Journal of Informatics and Web Engineering

Vol. 4 No. 3 (October 2024) eISSN: 2821-370X

Research Article Recommender Systems: A Comprehensive Review of Models, Approaches and Evaluation Metrics

Sir-Yuean Lim[1], Noramiza Hashim[2*] and Lanh Le Thanh[3]

[1,2]Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, Cyberjaya, Malaysia

[3]Faculty of Information Technology, Dong Nai Technology University, Bien Hoa City, Vietnam

**corresponding author: (noramiza.hashim@mmu.edu.my; ORCiD: 0000-0001-9838-2892)*

Abstract - With the advent of the current digital era, individuals across the developed world are commonly equipped with devices that can access vast amounts of information at their fingertips. What was considered an impossible feat was realized through remarkable technological advancements. This positive transformation has had a profound impact on education, where traditional knowledge management, such as libraries, are no longer a primary determinant of a student's academic success. Instead, it has been replaced by the internet as a medium for learning, practicing, and topic exploration. However, the sheer volume of the ever-increasing information available online can easily overwhelm a user, particularly when conducting detailed research on a specific topic. Therefore, the need for a reliable research article recommender system cannot be understated, helping students and researchers to navigate the expansive knowledge space better and achieve their learning and research objectives. This review paper aims to study the most common types of recommendation system techniques in research articles recommender systems (RS). A total of ten related works and relevant evaluation metrics written by other researchers will be studied and accessed rigorously using comparative analysis, granting further insights into the current work similar or related to the domain of this paper. Finally, this paper will identify and elaborate their current trends and gaps in the discussion section.

Keywords— Semantic, Research Article, Recommender System, Ontology, Graph, Research Article Recommender System.

Received: 25 February 2025; Accepted: 25 May 2025; Published: 16 October 2025

1. INTRODUCTION

Today, technology is widespread across world civilizations. Due to recent astounding technological advancements and achievements, society is becoming ever more dependent on the benefits technology offers. For example, modern individuals probably need two or more electronic devices, such as laptops or smartphones, to navigate everyday challenges. These include communication, online payments, digital identification, commodity purchases, education, and entertainment. The data is exponentially increasing as technologies continue to compile these available services onto the web, causing difficulties in discovering relevant information or services online without spending an exorbitant

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2025.4.3.10

Published by MMU Press. URL: journals.mmupress.com/jiwe

amount of time to find and evaluate these searches [1]. This scenario has evolved into an issue whereby users must endure longer periods to achieve their tasks.

This problem is prevalent in the realm of academia. This long-standing issue has been permeating throughout the history of academia and the evolution of knowledge management from big university libraries to the ever-expanding space of the internet. As university students shoulders countless demanding classes, meeting the high expectations of subjects becomes difficult. Students' performance may be negatively impacted due to the exhaustive and tedious information discovery process, leading to slower academic progress and ultimately, stagnation [2].

Considering this, decision-making support technologies such as RS were invented and subsequently utilized across the board. Leveraging advanced algorithms to help users find relevant information in the least amount of time despite its popularity among the public. Recommendation systems such as e-commerce, streaming services, and entertainment are especially prevalent in the industry. Directing users to their realm of interests to maximize user engagement, retention, and satisfaction.

While RS have been widely deployed in numerous sectors, their use within scholarly research itself has yet to be fully explored, less equipped to help students and researchers find suitable articles quickly through efficient search strategies. The available research article RS rely on citations, content-based approaches, or collaborative filtering (CF). Problems associated with them like information overload, generic nature of the recommendations, and the fast-changing nature of academic literature. Most importantly, most systems fail to adapt to changes in users' research interests; hence, they provide poorly tuned recommendations that fail to foster adequate scholarly exploration.

This paper attempts to provide a comprehensive literature review for the most recent works on research article RS by other researchers. Providing insights into current and future trends of research article recommendation systems such as common data preprocessing and implementation design methods. These insights could pave the way for future researchers to focus on impactful contemporary methods when designing their own recommendation systems and better understand the desired improvements by the research community within the domain rather than reinventing the wheel. Although numerous existing review journals within this domain are readily available, their relevance is often overshadowed by the rapid progress of contemporary research in this in-demand field. Therefore, these journals frequently rely on foundation RS concepts, lacking engagement with the latest advancements and discoveries, resulting in recurrent publications with minimal progressive impact within the domain.

The remainder of the paper is organized as follows. Section 2 provides a comprehensive overview of RS and its overall processes, describing the phases in RS and its ever-evolving techniques thus far in tackling various emerging challenges in meeting use cases. Section 3 is a study on a combination of related works written by other researchers in similar domains. Section 4 is the evaluation metrics used in various RS in gauging the performance of their models. Section 5 is the discussion of current and future trends of recommendation system. Finally, section 6 is the conclusion and suggestion for future research areas within the RS domain.

2. RECOMMENDER SYSTEM AND THE TYPICAL PROCESSES

2.1 Overview of Recommender Systems

Before the emergence of the World Wide Web (WWW) and affordable advanced technologies, namely personal computers, smartphones, and Internet of Things (IoT) devices, information was primarily organized using traditional methods such as record books, files, floppy disks, and CDs. While traditional methods of managing information served their purpose, they made it difficult and time-consuming to locate specific information, presenting a significant challenge for researchers and readers alike. With the advent of digital technologies, managing and storing vast amounts of information has become significantly more efficient. However, quickly finding relevant information remains a challenge, especially as data volumes grow. RS emerged as a solution to this long-standing challenge. These systems use algorithms to suggest items—such as books, movies, products, or content—tailored to users based on their preferences and behaviours [3]. The primary objective of RS is to help users shift through the ever-growing volume of information online while enhancing user experience by delivering personalized content, thus increasing engagement and satisfaction [4].

Prior to the worldwide adoption of advanced technologies, there were documented efforts to achieve personalized recommendations. In 1979, Elaine Rich, then a computer librarian and later a renowned computer scientist, sought to develop a method for suggesting books based on categorizing user preference through interviews [5]. This development marked a turning point, inspiring other researchers to expand upon the concept. According to [3], Jussi Karlgren, a Swedish computational linguist, conceptualized the "digital bookshelf" in 1990—a prototype of a recommendation system, which researchers at SICS, MIT, and Bellcore later expanded upon. Despite numerous contributions from researchers over time, foundational theories for RS were established by scholars such as Gediminas Adomavicius, Jonathan L. Herlocker, and Joeran Beel [3].

As RS continues to evolve and expand over time, so does the range of experimental and well-engineered methods develop in academia and industry. Academia focuses on advancing theories, methods, and algorithms for RS, while industry prioritizes practical applications, scalability, and driving business impact. Therefore, the challenges encountered in these sectors are distinctively different. Machine Learning (ML) techniques have been introduced to enhance RS algorithms, including clustering methods, neural networks, decision trees, support vector machines, Naïve Bayes classifiers, and performance evaluation metrics. In recent years, RS has increasingly integrated with state-of-the-art (SOTA) methods, such as deep learning, GNNs, and Large Language Models (LLMs). However, it is essential to note that new technologies are built upon the foundation laid by their predecessors that have adapted to evolving user demands and environments.

Today, RS are broadly categorized into three primary groups: Content-based filtering (CB), CF, and Hybrid-based filtering [6]. Figure 1 is a structured overview of the main types of RS. Figure 1 is a high-level overview of the RS ecosystem. Still, researchers have further investigated new ways such as more advanced deep learning models, factorization machines (FM), more efficient graph-based methods, association rule mining, and reinforcement learning (RL). These are testaments to the substantial investment in RS, which has been made to improve the user experience in various legs of the market such as e-commerce, finance, healthcare, content streaming, and education.

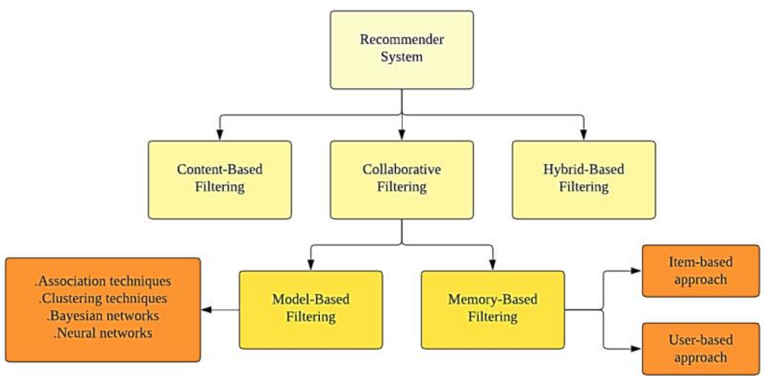

Figure 1. Recommender Systems Overview

This report is concerned with RS in education, focusing on research article recommender systems (RARS). With the increasing diversity and volume of data coming from the internet, new researchers have found themselves overwhelmed by information that is readily available to them, unable to determine what information they need and, even worse, cannot find it [7]. According to [8], approximately 64 million academic papers have been published since 1996, with over 5.14 million academic articles published in 2022. It is extremely difficult for contemporary researchers to filter through huge amounts of information when they do not want to miss any crucial details [9]. Recommendation systems created to solve this challenge have been created on popular platforms such as Google Scholar, Microsoft Academic Search, and ResearchGate. Unfortunately, these systems often fail to properly produce results, as keyword-based searches are unable to pick up the underlying relationships between articles, which causes users' confusion and inefficiency [10].

