernstein, P. Grace, M. Klusch, & M. Paolucci (Eds.), *Proceedings of the 4th International Workshop on Service Matchmaking and Resource Retrieval in the Semantic Web (SMR2 2010)* (vol. 667, pp. 115-130). CEUR Workshop Proceedings.
- Deshmukh, A. A., Nair, P., & Rao, S. (2018, November 17-20). A scalable clustering algorithm for serendipity in recommender systems. In H. Tong, Z. Li, F. Zhu, & J. Yu (Eds.), *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 1279-1288). IEEE.
- Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M. (2012, September 5-7). Linked open data to support content-based recommender systems. In H. Sack, & L. T. Pellegrini (Eds.), *Proceedings of the I-SEMANTICS 2012: 8th International Conference on Semantic Systems* (pp. 1-8). Association for Computing Machinery.
- Di Noia, T., Ostuni, V. C., Tomeo, P., & Di Sciascio, E. (2016). SPrank: Semantic path-based ranking for top-N recommendations using linked open data. *ACM Transactions on Intelligent Systems and Technology*, 8(1), 9. <https://doi.org/10.1145/2899005>.
- Erion, Ç., & Maurizio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6), 1487-1524. <https://doi.org/10.3233/IDA-163209>.
- Feng, X., Zhang, H., Ren, Y., Shang, P., Zhu, Y., Liang, Y., Guan, R., & Xu, D. (2019). The deep learning-based recommender system "Pubmender" for choosing a biomedical publication venue: Development and validation study. *Journal of Medical Internet Research*, 21(5), e12957. <https://doi.org/10.2196/12957>.
- Ferré, S. (2017). SPARKLIS: An expressive query builder for SPARQL endpoints with guidance in natural language. *Semantic Web*, 8(3), 405-418. <https://doi.org/10.3233/SW-150208>.
- Figuroa, C., Vagliano, I., Rocha, O. R., Torchiano, M., Zucker, C. F., Corrales, J. C., & Morisio, M. (2019). Executing, comparing, and reusing linked-data-based recommendation algorithms with the allied framework. In M. Lytras, N. Aljohani, E. Damiani, & K. Chui (Eds.), *Semantic Web science and real-world applications* (pp. 18-47). IGI Global.
- Franz, T., Koch, J., Dividino, R., & Staab, S. (2010, March 22-24). LENA-TR: Browsing linked open data across knowledge-aspects. In D. Brickley, V. K. Chaudhri, H. Halpin, & D. McGuinness (Eds.), *Proceedings of the AAAI 2010 Spring symposium* (Technical Report SS-10-07, pp. 46-51). AAAI Press.
- Gmerek, N. (2015). The determinants of Polish movies' box office performance in Poland. *Journal of Marketing and Consumer Behaviour in Emerging Markets*, 1(1), 15-35. <https://doi.org/10.7172/2449-6634.jmcbem.2015.1.2>.
- Hammou, B. A., Lahcen, A. A., & Mouline, S. (2019). A distributed group recommendation system based on extreme gradient boosting and big data technologies. *Applied Intelligence*, 49(12), 4128-4149. <https://doi.org/10.1007/s10489-019-01482-9>.
- Heitmann, B., & Hayes, C. (2010, March 22-24). Using linked data to build open, collaborative recommender systems. In D. Brickley, V. K. Chaudhri, H. Halpin, & D. McGuinness (Eds.), *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium 2010: Linked Data Meets Artificial Intelligence* (pp. 76-81). AAAI

- Publications.
- Hidayat, R., Yahya, Y., Mohd Noah, S. A., Ahmad, M. Z., & Hamdan, A. R. (2012). Semantic web portal in university research community framework. *International Journal on Advanced Science, Engineering and Information Technology*, 2(6), 449-453. <https://doi.org/10.18517/ijaseit.2.6.242>.
- Hong, M., Jung, J. J., & Lee, M. (2015, November 26-27). Social affinity-based group recommender system. In P. C. Vinh, & V. Alagar (Eds.), *Proceedings of the 4th International Conference, ICCASA 2015* (vol. 165, pp. 111-121). Springer.