According to the Malaysian Education Blueprint 2015-2025, Malaysia is projected to have an increase in the number of students in private higher learning institutions, growing from 455,000 in 2012 to 867,000 by 2025 [11]. In 2024, the number of tertiary education students enrolment has reached over 1.3 million and average annual graduate output of 294,698 [12]. Concurrently, we have witnessed an increase in the scale of the academic population and concomitantly accrued an increase in the production of research articles, academic textbooks, and reports in digital libraries. This emphasizes the upward trend for robust, dependable RS ahead of progressing university staff, lecturers, and students' needs.

According to [6], the functionality of RS typically involves two primary elements: users and items. Different methods can let users rate items (preference values): implicit or explicit. In general, it distinguishes implicit ratings (inferred indirectly from user interactions with items) from explicit ratings (directly provided by users under finite scale or labelled intervals). Based on [13]'s extensive research in 2022, RARS are predominantly defined into four main categories: graph-based methods, hybrid-based systems, CF, and CB.

CB method generates recommendations by analysing the historical interactions of a user with research papers and their features, such as keywords or abstracts [14]. This method assumes that users should prefer similar papers if they like a given paper. CB needs a user profile to achieve recommendations and could not propose new or diverse content [7]. On the other hand, CF method uses the preferences of a user's profile and the paper's properties, comparing it with other users' preferences to find similarities [14]. CF predicts the relevance of research papers through patterns of user behaviour interest. However, this method often suffers from its 'cold start' problem, which requires strong user-profiles and enough data to be able to make accurate recommendations [7]. Hybrid-based systems achieve this as a combination of CB and CF to mitigate the drawbacks, as they utilize both features at once and improve the recommendation accuracy while tending to the drawbacks of using either method during runtime [15]. Graph-based approaches connect papers, papers, authors, or citations to find some measure of relevance using graph algorithms. The intention is to model relationships within a graph (co-authorship or citation links) to suggest articles. Common algorithms include random walks (an estimation of the probable location or relevance of papers or authors in a graph-based of the simulated traversal), bibliographic coupling (identification of that third paper, given its co-citation, which shares focus or theme), and co-citation analysis (recommending other papers that are frequently cited together as a part of other works) [14].

Other approaches in this are not directly applicable to pure RARS. For instance, [16] proposed to combine matrix factorization with content-based filtering (CBRec) to recommend good reading habits for children by suggesting children's book. Though generally similar, it serves a different pool of domains and targets. RARS lacks emphasis and the development of more specialized methods. Slow progress in addressing the needs of academic users has been facilitated by a lack of dedicated research and authoritative solutions [13]. In addition, as the demand for more positive educational impact increases, innovations such as accommodating dynamic content and user behaviours, maintaining contextual relevance and using suitable evaluation metrics to measure the level of educational outcomes of students were made [17].

One of the key challenges in RARS is the overemphasis on accuracy as a sole measure of system performance. Equating user satisfaction with predefined recommendation accuracy levels often fails to truly reflect the complexity of user expectations and interactions [13]. Beyond precision and accuracy, incorporating diverse evaluation metrics such as maximal marginal relevance (which avoids redundancy while retrieving relevant yet varied items), popularity, serendipity, and click-through rates are essential to capturing the broader dimensions of user satisfaction and system efficacy [13]. Moreover, there were long-standing missing translations from research into practice. Prototypes exploring innovative methods are seldom implemented in real-world systems or deployed at scale within the industry [13]. This starkly contrasts RS's widespread adoption in domains such as social media, e-commerce, and video streaming services. Recommendations are key factors in driving growth for Amazon, eBay, Uber, and Lyft, while YouTube reported that 70% of its watch time is attributed to recommendations and Netflix stated that 80% of content consumption stems from personalized suggestions [18]. The inconsistency in cooperation among research groups further exacerbates this challenge, resulting in prior works not being extended for further improvements, thus hindering the advancements of the field [19].

Additionally, RARS research literature contains the problem of information scarcity, where its approaches do not offer sufficient detail to replicate its findings [13]. Critical implementation details are often excluded from most studies, further hindering progress, and slowing the development of RARS [13]. In addition, there are still many technical challenges to be overcome with RARS development. An infamous problem is the well-known "cold start" problem, where a lack of historical data allows the system to develop relevant recommendations [20]. Consequently, deep learning methods have been widely pursued as possible solutions. One study introduced the evaluation metrics totNP_EU and avgNP_EU to quantify a system's ability to overcome this challenge [21]. Not only that, RARS are also confronted with the coverage problem, which makes it difficult to evaluate the viability of infrequently rated papers [5]. Despite the connection with this problem being mostly due to CB methods, contemporary approaches often utilizing deep learning methods have likely mitigated the influence of this problem [20].

Additionally, the ever-growing set of users brings forth the issue of scalability issue [22]. Depending on the dataset's complexity and size, the strategies to address scaling are different. One such study circumvented this problem by using a clustering algorithm to reduce the number of times that otherwise needed computations would have been necessary [22]. For instance, [22] mentioned that scalability problems arise when we handle up to 25,000 publications and their citation relations. Furthermore, privacy is a problem of great magnitude in RARS, which warrants careful study. Users decline to disclose sensitive information, such as habits and preferences that could've been included in the system [22]. Google's Personalized Search indicates its use of personal data through prior agreements and represents an explicit part of some recommendation systems. However, such approaches may not face this issue for certain users. Additionally, serendipity in RARS may neglect other relevant research articles while overemphasizing popular relevant papers [23]. The balance between clearly relevant and serendipitous recommendations is crucial in maintaining user trust in the system, especially for newer researchers who may rely on the system's guidance to discover and learn research work. Finally, the issue of unified scholarly data standards poses as a challenge for most modern RARS. Digital libraries containing relevant information on the web may use varying data formats, complicating recommendation systems' training and evaluation processes [13]. Some approaches such as web crawling are exempted from this problem [24].

Other miscellaneous challenges in RARS include issues such as synonym (similar meanings among words), inconsistent user ratings (Gray sheep), and false user ratings (Black sheep) are mentioned. These issues have largely been solved without machine learning methods. Finally, cyberattacks on RARS like the Shilling attack, are a known problem encountered in CF methods [22]. Nevertheless, the influence of this issue has largely been dwindled by the massive adoption of deep learning methods, making it no longer relevant to most modern RARS approaches in this space [13].

2.1.1 Phases in Recommendation Systems

Numerous RS have been developed based on the overview of recommendations systems, ranging from experimental prototypes to widely deployed industry solutions. To better understand how these systems function, the general processes can be segmented into three distinct phases: the Data Acquisition Phase, the Learning Phase, and the Prediction/ Recommender Phase. Figure 2 provides a clear framework for understanding their operational workflow, which is commonly highlighted in in-depth literature reviews on RS [25].

The RS begins by gathering essential information from the user's characteristics, preferences, and item contents to build user profiles [26]. This data is necessary for the RS to supply specific recommendations for every distinct user profile. To achieve this, the RS primarily uses three feedback methods to receive user input: explicit feedback, implicit feedback, and hybrid feedback [25].

Explicit feedback comes explicitly from users through system-prompted actions. For example, a streaming platform might ask its users to rate a placed video while watching it, providing them with clear and exact information about their preferences [27]. Explicit feedback provides good insights, but users' participation is crucial as it can result from personal bias and reluctance to give input.

For implicit feedback, it automatically receives feedback in the background by using user behaviour such as browsing history, purchase history, the links clicked, time spent on web pages or even email interactions [27]. In contrast to explicit feedback methods, implicit methods have minimal user intervention and a continuous and less intrusive data stream. Although this data is indirect, interpreting it accurately can be difficult.

Hybrid reviews combine explicit and implicit reviews to exploit the strengths of both forms of reviews to improve recommendations. This review method is particularly popular as the most effective way to get feedback from users since combining explicit user ratings with implicit data overcomes the shortcomings of each method and thus improves both accuracy of predictions [27].

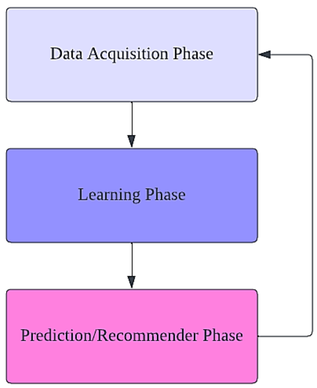

Figure 2. Framework in a Recommender System

The RS collects input data from users, taking this data and processing the collected features using machine learning models or other algorithms to analyse the characteristics of this data and predict recommendations [25], [26]. During phase one first identifies patterns, correlations and user preferences of input data. Most of these patterns are captured using various machine learning models, such as CF, content-based models, and deep learning techniques [28]. Each model is leveraging the distinctive attributes of the data in order that the recommendations are as closely aligned with user needs.

Once the RS learns patterns on the user data during the learning phase, it recommends the items with the minimum error according to the recommendation system [26]. These recommendations aim to follow user preferences as closely as possible for a better user experience. Next, the selection of explicit and implicit feedback from the recommendations offered is sent back into the system for re-evaluation [25]. This feedback loop gives RS a chance to refine its predictions and adapt to user behaviour, restarting' the recommendation process again.

This popular framework from the academic domain has been slightly adapted by the machine learning industry, which is working on recommendation systems to suit real-world use cases better. For instance, [28] made their recommender system process a way of forecasting user ratings even before explicit feedback is provided through predictive analytics. They have a version of the framework that consists of collecting, storing, analysing, and filtering to provide recommendations to users.

The data collected using explicit and implicit methods are stored in a centralized database for future use as stated by [28]. It then analyses the data to find items with similar engagement patterns among users. Finally, filtered results are recommended according to each person's preferences. This avoids having to re-collect user profile data each time for a different application, without having to reinvent that data for each domain.

During this process, evaluation metrics are used to compare the quality and performance of the RS at generating predictions. The standard predictive accuracy metrics are Mean Squared Error (MSE), Root Means Squared Error (RMSE), Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE) [25]. These metrics are designed to calculate their recommendations' accuracy based on deviating or differing between our predicted and actual values.

Another metric used to evaluate the RS is classification accuracy metrics, which assess the RS's ability to correctly match users with relevant items, which are precision, recall, F1-score measurement, and receiver operating (ROC) curves [25]. These are important to define how accurately the RS understands when and what will be generated to fulfil a correct user's needs.

Furthermore, ranking accuracy metrics evaluate the closeness in the predicted item order by RS and user preference order from the same set of items [25]. These metrics are based on how close the item ranking is (proximity), and metrics based on the correctness of item ranking and user's expectations.

2.2 RS Techniques

Content-based is a widely utilized recommendation technique based on two primary components: item descriptions and user preference profiles. They operate on associative keywords pulled from the user's history on the platform that captures what the user prefers and is interested in [25]. These keywords are leveraged within the system to predict and suggest future items that match the user's interests. For example, a user frequently watching superhero movies such as Iron Man would likely receive recommendations for other superhero genre movies or terms associated with Tony Stark [28].

The Term Frequency-Inverse Document Frequency (TF-IDF) is a well-known natural language processing (NLP) method widely utilized in CB. It is used to model the significance of the terms being within a document against the corpus. This technique, also known as the vector space model, translates item descriptions into weighted feature vectors [25].

Weights are calculated using mean values of the vectors for items rated by the user. Once the weights are calculated, they are evaluated to retrieve the likelihood of the item preferences by the user [25]. After evaluation, the likelihood of the item preferences is added to the list of top-ranked items [29]. Most domains implement this approach to recommend travelling activities, e-commerce, and video streaming platforms. However, this method is particularly popular with domains relating to textual information such as websites, news and articles [29]. Figure 3 is the general framework of a CB system.

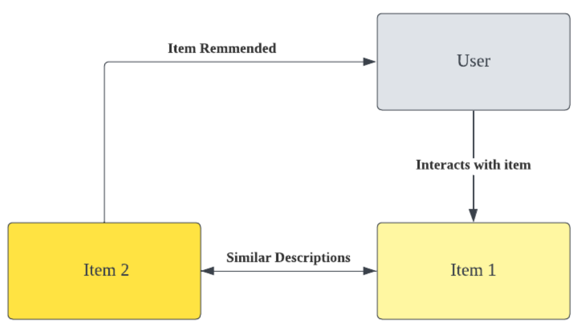

Figure 3. Framework of Content-based Filtering System

The advantages of this method include a high-level of personalized recommendations, a scalable number of users, peculiar interest, and high security from malicious item creation to boost specific and biased recommendations [3]. However, the disadvantages arises with the high cost of computation, as every item are examined to generate recommendations, which are often error-prone and time-consuming [28]. Static characteristics of items could lead to lack of serendipity whereby different users have similar profiles while only receiving overspecialized recommendations from their past interests [30].

CF methods also utilize item descriptions and user preference profiles but address the limitations of CB by adopting a distinct approach. The method uses a different technique to address the disadvantages of CB by gathering and analysing data on user's behaviour, such as their online activities. This method compares user profiles or item descriptions to identify user relationships and interdependencies between items, hence the term "collaborative" [31]. The prevailing idea is that users' preferences in the past have influence in the future.

The first CF framework was developed by researchers in Grouplens Research Institute [32]. Today, memory-based CF contains two primary algorithms. user-based filtering predicts a user's preferences by finding similar users, and item-based filtering predicts an item preference by finding similarities between items [28]. However, model-based CF attempts to predict user ratings for unlabelled products through models like Bayesian networks or Markov decision processes (MDP) [28]. The improved scalability and ability to handle large datasets with extra efficiency over the memory-based methods demonstrates the approach. Figure 4 is the framework of a CF system.

CF is advantageous for its serendipity in recommending items without necessarily having to know the item itself that deeply. Unlike direct analysis of every item description, CF identifies nuanced relationships between items and user preferences, making it a flexible solution available across a broad range of domains [28].

However, there are certain disadvantages found in CF, such as data sparsity, where the system is unable to discover the relations between the variables and inevitably leads to poor recommendations, cold start problems affecting the new users or new items which do not have enough data to interact, scalability issues which is due to the system performance decrease as there incrementally more users and items and rating biases which are possible in the uneven or variable ratings [28]. Although there are existing challenges, CF is used widely for its capability to provide diverse and personalized recommendations on different domains, such as commerce and entertainment platforms.

The difficulties associated with both CB and CF methods, hybrid-based filtering combines both advantages to make an accurate prediction and a broader recommendation [28]. User preference profiles and item ratings may be run in parallel or separately before combining the results to produce the final recommendations [25].

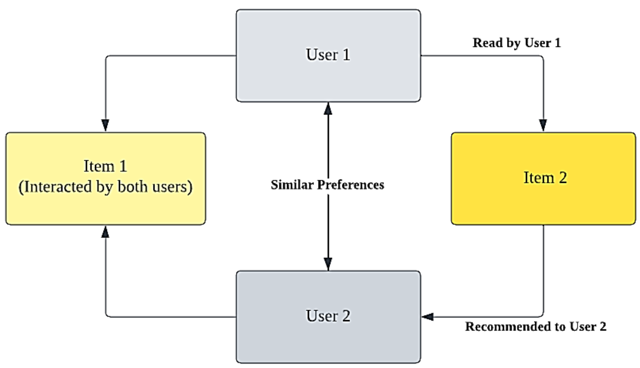

Figure 4. Framework of Collaborative Filtering System

For hybrid-based filtering methods, two key considerations are addressed. The first step is choosing the recommendation models that identify which inputs the recommendation models require and lay the foundation for the hybrid recommender system, while the other is choosing the appropriate strategy within the hybrid system [33]. A classic example of hybrid recommender system is Netflix's CF to determine what to recommend to users based on their search and view behaviour. It then switches to CB to show similar shows or movies based on the users rated highest [17].

According to [25], several hybrid-based recommender techniques are commonly used:

- **Weighted -** The weights of the various recommendation components are added numerically.
- **Switching -** Generating recommendations by selecting from various components based on the current context.
- **Mixed -** Making a recommendation by combining predictions from different recommenders.
- **Feature Combination -** Combining features from many knowledge sources to feed into a single recommendation system.
- **Feature Augmentation -** Use one recommendation technique's output as a source for another.
- **Cascade -** Priority is assigned to recommenders, with lower-priority ones breaking ties in higher-priority ones' scores.
- **Meta-level -** Generating a model from a recommendation method and using it as input for another method.

Its advantages overcome CB and CF methods' limitations with accuracy, but it also has disadvantages. These include the high implementation costs, the increased complexity, and a dependence on external datasets that may be difficult to obtain [28]. Figure 5 is a framework for Hybrid-based Filtering system.