- Iana, A., Jung, S., Naeser, P., Birukou, A., Hertling, S., & Paulheim, H. (2019, September 9-12). Building a conference recommender system based on SciGraph and WikiCFP. In M. Acosta, P. Cudré-Mauroux, M. Maleshkova, T. Pellegrini, H. Sack, & Y. Sure-Vetter (Eds.), *Proceedings of the 15th International Conference on Semantic Systems, SEMANTiCS 2019* (vol. 11702, pp. 117-123). Springer.
- Jamil, N., Mohd Noah, S. A., & Mohd, M. (2020). Collaborative item recommendations based on friendship strength in social network. *International Journal of Machine Learning and Computing*, 10(3), 437-443. <https://doi.org/10.18178/ijmlc.2020.10.3.954>.
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: An introduction*. Cambridge University Press.
- Kim, M. H. (2013). Determinants of revenues in the motion picture industry. *Applied Economics Letters*, 20(11), 1071-1075. <https://doi.org/10.1080/13504851.2013.783678>.
- Ko, H., Kim, E., Ko, I.-Y. & Chang, D. (2014, January 15-17). Semantically-based recommendation by using semantic clusters of users' viewing history. In S.-M. Choi, & J.-C. Chen (Eds.), *Proceedings of the 2014 International Conference on Big Data and Smart Computing (BIGCOMP)* (pp. 83-87). IEEE.
- Kuchař, J. (2015, August 2-5). Augmenting a feature set of movies using linked open data. In N. Bassiliades, P. Fodor, A. Giurca, G. Gottlob, T. Kliegr, G. J. Nalepa, M. Palmirani, A. Paschke, M. Proctor, D. Roman, F. Sadri, & N. Stojanovic (Eds.), *Proceedings of the RuleML 2015 Challenge, the Special Track on Rule-based Recommender Systems for the Web of Data* (vol. 1417, pp. 1-9). CEUR Workshop Proceedings.
- Kushwaha, N., & Vyas, O. P. (2014, October 9-11). SemMovieRec: Extraction of semantic features of DBpedia for recommender system. In P. Bhattacharya, P. J. Narayanan, & S. Padmanabhuni (Eds.), *Proceedings of the COMPUTE '14: 7th ACM India Computing Conference* (pp. 1-9). Association for Computing Machinery.
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., & Bizer, C. (2015). DBpedia – a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2), 167-195. <https://doi.org/10.3233/SW-140134>.
- Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation. *User Modeling and User-Adapted Interaction*, 29(2), 239-249. <https://doi.org/10.1007/s11257-019-09231-w>.
- Maillo, J., Ramírez, S., Triguero, I., & Herrera, F. (2017). kNN-IS: An iterative spark-based design of the k-nearest neighbors classifier for big data. *Knowledge-Based Systems*, 117, 3-15. <https://doi.org/10.1016/j.knosys.2016.06.012>.
- Meymandpour, R., & Davis, J.G. (2015, January 27-30). Enhancing recommender systems using linked open data-based semantic analysis of items. In J. G. Davis, & A. Bozzon (Eds.), *Proceedings of the 3rd Australasian Web Conference (AWC 2015)* (vol. 166, pp. 11-17). Australian Computer Society.
- Middleton, S. E., Shadbolt, N. R., & De Roure, D. C. (2004). Ontological user profiling in recommender systems. *ACM Transactions on Information Systems*, 22(1), 54-88. <https://doi.org/10.1145/963770.963773>.
- Mirizzi, R., Di Noia, T., Ragone, A., Ostuni, V. C., & Di Sciascio, E. (2012). Movie recommendation with DBpedia. *IIR*, 101-112. <https://dblp.org/db/conf/iir/iir2012.html#MirizziNROS12>.
- Mohamed, M. H., Khafagy, M. H., Elbeh, H., & Abdalla, A. M. (2019). Sparsity and cold start recommendation system challenges solved by hybrid feedback. *International Journal of Engineering Research and Technology*, 12(12), 2735-2742. https://www.ripublication.com/irph/ijert19/ijertv12n12_87.pdf.