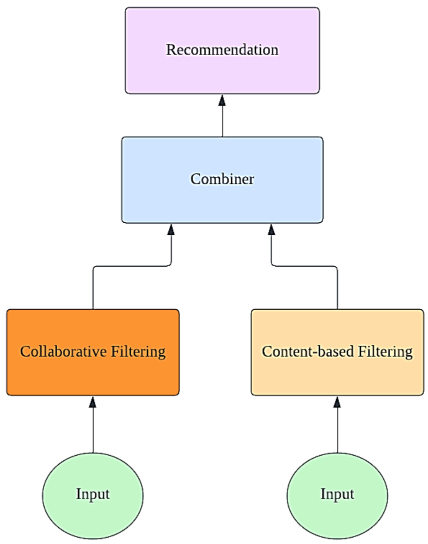

Figure 5. Framework for Hybrid-based Filtering System

From these methods, the typical approach is to take keywords and match items to a user profile by that keyword. Despite their respective advantages and disadvantages, a persistent issue remains: the possibility of a high miss rate. The miss rate is high because the items considered by the method do not have the same keywords but hold similar meanings in context, for instance, 'Romance' and 'Wedding' [34]. Mistakes such as incorrect spelling, multiple languages, and missing data worsen this issue. While the issue is not severe, a high miss rate may still miss correlations between things and people, potentially leaving important data out of your recommendations even if they are unbiased and appropriately specialized.

This complexity has led researchers and practitioners to explore and implement deep learning techniques to address it and continue the progress that CB, CF and hybrid methods have made. Deep learning technique is the semantic filtering approach that utilizes the semantic similarity of items and their relationships to 'understand' the contextual meaning of these items and their relationships, which traditional methods concentrated in the use of lexicographic similarity usually miss [34]. Deep learning brought the field to the present day of graph-based social recommendations [35]. with each augmentation of the idea building on graph algorithms.

Semantic representation and querying are the two essential components of the semantic filtering approach. The following words are embedded from a high-dimensional vector space to a lower-dimensional space as part of the semantic representation processes [34]. We call this process dimension reduction, which reduces data from high-dimensional to low-dimensional embeddings that are more efficient, easier, and faster to compute. Machine learning models for text like word2Vec, GloVe, ELMo and Transformer models learn to generate synthetic tasks to guide the learning process [34]. The common synthetic tasks for training these embeddings include classification, language modelling, and masked language modelling [34]. Figure 6 is a dimensionality reduction process and Figure 7 is an example of semantic representation.

The retrieval of these stored embeddings in vector databases based on nearest neighbour search is referred to as querying. Approximate Nearest Neighbor (ANN) is a commonly used method of real-time querying, whereby it computes item similarities with methods such as cosine similarity or Euclidean distance [34]. The retrieval process can be implemented using two common techniques: tree-based approach and the hashing-based approach. Data is recursively partitioned in a tree-based manner or can be turned into codes such that similar vectors produce the same code, resulting in hash collisions [34].

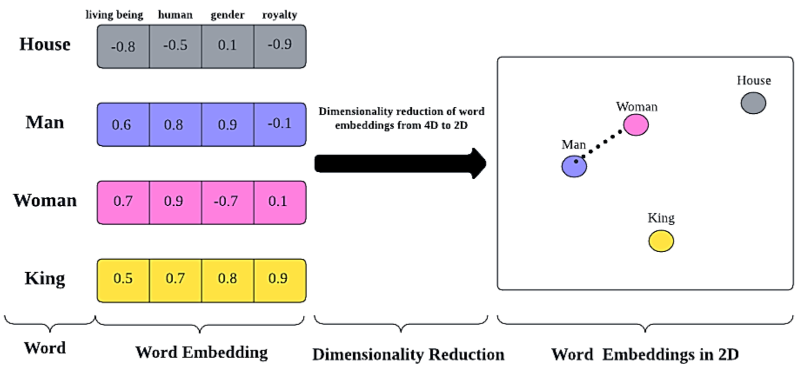

Figure 6. Dimensionality Reduction Process [36]

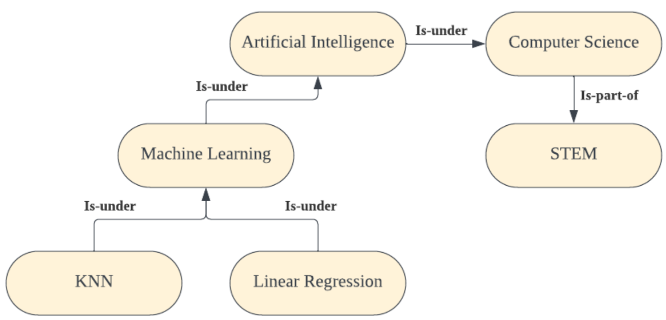

Figure 7. Semantic Representation Example

Semantic-based filtering techniques have spawned two popular approaches: ontology-based and graph-based recommendations.

Ontology-based filtering is a clear formal definition of a common conception [37]. This means that the method relies on a clear, structured model of how knowledge about a particular domain is understood and represented, which could be understood by both humans and machines. For example, a student may look for keywords such as "machine learning" for a research paper, an ontology-based filtering understands that "neural networks," "supervised learning," and "deep learning" are subfields of "machine learning" and recommends those resources as well, whereas a traditional filtering may look for that exact same keyword. Therefore, ontology-based approaches are crucial and beneficial for the advancement of semantic-based filtering.

Knowledge-based RS employ ontology to describe knowledge are known as ontology-based recommenders. Application, domain, reference, general, and (top-level) generic ontologies are the different categories of ontologies based on their domain scope [38]. Domain ontologies are general application ontologies which focus in representing a specific area of knowledge from the point of view of a particular user or developer. A domain ontology is knowledge regarding some subject or a general area. Reference ontologies take the broader, more objective approach by offering the same subject from a broader and more general view than a specific purpose would require, with a view towards the domain [38]. However, general ontologies are broad and are irrelevant with a specific domain. Both cover a wide range of general knowledge ranging from common sense reasoning (CYC), a general knowledge base and commonsense reasoning engine, to DBpedia, an open-source project on structuring Wikipedia information and making it public on the web [39]. The top-level ontologies are frameworks across many domains. Defining objects, relationships, events and processes [38].

Ontology is finding meaning in the place of diverse data [40]. It establishes a framework for learning object classification, storage, and semantic recommendation generation in genomics. A comparison with basic models suggests that ontology-based filtering can detail user profiles and item descriptions [41]. Implementing an ontology-based filtering recommender system requires three components for organizing and representing knowledge: class, relation, and properties [42]. Classes are labels defining categories representing various instances, such as "car" or "pedestrian". Relations represent the connections between descriptions that can be graphed using directed or undirected graphs. For example, such relationships include "parallel" or "overtaking" between vehicles. Finally, properties represent these attributes with a node or relation. For instance, an attribute of a car could be "colour" or "window open/closed", while an attribute of a pedestrian could be "gender" or "mask on/off". These components interact together to form a comprehensive, structured representation of knowledge in a system. Figure 8 is the framework for an ontology-based filtering system.

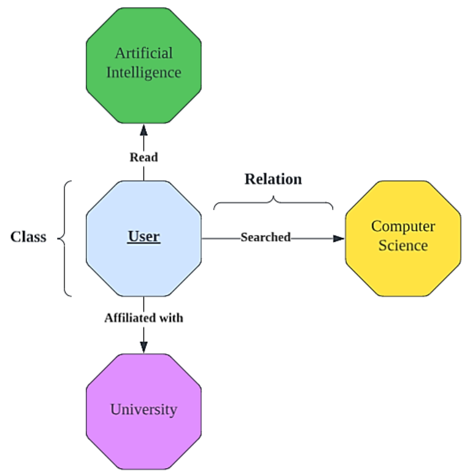

Figure 8. Framework of Ontology-based Filtering System.

Ontology-based filtering techniques offer several benefits, including addressing the cold-start issue, reducing rating sparsity, and preventing overspecialization in recommendations. These advantages stem from the method's strong reliance on domain expertise rather than user-generated data [38]. However, ontology-based filtering approach introduces its own set of challenges. These include the complex and expensive process of developing such systems [43], limited availability of datasets for specific domains and appropriate evaluation tools [38], and the need for expertise in knowledge engineering [41].

Graph-based filtering is a technique that utilizes the complex structure of user-item networks. This approach represents users and items as nodes, while their interactions, such as ratings, purchases, and social connections, are depicted as edges connecting these nodes. Semantic filtering techniques that utilize graphs that fall into three primary categories: traditional CF techniques, sophisticated graph embedding-based social recommendation systems, and graph neural network (GNN)-based social recommendation systems.

In conventional CF approaches, two distinct networks are established: one depicting user-item interactions, and another illustrating social connection among users. Similarity measures are employed to identify users most comparable to a target individual. Subsequently, items are assigned weighted ratings based on these identified similarities. By combining evaluations from users with comparable tastes, the algorithm forecasts a target user's possible interest in particular things. Items can be ranked based on such weighted predictions, which in turn can provide more personalized recommendations [34]. Figure 9 is the framework for a traditional CF method.

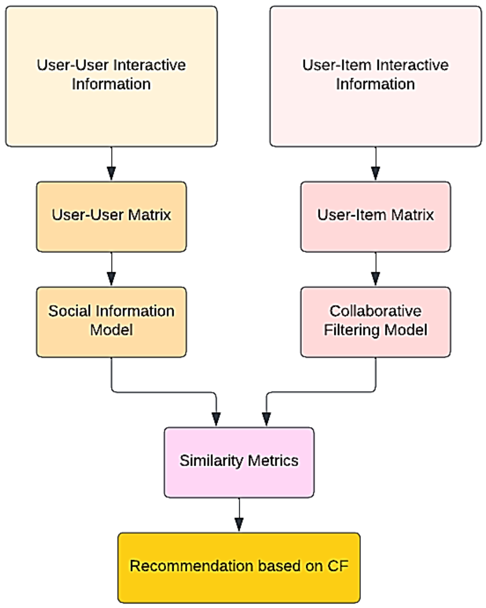

Figure 9. Framework of Traditional Collaborative Filtering Method

The deep social recommendation method depicts the relation between the entities. The Nodes are reduced with dimensionality reduction into vector embeddings by retaining the relationship between the nodes in a Euclidean space. Utilizing trust relationships that can be inferred using these embeddings, the system can produce recommendations by weighing ratings of trusted individuals who exhibit similar likes [34]. Figure 10 is an illustration of Euclidean space of entities and relationships.

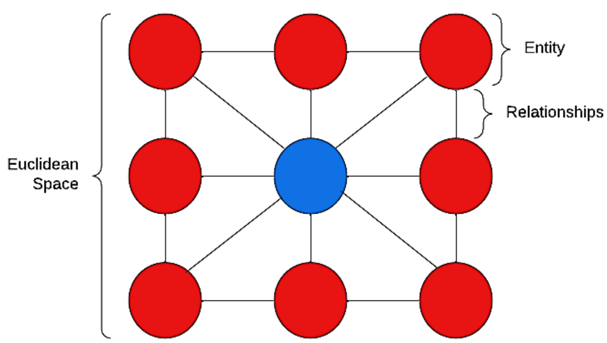

Figure 10. Illustration of Euclidean Space of Entities and Relationships [44]

Based on social recommendation, GNNs can learn and extract features from the graph structure. The key point of GNN is to simulate human relationships with the help of social graphs. They model social-like interactions among users by constructing social networks and searching for relevant items via social recommendation models and algorithms based on strong social ties and weak social-ties [45]. A social recommendation model mainly consists of three components: encoder, decoder, and loss function. The encoder applies GNNs to turn users and items into embeddings [46]. These embeddings would be fed into the decoder, which would predict user preferences. Thus, optimization is ultimately performed by minimizing the prediction error of user preferences [47]. Figure 11 is an illustration of GNN-based social recommendation system.

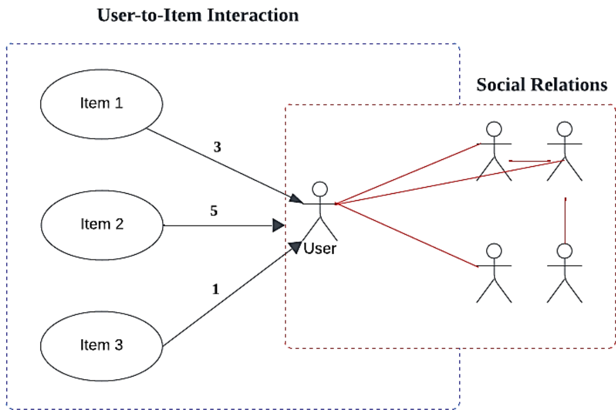

Figure 11. Illustration of GNN-based social recommendation system. [47]

These ideas were suggested by researchers who observed how their social circles impact users' decision-king [34]. Therefore, by incorporating users' social information into RS, the accuracy of personalized recommendation generated would increase significantly, enabling complex interactions and relationships between users and items being identified [48].

On the other hand, the GNN techniques, Graph Convolutional Matrix Completion (GC-MC) and Neural Graph Collaborative Filtering (NGCF), have been developed for enhancing user-item interaction analysis concerning social network data. The process may require extensive incorporation of social relationships into recommendations, making them very precise. In addition, it entails other conditions under which this failure occurs due to the absence of social network data on the experimental cases [49], [50], [51].

To summarize, the capabilities of semantic-based filtering are enhanced by ontology-based and graph-based methods due to the deeper understanding of context and discovery of unseen relationships. Turning them into capable instruments for better recommendation accuracy and user satisfaction. The methods may introduce newer difficulties, such as the mandatory extensive domain knowledge to build ontologies and the high computational complexity with graph processing, especially in large and specialized datasets. Nonetheless, the integration of deep learning into semantic-based filtering will continue to spearhead advancements in the domain of personalized RS.

Over recent years, the exploration for Generative AI boomed in the realm of RS. Generative AI technique's purpose is to generate high-quality synthetic data akin to real-world data using sophisticated algorithms and models [52]. This is exceptionally attractive for approaches needing rare and specialized datasets such as the semantic-based filtering method. If achieved, Generative AI could produce its own original content rather than relying on external labelled datasets used to train traditional AI models [53]. In this space, Generative Adversarial Networks (GANs) and Variational Autoencoders (VAES) are the main foundational models used by contemporary researchers and industry practitioners.

A discriminator and a generator are the two neural networks that make up a standard GAN. While a discriminator's job is to evaluate the correctness of the real data and the data produced by the generator neural network, the generator's job is to produce high-quality synthetic data [54]. This is also known as adversarial learning, whereby the model refines knowledge by feeding it the wrong information on purpose to make the necessary corrections for a better prediction [52]. GANs work by having a generator and a discriminator engage in a two-player adversarial game, constantly reevaluating its generated synthetic data until it achieves the real-world quality [54]. Data noise and sparsity are less of a concern in GAN-based recommendation models [55].

Equation (1) is a loss function designed to detect if the quality of the synthetic data generated by GAN-based models is as realistic as the actual dataset. The (u, i) is the value of user-item interaction taken from real data, while G(z, i) is the generated interaction based on hidden variables z and the item i. GANs are a promising approach for this task, as they are designed to produce synthetic data samples that closely resemble real data samples [56]. GANs are served as the ideal foundational model for researchers to experiment with, spawning a variety of modified GANs according to their use cases, such as Conditional GANs (cGans), Convolutional GANs, Wassersteuig GAN, FairGAN, STRGAN (Social Trust Relationships GENERATIVE Adversarial Network), and MRNGAN model based.

$$\min_{G}\max_{D}\left[E_{((u,i)\sim p_{\mathrm{dt}}(u,i))}\left(\log D\,(u,i)\right)+E_{(z\sim p(z))}\left(1-\log D\left(G(z,i)\right)\right)\right] \tag{1}$$

Aside from this, VAES is an autoencoder that can capture the underlying distribution of input data by learning to encode data into a lower-dimensional latent space and producing new data samples that resemble the input [57]. According to [19], VAE converts interactions between items and users into a latent space (group of stored embedded data), allowing a deeper understanding of user preferences. VAE arranges, condenses, and discovers high-level features to enable its unsupervised learning process based on its extracted nonlinear features. RS will be able to expand the focus of user recommendations, enhance user satisfaction and encourage user engagement [56]. For VAES, the use of Autoencoders (AEs) cannot be overlooked. Its task is to lower data occupancy in data storage, enhance interoperability by discovering essential data features, generalize new data, and handle noisy or unfinished datasets [53].

3. RELATED WORKS

3.1 Work done by Haruna et al.

[58] proposed a system for recommending scientific papers based on public contextual data. This system aims to address the cold start issues for new users without rich user profile information and expand the possible scope of publicly available recommended papers.

This proposed recommender system has five stages: extraction stage, synthetization stage, content similarity calculation stage, combination and aggregate stage using a top N ranking method. The digital bibliography and library project (DBLP) dataset, which contains over 50,000 research publications including title, abstract, venue, authors, citations, and key terms, is used to assess this suggested system.

The evaluation metrics utilized are precision, recall, and F1-score. This study concluded its success in its alternative approach to using public contextual metadata in generating research papers for users without rich user profile information. However, the study stated its limitation around limited evaluation process and the system does not consider the reasons behind the input paper, only recommends the most relevant ones.

3.2 Work done by Kartheek et al.

[59] suggested a recommender system that uses knowledge graph embedding and is based on semantics. This proposed system attempts to solve cold start and sparsity issues by utilizing knowledge graphs to explain semantic explanation of recommendations.

This proposed system will construct a knowledge graph using embedded lower dimension data. Next, using graph embedding, the system will predict any missing links in the knowledge graph, thus generating a knowledge base. When representing a fact in the knowledge base, scoring functions such as translation-based scoring functions and factorization-based scoring functions will work to improve its accuracy. This system is evaluated using the Movielens dataset.