- Musto, C., Basile, P., Lops, P., de Gemmis, M., & Semeraro, G. (2014, October 6). Linked open data-enabled strategies for top-N recommendations. In T. Bogers, M. Koolen, & I. Cantador (Eds.), *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014* (vol. 1245, pp. 49-55). CEUR Workshop Proceedings.
- Musto, C., Basile, P., Lops, P., de Gemmis, M., & Semeraro, G. (2017). Introducing linked open data in graph-based recommender systems. *Information Processing & Management*, 53(2), 405-435. <https://doi.org/10.1016/j.ipm.2016.12.003>.
- Nawi, R. M., Mohd Noah, S. A., & Zakaria, L. Q. (2020). Evaluation of group modelling strategy in model-based collaborative filtering recommendation. *International Journal of Machine Learning and Computing*, 10(2), 330-338. <https://doi.org/10.18178/ijmlc.2020.10.2.1000>.

- doi.org/10.18178/ijmlc.2020.10.2.939.
- Nguyen, P. T., Tomeo, P., Di Noia, T., & Di Sciascio, E. (2015, October 11-15). Content-based recommendations via DBpedia and freebase: A case study in the music domain. In M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunaryan, K. Thirunarayan, & S. Staab (Eds.), *Proceedings of the 14th International Semantic Web Conference* (vol. 9366, pp. 605-621). Springer.
- Nilashi, M., Asadi, S., Abumalloh, R. A., Samad, S., & Ibrahim, O. (2020). Intelligent recommender systems in the COVID-19 outbreak: The case of wearable healthcare devices. *Journal of Soft Computing and Decision Support Systems*, 7(4), 8-12. https://www.jsdss.com/index.php/files/article/view/233/pdf_286.
- Nilashi, M., Ibrahim, O., & Bagherifard, K. (2018). A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. *Expert Systems with Applications*, 92, 507-520. <https://doi.org/10.1016/j.eswa.2017.09.058>.
- Nozari, R. B., & Koochi, H. (2020). A novel group recommender system based on members' influence and leader impact. *Knowledge-Based Systems*, 205, 106296. <https://doi.org/10.1016/j.knsys.2020.106296>.
- Oliveira, J., Delgado, C., & Assaife, A. C. (2017). A recommendation approach for consuming linked open data. *Expert Systems with Applications*, 72, 407-420. <https://doi.org/10.1016/j.eswa.2016.10.037>.
- Ortega, F., Hernando, A., Bobadilla, J., & Kang, J. H. (2016). Recommending items to group of users using matrix factorization based collaborative filtering. *Information Sciences*, 345, 313-324. <https://doi.org/10.1016/j.ins.2016.01.083>.
- Osman, N. A., Noah, S. A. M., & Darwich, M. (2020). Contextual sentiment based recommender system to provide recommendation in the electronic products domain. *International Journal of Machine Learning and Computing*, 9(4), 425-431. <https://doi.org/10.18178/ijmlc.2019.9.4.821>.
- Ostuni, V. C., Di Noia, T., Di Sciascio, E., & Mirizzi, R. (2013, October 12-16). Top-N recommendations from implicit feedback leveraging linked open data. In Q. Yang, I. King, Q. Li, & P. Pu (Eds.), *Proceedings of the RecSys '13: 7th ACM conference on Recommender systems* (pp. 85-92). Association for Computing Machinery.
- Palumbo, E., Rizzo, G., & Troncy, R. (2017, August 27-31). entity2rec: Learning user-item relatedness from knowledge graphs for top-N item recommendation. In P. Cremonesi, F. Ricci, S. Berkovsky, & A. Tuzhilin (Eds.), *Proceedings of the RecSys '17: 11th ACM Conference on Recommender Systems* (pp. 32-36). Association for Computing Machinery.
- Peralta, V. (2007). Extraction and integration of MovieLens and IMDb data, Technical Report. *Accès Personnalisé à des Masses de Données*. https://www.researchgate.net/publication/228429288_Extraction_and_Integration_of_MovieLens_and_IMDb_Data.