The evaluation metrics included in this proposed system are Mean Rank (MR), Mean Reciprocal Rank (MRR), and Hits@N, a proportion of predictions with a rank which is defined by a certain threshold. The proposed model has successfully decreased the likelihood of the mentioned issues due to its semantically meaningful recommendations produced by the knowledge base although there are no mentions of any known disadvantages.

3.3 Work done by Chew et al.

[60] proposed RS, they experimented with a hybrid ontology-based recommender system by employing an improved version of data enrichment process and utilizing Singular Value Decomposition (SVD). This project aims to improve recommendation accuracy by substantially improving upon a previously experimented Hybrid-based RS using ontology modelling by [61].

The applied improvement is largely centred around the data enrichment process, whereby the system enhances the existing data with additional information such as extra Book Attributes from Google Books APIs using matching ISBN into its existing ontology construction along with several algorithm optimizations such as weighted average to calculate the final item-item similarity matrix. The ontology is constructed using Neo4j graph database platform. The Book-Crossing Dataset was used for this model, which includes demographic information, books, and ratings.

Root Mean Square Error (RSME) and benchmark datasets to gauge model performance. The proposed system's overall results showed that their methods have successfully increased the model's accuracy using extra information from Google Books API, with a comparatively lower RSME score and computation time than the previously experimented model by [62].

3.4 Work done by Chaudhuri et al.

[62] proposed a RS using a Systematic Hidden Attribute-based Recommendation Engine (SHARE) model. To produce recommendations, this system considers a research paper's originality, applicability, complexity, variety, and user purpose.

This paper is focused on solving cold start issues for newer users in the system by filtering using Keyword Diversification Measurement (KDM), Sentence Complexity Analysis (SCA), Citation Analysis Over Time (CAOT), Scientific Quality Measurement (SQM), and Topic, the system would identify pertinent articles depending on the user's search and compile them into a list of possible choices. The system would rank the best papers by using a Multi-Criteria Decision Analysis (MCDA). This model is evaluated based on Cite-U-Like and Scopus datasets.

The evaluation metrics included are relevancy, precision, novelty, diversity, quality (user given scores), CTR (how many of the user's suggested papers were clicked), and response time. The paper concluded that the proposed system could recognize the targeted features from the papers, capture users' dynamic activity, and predict personalized quality recommendations based on these attributes.

3.5 Work done by Zhang et al.

[63] proposed a RS by using semantic representation of cited papers' relations and content to generate citation recommendations. This proposed system aims to help researchers swiftly find alternative or supplementary references by recommending specific relevant and suitable references, reducing potential missing citations when writing a paper.

The system would first generate co-citation relationships and then extract citation content from the papers through corresponding sentences and represent them according to four citation content criteria such as automatic summarizing of CS&SS (SuCS&SS), current sentences (CS), current sentences and surrounding sentences (CS&SS), and current sentences and surrounding sentences that do not include additional references (CS&SS-). Afterward, semantic representation based on the extracted content will be generated using network representation learning algorithms. Finally, a cosine similarity calculation will be conducted to generate recommendations. This proposed system is evaluated using PLOS ONE dataset extracted in 2018 under the heading of artificial intelligence.

This system is evaluated using AUC, MAP, and case study of the recommender system for qualitative evaluation. Although their paper concluded the satisfactory results of the system's framework on tackling their intended problem, further improvements could still be made since the effectiveness of citation recommendation is not greatly improved.

3.6 Work done by Pinedo et al.

[64] introduced the Artikulo Zientifikoen Gomendio-sistema (ArZiGo), a web-based prototype system for recommending scientific articles such as the Semantic Scholar Open Research Corpus to users on a higher level of personalization with a heavy focus on user experience, user interests, long-term activities, and user feedback of recommended scientific articles.

The User Interface, Knowledge Bases, Search Module, Interaction Processing Module, and Recommendation Module are the five primary parts of the ArZiGo system. Firstly, users would send queries to the search module, the search module would display the results filtered from the knowledge base. Once this interaction is recorded by the interaction processing module, the interaction will be converted and stored as a form of implicit feedback. Additionally, this prototype system implements more than one recommendation algorithm such as CBF, ALS-CF (Alternating Least Squares), BRP-CF (Bayesian Personalized Ranking), and Hybrid-based algorithms.

Due to the lack of datasets for implicit user feedback, a synthetic data generator was implemented to simulate a real-life dataset for evaluation. The evaluation consists of 30 experts to evaluate the recommendations generated with MRR and precision.

3.7 Work done by Xiao et al.

[65] proposed a TCRec method for predicting co-authorship interactions for a recommender system. This system recommends research papers by predicting co-authorship occurrence by learning the connections between papers and authors.

Coauthor Interactive Ternary, Tripartite Heterogeneous Academic Graph with Dual-attention, and Paper extractor are the three stages of the system. This system is evaluated using two datasets: Web of Science (WOS) and DBLP.

The evaluation metrics included are component evaluation and metapath evaluation. These metrics are specifically created to assess the overall effectiveness of the component design of the suggested model. They have concluded with superior results with the proposed model; however, it is limited to the domain of research paper datasets.

3.8 Work done by Gharibi et al.

[66] suggested a convolutional neural network (CNN)-based ontology-based recommender system that uses a deep learning methodology. The goal of this suggested system is to prevent the reduction of recommendation accuracy when incorporating numerical data, categorical data, and image features whilst attempting to improve the overall recommendation process.

There are 4 phases in this proposed recommender model: product image extraction, image scoring, user preferences profile creation, similarity calculation, and generate suggestions. This proposed system is evaluated based on Netflix movie dataset.

The evaluation metrics for this model include MAE, RSME, Precision, Recall, F1-score, and User Score. Overall, they concluded that using the included metrics, their proposed system has achieved about 25-30% improvement over other system methods. This proposed system has potential scalability issues when dealing with larger datasets or real-time recommendations.

3.9 Work done by Bahrani et al.

[67] suggested a hybrid semantic recommender system that uses imputation to improve data. Aside from improving recommendation accuracy and increasing system performance, it strives to manage the difficulties of sparsity, scalability, and cold start.

This paper uses a hybrid-recommender system containing CB and collaborative using a modified k-nearest neighbors (KNN) algorithm and clustering on its ontology graph construction to produce its recommendations. This proposed model comprises two main components: a pair of distinct RS and an aggregator.

The evaluation metrics included in this proposed system are MAE, Rating Correction (RC), Precision, Recall, and F-measure. An extensive dataset benchmark was also conducted on the same dataset. In summary, they concluded that their proposed model has successfully tackled the issues with an improved WordNet ontology to resolve the cold start issue, while scalability and sparsity issues were resolved using their modified KNN within the collaborative recommender system. These improvements did not reduce their model's overall predictive precision. Nonetheless, their current proposed model's result is only applied for domains in movie recommendations.

3.10 Work done by Nura et al.

[68] proposed a personalized scholarly-based RS (IPSPR), designed to address the cold start issues often faced by newly published research papers.

This suggested methodology is to determine the similarity between candidate papers and the Paper of Interest (POI) based on contextual information using a Content-Based Recommendation. This method is evaluated on an experiment dataset containing 50 researchers in different fields. The evaluation metrics included in this proposed system are Mean Average Precision (MAP), Precision, Recall, F1-score, and algorithm benchmarking.

Table 1 shows a compiled summary of related works examined in the previous section. A few notable observations here worth mentioning are the widely used graphing methods to get the most of their available data to achieve semantic and criteria optimizations to improve recommendation accuracy regardless of the data domain.