- Pereira, C. K., Campos, F., Ströele, V., David, J. M. N., & Braga, R. (2018). BROAD-RSI – educational recommender system using social networks interactions and linked data. *Journal of Internet Services and Applications*, 9, 7. <https://doi.org/10.1186/s13174-018-0076-5>.
- Petrova, E., Pauwels, P., Svidt, K., & Jensen, R. L. (2019, July 10-12). Semantic data mining and linked data for a recommender system in the AEC industry. In J. O'Donnell, A. Chassiakos, D. Rovas, & D. Hall (Eds.), *Proceedings of the 2019 European Conference on Computing in Construction* (pp. 172-181). European Council on Computing in Construction.
- Piao, G., & Breslin, J. G. (2016, April 4-8). Measuring semantic distance for linked open data-enabled recommender systems. In S. Ossowski (Ed.), *Proceedings of the SAC '16: 31st Annual ACM Symposium on Applied Computing* (pp. 315-320). Association for Computing Machinery.
- Piao, G., & Breslin, J. G. (2018, June 3-7). Transfer learning for item recommendations and knowledge graph completion in item related domains via a co-factorization model. In A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, & M. Alam (Eds.), *Proceedings of the 15th International Conference, ESWC 2018* (vol. 10843, pp. 496-511). Springer.
- Rawal, M., & Saavedra, J. L. (2017). Mediating effect of word-of-mouth in movie theatre industry. *Journal of Media and Communication Studies*, 9(3), 17-23. <https://doi.org/10.5897/JMCS2016.0529>.
- Raynaud, T., Subercaze, J., & Laforest, F. (2018, October 22-26). Fouilla: Navigating DBpedia by topic. In A. Cuzzocrea, J. Allan, N. Paton, D. Srivastava, R. Agrawal, A. Broder, M. Zaki, S. Candan, A. Labrinidis, A. Schuster, & H. Wang (Eds.), *Proceedings of the CIKM '18: 27th ACM International Conference on Information and Knowledge Management* (pp. 1907-1910). Association for Computing Machinery.
- Rowe, M. (2014, August 11-14). SemanticSVD++: Incorporating semantic taste evolution for predicting ratings. In D. Ślęzak, B. Dunin-Łępicz, M. Lewis, & T. Terano (Eds.), *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (pp. 213-220). IEEE.
- Roy, A., Banerjee, S., Bhatt, C., Badr, Y., & Mallik, S. (2018). Hybrid group recommendation using modified termite colony

- algorithm: A context towards big data. *Journal of Information & Knowledge Management*, 17(2), 1850019. <https://doi.org/10.1142/S0219649218500193>.
- Saat, N. I. Y., Noah, S. A. M., & Mohd, M. (2018). Towards serendipity for content-based recommender systems. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4-2), 1762-1769. <https://doi.org/10.18517/ijaseit.8.4-2.6807>.
- Sankar, A., Zhang, X., Krishnan, A., & Han, J. (2020, February 3-7). Inf-VAE: A variational autoencoder framework to integrate homophily and influence in diffusion prediction. In J. Caverlee, X. B. Hu, M. Lalmas, & W. Wang (Eds.), *Proceedings of the WSDM '20: 13th International Conference on Web Search and Data Mining* (pp. 510-518). Association for Computing Machinery.
- Sansonetti, G., Gasparetti, F., Micarelli, A., Cena, F., & Gena, C. (2019). Enhancing cultural recommendations through social and linked open data. *User Modeling and User-Adapted Interaction*, 29(1), 121-159. <https://doi.org/10.1007/s11257-019-09225-8>.
- Srinivasan, U. S., & Mani, C. (2018). Diversity-ensured semantic movie recommendation by applying linked open data. *International Journal of Intelligent Engineering & Systems*, 11(2), 275-286. <https://doi.org/10.22266/ijies2018.0430.30>.
- Stratigi, M., Kondylakis, H., & Stefanidis, K. (2020). Multidimensional group recommendations in the health domain. *Algorithms*, 13(3), 54. <https://doi.org/10.3390/a13030054>.