Table 1. Summary of Related Works

Article	Description	Dataset	Evaluation Metrics
[58] Research paper recommender system based on public contextual metadata	This study proposed a RS that emphasizes adaptation.	• DBLP dataset	• Precision • Recall • F1-score
[59] Building Semantic Based Recommender System Using Knowledge Graph Embedding	This system uses knowledge graphs to overcome cold start and sparsity issues. It is evaluated using non-traditional metrics.	• Movielens dataset.	• MR • MRR • Hits@N
[60] A Hybrid Ontology-based Recommender System Utilizing Data Enrichment and SVD Approaches.	Based on an existing iteration of the model, gather additional information for the ontology construction and use SVD to convert stored data from the ontology construction to improve accuracy of the model and reduce computation time.	• Book-Crossing dataset	• RSME • Benchmark Datasets
[62] SHARE: Designing multiple criteria-based personalized research paper recommendation system	This paper proposed a RS using SHARE. It considers a research paper's novelty, relevancy, complexity, diversity, and user's intention to generate recommendations.	• Cite-U-Like dataset • Scopus dataset	• Relevancy • Precision • Novelty • Diversity • Quality • CTR • Response Time
[63] Citation recommendation using semantic representation of cited	This paper proposed a recommender system that is based on citation, co-citation relationships, and citation content.	PLOS ONE dataset was extracted under the domain of	• AUC • MAP • Case Study

papers' relations and content		Artificial Intelligence in 2018.	
[64] TCRec: A novel paper recommendation method based on ternary coauthor interaction	This proposed model is called TCRec. It operates using a paper extractor, Tripartite Heterogeneous Academic Graph, and a coauthor interactive ternary.	• WOS dataset • DBLP dataset	• MR • MRR • Hits@N
[65] Ontology-based recommender system: a deep learning approach	The proposed system utilizes CNN to capture user data and construct its ontology to improve RS for numerical, categorical, and visual data. Based on ontology construction, find similarities and recommend product to user accordingly	• Netflix movie dataset	• MAE • RSME • Precision • Recall • F1-score • Score of the User
[66] A hybrid semantic recommender system enriched with an imputation method	This proposed model is a hybrid semantic recommender system made to overcome problems such as cold start, scalability, and sparsity. This is achieved using a modified KNN algorithm and missing value imputation mechanism.	• TN dataset • ICS dataset • UCS dataset	• MAE • RC • Precision • Recall • F-measure
[67] An Author-Centric Scientific Paper Recommender System to Improve Content-Based Filtering Approach	This proposed system uses Top N Recommended Papers and cosine similarity scores to solve cold start issues with newly published research papers.	• Experiment Dataset	• MAP • Precision • Recall • F1-score Algorithm Benchmark
[68] ArZiGo: A recommendation system for scientific articles	The ArZiGo prototype recommender system has five components for tracking user behaviour and uses four recommendation algorithms. A custom evaluation was used to evaluate the prototype	• Semantic Scholar Open	• Precision (with 30 experts) • MRR

These datasets employed by researchers within this domain have rich metadata and open access to the public (Table 2). Detailed metadata such as title, abstracts, tags, citations are instrumental for the filtering process. However, not all datasets such as Movielens and Netflix Movie are applicable to developing RARS, but they hold similar dataset structure as pure research paper dataset to simulate and model user preferences and recommendations.

Despite their advantages, there are certain drawbacks and limitations to some of these datasets. Domain-Specific Bias such as Movielens and Netflix Movie are not designed to RARS. This could lead to inconsistencies when modelling recommendations purely on academic preferences. Furthermore, synthetic or simulated data from the TN dataset may not represent real-world behaviour. Finally, some of the datasets from related works are exempted from further elaboration due to accessibility issues and experimental nature of the dataset structure.

In summary, this section has reviewed ten relevant literatures regarding academia's recent work on semantic-based article RS. Various strengths and weaknesses of each implementation in the paper are rigorously analysed and discussed. Additionally, evaluation methods are also taken note of according to each distinctive dataset.

Table 2. Datasets of Related Works

Dataset	Description
DBLP Dataset	A bibliographic database of computer science research papers and proceedings. Available in XML, JSON, and plain text format.
Movielens Dataset	A rating and free-text tagging from the MovieLens platform. Available in different.
Book-Crossing Dataset	A rating record for books by users.
Cite-U-Like Dataset	A collection of academic papers with their respective citations from the CiteULike platform.
Scopus Dataset	A collection of citation and academic papers from the Elsevier platform.
PLOS ONE Dataset	A collection of academic papers from the open-access journal PLOS ONE platform.
WOS Dataset	A collection of document classification dataset from the WOS database.
Netflix Movie Dataset	A collection of movie data from Netflix's movie catalogue.
TN (Tennessee) Dataset	A collection of research paper related to Tennessee or the Tennessee Eastman process.
ICS (Industrial Control Systems) Dataset	A collection of research paper related to cybersecurity, anomaly detection, and system optimization.
UCS (Unconfined Compressive Strength) Dataset	A collection of research paper related to geotechnical engineering.
Semantic Scholar Open Dataset	A collection of scientific papers gathered from the Semantic Scholar Public API. It is part of the Research Corpus Dataset.

4. EVALUATION METRICS

Common ground truth-based metrics typically employed by researchers are precision, recall, and F1-score. Equation (2), (3), and (4) are the formulas for precision, recall, and F1-score. These metrics are used to measure the likeness ratio between retrieved similar recommendation results and truly similar results. Equation (2) measures how many predicted positive instances are true against all actual positive instances, whereas Equation (3) measures how many actual positive instances were correctly identified against false positive instances. High precision value entails fewer false positives, a critical evaluation metric for spam detections. On the other hand, high recall value signals fewer false negatives, an important evaluation for medical diagnosis applications. Equation (4) is the mean of precision and recall, useful for more complex applications such as image classification or RS, where a balance of both is crucial. Despite their expendable utilities, ground-truth data is required for these metrics to be reliable.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

While MRR and MAP are also used extensively by researchers from related works, they require further ground truth modification such as recommendation rankings to determine the correctness of the system's accuracy. It is noteworthy that ground truth-based metrics can be tailored according to specific use cases of the system.

In addition, ground truth-free metrics such as serendipity, novelty, diversity, and user satisfaction of the received recommendations can also be used to evaluate its accuracy. This metric is necessary when ground truth data is absent when creating the RARS, which is often the case for most RS. Notwithstanding, viable evaluation methods, namely third-party expert evaluation and engagement metrics like Click-Through Rate, Dwell Time, User Feedback/Surveys and Conversion Rate can also be integrated with system and observe accuracy improvements over time.

5. DISCUSSIONS

By intently examining the ten of the related works, three notable trends emerge. The first trend is in the deviation of traditional RARS to develop a more sophisticated approach in delivering more accurate and engaging results. To illustrate, aside from processing the contents of the research paper with NLP methods to create a recommendation model, these related works utilize metadata, user criteria, author specific relations, and imputed data to generate recommendations. The second trend is the prevalent use of deep learning techniques in the domain of RARS such as SVD by [62], CNN by [65], and modified KNN by [66]. While few of these methods are used for other use cases such as image processing, computer vision, and data analysis, they are being modified by researchers to improve recommendation results. The third trend is the common utilization of ontology methods. Almost all the related works utilize some form of relationship such as coauthors, public context, criteria, citation and semantic relationships. It is evident that relationship among entities within a RARS is perceived to have a major impact on its performance.

It is certain that researchers within this domain are actively exploring new data and methods to create robust RARS, however, ontology and semantic methods are commonly utilized regardless of the type of data being used as shown by the related works. In the future, it can be expected that a new layer of foundation will be laid upon the traditional RS processes with a heavy emphasis on ontology and semantic techniques. Additionally, this layer will also focus upon the adaptability of the system across varying use cases aside from recommending research papers, able to integrate within unrelated designed systems to deliver recommendations. As an example, according to [69], the recent advances from LLMs have allowed RS to treat user actions as a language itself to deliver recommendations while simultaneously achieving user engagement, optimal computing performance, scalability, and continuous self-improvement such as the Generative Recommenders (GRs), pioneered by Meta's researchers.

RS's main objective is to assist users in finding accurate and trustworthy information in the shortest period possible so they can concentrate on much broader objectives. The goals could vary from enjoying fun entertainment, receiving quality educational content, or making cost-effective and high-value purchases. The interacting user satisfaction ultimately measures the effectiveness of an RS.

The evolution from traditional settings, from recommending books in libraries to contemporary use-cases like recommending products on an e-commerce platform has been marked with extraordinary achievements. However, the challenges mentioned in this paper persist, namely high computational costs, susceptibility to errors, time consumption, serendipity issues, data sparsity, cold start, scalability potential, rating biases, high complexity, and availability of specialized datasets. Although promising methods from experimental systems have emerged, lingering concerns such as developing specialized evaluation metrics, advanced knowledge engineering, and context-specific solutions remain unsolved. Addressing these issues will require more research and refinement to increase the RS' accuracy, scalability, and application potential across all domains, especially in academia.

6. CONCLUSION

In closing, this review paper has described how different approaches are utilized for varying levels of RS methods to achieve specific objectives. Overall, ontology-based and graph-based filtering methods are better candidates for implementation due to their capabilities in managing domain-specific knowledge, complex relationships, and sparse data. Ontology-based methods are suitable for representing semantic relationships between research topics, while graph-based methods can model interactions among authors, citations, and research areas.

To avoid the potential challenges involved in ontology-based and graph-based filtering methods, RL, such as the MDPs, can be considered for integration. By modelling the user interactions as a sequence of states, actions, and rewards, the system could simulate and adapt to dynamic user behaviours, continuously learning from their feedback over time. The integration of RL into RS models could improve its decision-making process, decreasing its likelihood of errors from a complex recommendation environment while raising its overall performance.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Sir-Yuean Lim: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Noramiza Hashim: Writing – Review & Editing;
Lanh Le Thanh: Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflicts of interest were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] G. Foley, "How information overload is killing your progress and what to do about it," *Medium*, Jan. 30, 2024. [Online]. Available: https://medium.com/@gregzfoley/how-information-overload-is-killing-your-progress-and-what-to-do-about-it-9c28848871c6

[2] I. Bouchrika, "Overcoming Information Overload in Higher Education: The Power of Document Summarization," *Research.com*, Sep. 26, 2024.