- Tomeo, P., Di Noia, T., de Gemmis, M., Lops, P., Semeraro, G., & Di Sciascio, E. (2015, September 16-20). Exploiting regression trees as user models for intent-aware multi-attribute diversity. In T. Bogers, & M. Koolen (Eds.), *Proceedings of the 2nd Workshop on New Trends on Content-Based Recommender Systems co-located with 9th ACM Conference on Recommender Systems (RecSys 2015)* (vol. 1448, pp. 1-8). CEUR Workshop Proceedings.
- Vagliano, I., Monti, D., Scherp, A., & Morisio, M. (2017, December 4-6). Content recommendation through semantic annotation of user reviews and linked data - an extended technical report. In Association for Computing Machinery (Ed.), *Proceedings of the 9th International Conference on Knowledge Capture* (pp. 1-4). Association for Computing Machinery.
- Vall, A., Dorfer, M., Eghbal-zadeh, H., Schedl, M., Burjorjee, K., & Widmer, G. (2019). Feature-combination hybrid recommender systems for automated music playlist continuation. *User Modeling and User-Adapted Interaction*, 29(2), 527-572. <https://doi.org/10.1007/s11257-018-9215-8>.
- Vilakone, P., Park, D.-S., Xinchang, K., & Hao, F. (2018). An efficient movie recommendation algorithm based on improved k -clique. *Human-centric Computing and Information Sciences*, 8(1), 38. <https://doi.org/10.1186/s13673-018-0161-6>.
- Wang, H., Chen, D., & Zhang, J. (2020). Group recommendation based on hybrid trust metric. *Automatika*, 61(4), 694-703. <https://doi.org/10.1080/00051144.2020.1715590>.
- Wenige, L., & Ruhland, J. (2018). Retrieval by recommendation: Using LOD technologies to improve digital library search. *International Journal on Digital Libraries*, 19(2-3), 253-269. <https://doi.org/10.1007/s00799-017-0224-8>.
- Xu, J., Yan, Z., Cao, G., & Zhao, J. (2018, August 17-19). Family shopping recommendation system using behavior sequence data and user profile. In ICIMCS (Ed.), *Proceedings of the ICIMCS '18: 10th International Conference on Internet Multimedia Computing and Service* (article no.: 41). Association for Computing Machinery.
- Yadav, U., Duhan, N., & Bhatia, K. K. (2020). Dealing with pure new user cold-start problem in recommendation system based on linked open data and social network features. *Mobile Information Systems*, 2020, 8912065. <https://doi.org/10.1155/2020/8912065>.
- Yang, R., Hu, W., & Qu, Y. (2013). Using semantic technology to improve recommender systems based on slope one. In J. Li, G. Qi, D. Zhao, W. Nejdl, & H.-T. Zheng (Eds.), *Semantic web and web science* (pp. 11-23). Springer.
- Yochum, P., Chang, L., Gu, T., & Zhu, M. (2020). Linked open data in location-based recommendation system on tourism domain: A survey. *IEEE Access*, 8, 16409-16439. <https://doi.org/10.1109/ACCESS.2020.2967120>.
- Zarzour, H., Maazouzi, F., Soltani, M., & Chemam, C. (2018, May 8-10). An improved collaborative filtering recommendation algorithm for big data. In A. Amine, M. Mouhoub, O. A. Mohamed, & B. Djebbar (Eds.), *Proceedings of the 6th IFIP TC 5 International Conference, CIIA 2018* (vol. 522, pp. 660-668). Springer.
- Zaveri, A., Maurino, A., & Equille, L.-B. (2014). Web data quality: Current state and new challenges. *International Journal on Semantic Web and Information Systems*, 10(2), 1-6. <http://doi.org/10.4018/ijswis.2014040101>.
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1-38. <https://doi.org/10.1145/3285029>.
- Zhao, W. X., He, G., Yang, K., Dou, H., Huang, J., Ouyang, S., & Wen, J.-R. (2019). KB4Rec: A data set for linking knowledge bases with recommender systems. *Data Intelligence*, 1(2), 121-136. https://doi.org/10.1162/dint_a_00008.