[3] S. Raza et al., "A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice," 2024, arXiv preprint arXiv:2407.13699.

[4] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *Applied Sciences (Switzerland)*, vol. 10, no. 21, pp. 1–20, 2020, doi: 10.3390/app10217748.

[5] S. Apathy, "History of recommender systems: overview of information filtering solutions," *Onespire - SAP and IT Services*, Sep. 12, 2023. [Online]. Available: https://onespire.net/history-of-recommender-systems/.

[6] D. Roy, and M. Dutta, "A systematic review and research perspective on recommender systems," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.

[7] M. Nura, and Z.A. Hamisu, "Author-Centric Scientific Paper Recommender System to Improve Content-Based Filtering Approach," *Int. J. Softw. Eng. Comput. Syst.*, vol. 10, no. 1, pp. 40–49, 2024, doi: 10.15282/ijsecs.10.1.2024.4.0122.

[8] D. Curcic, "Number of Academic Papers Published Per Year – WordsRated," *Wordsrated*, Jun. 01, 2023. [Online]. Available: https://wordsrated.com/number-of-academic-papers-published-per-year/.

[9] R. Habib, and M.T. Afzal, "Sections-based bibliographic coupling for research paper recommendation," *Scientometrics*, vol. 119, no. 2, pp. 643–656, 2019, doi: 10.1007/s11192-019-03053-8.

Formatted: Not Highlight

Formatted: Not Highlight

[10] T. Dai, L. Zhu, Y. Wang, H. Zhang, X. Cai, and Y. Zheng, "Joint model feature regression and topic learning for global citation recommendation," *IEEE Access*, vol. 7, pp. 1706–1720, 2019, doi: 10.1109/ACCESS.2018.2884981.

[11] Ministry of Higher Education Malaysia, *Malaysia Education Blueprint 2015–2025 (Higher Education)*, 2015. [Online]. Available: https://www.mohe.gov.my/en/download/publications-journals-and-reports/pppm-2015-2025-pt/102-malaysia-education-blueprint-2015-2025-higher-education/file.

[12] Ministry of Higher Education Malaysia, *Statistik Pendidikan Tinggi 2024: Bab 1 – Makro*, 2024. [Online]. Available: https://www.mohe.gov.my/muat-turun/statistik/2024-4/1702-bab-1-makro-2024-update-pdf/file

[13] C.K. Kreutz, and R. Schenkel, "Scientific paper recommendation systems: a literature review of recent publications," *Int. J. Digit. Libr.*, vol. 23, no. 4, pp. 335–369, 2022, doi: 10.1007/s00799-022-00339-w.

[14] Z. Ali, G. Qi, K. Muhammad, B. Ali, and W.A. Abro, "Paper recommendation based on heterogeneous network embedding," *Knowledge-Based Systems*, vol. 210, 2020, doi: 10.1016/j.knosys.2020.106438.

[15] N. Sakib et al., "A hybrid personalized scientific paper recommendation approach integrating public contextual metadata," *IEEE Access*, vol. 9, pp. 83080–83091, 2021, doi: 10.1109/ACCESS.2021.3086964.

[16] Y.-K. Ng, "CBRec: a book recommendation system for children using the matrix factorisation and content-based filtering approaches," *Int. J. Business Intell. Data Min.*, vol. 16, no. 2, 2020, doi: https://doi.org/10.1504/ijbidm.2020.104738.

Commented [G1]: Missing doi

[17] N.W. Rahayu, R. Ferdiana, and S.S. Kusumawardani, "A systematic review of ontology use in E-Learning recommender system," *Computers and Education: Artificial Intelligence*, vol. 3, 2022, doi: 10.1016/j.caeai.2022.100047.

[18] "A better way to make the recommendations that power popular platforms," Stanford Graduate School of Business, Sep. 4, 2024. [Online]. Available: https://www.gsb.stanford.edu/insights/better-way-make-recommendations-power-popular-platforms#:~:text=YouTube%20has%20attributed%2070%25%20of,to%2080%25%20of%20content%20consumption.

[19] T. Zhong, Z. Wen, F. Zhou, G. Trajcevski, and K. Zhang, "Session-based recommendation via flow-based deep generative networks and Bayesian inference," *Neurocomputing*, vol. 391, pp. 129–141, 2020, doi: 10.1016/j.neucom.2020.01.096.

[20] Z. Li, and X. Zou, "A Review on Personalized Academic Paper Recommendation," *Comput. Inf. Sci.*, vol. 12, no. 1, p. 33, 2019, doi: 10.5539/cis.v12n1p33.

[21] S. Lin, G. Lee, and S.-L. Peng, "Academic article recommendation by considering the research field trajectory," in *Proc. Int. Conf. Artificial Intelligence and Soft Computing*, pp. 447–454, 2021, doi: 10.1007/978-3-030-65407-8_39.

[22] A. Shahid et al., "Insights into relevant knowledge extraction techniques: a comprehensive review," *J. Supercomput.*, vol. 76, no. 3, pp. 1695–1733, 2020, doi: 10.1007/s11227-019-03009-y.

[23] btd, "Serendipity: A new dimension in recommender systems," *Medium*, Nov. 16, 2023. [Online]. Available: https://baotramduong.medium.com/recommender-system-serendipity-c40c052f8199.

[24] S. Ahmad, and M.T. Afzal, "Combining metadata and co-citations for recommending related papers," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 28, no. 3, pp. 1519–1534, 2020, doi: 10.3906/elk-1908-19.

[25] M.P. Geetha, and D.K. Renuka, "Research on recommendation systems using deep learning models," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, pp. 10544–10551, 2019, doi: 10.35940/ijrte.D4609.118419.

[26] S. Malik, A. Rana, and M. Bansal, "A survey of recommendation systems," *Inf. Resour. Manage. J.*, vol. 33, no. 4, pp. 53–73, 2020, doi: 10.4018/IRMJ.2020100104.

[27] W.-E. Kong, S.-C. Haw, N. Palanichamy, and S. H. A. Rahman, "An e-learning recommendation system framework," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 1, pp. 10–19, 2024, doi: 10.18517/ijaseit.14.1.19043.

[28] S.M. Al-Ghuribi, and S.A. Mohd Noah, “Multi-criteria review-based recommender system–the state of the art,” *IEEE Access*, vol. 7, pp. 169446–169468, 2019, doi: 10.1109/ACCESS.2019.2954861.

[29] L. Chen, G. Chen, and F. Wang, “Recommender systems based on user reviews: the state of the art,” *User Modeling and User-Adapted Interaction*, vol. 25, no. 2, pp. 99–154, 2015, doi: 10.1007/s11257-015-9155-5.

[30] R.D’Addio, M. Conrado, S. Resende, and M. Manzato, “Generating recommendations based on robust term extraction from users’ reviews,” in *Proc. 20th Brazilian Symposium on Multimedia and the Web*, 2014, pp. 55–58, doi: 10.1145/2664551.2664583.

[31] C. Yang, X. Yu, Y. Liu, Y. Nie, and Y. Wang, “Collaborative filtering with weighted opinion aspects,” *Neurocomputing*, vol. 210, pp. 185–196, 2016, doi: 10.1016/j.neucom.2015.12.136.

[32] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens,” in *Proc. 1994 ACM Conf. Comput. Support. Coop. Work (CSCW ’94)*, pp. 175–186, 1994, doi: 10.1145/192844.192905.

[33] F. Hdioud, B. Frikh, and B. Ouhbi, “Multi-criteria recommender systems based on multi-attribute decision making,” in *Proc. Int. Conf. Information Integration and Web-Based Applications & Services*, pp. 203–210, 2013, doi: 10.1145/2539150.2539176.

[34] K. Pachauri, “Semantic similarity for recommender system,” *Medium*, Dec. 15, 2021. [Online]. Available: https://medium.com/analytics-vidhya/semantic-similarity-for-recommender-system-d72c58dfe686.

[35] H. Zhou, F. Xiong, and H. Chen, “A comprehensive survey of recommender systems based on deep learning,” *Appl. Sci.*, vol. 13, no. 20, p. 11378, 2023, doi: 10.3390/app132011378.

[36] “word_embeddings” [Online]. Available: https://web.engr.oregonstate.edu/~huanlian/teaching/ML/2024fall/unit4/word_embeddings.html

[37] T.R. Gruber, “A translation approach to portable ontology specifications,” *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, 1993, doi: 10.1006/knac.1993.1008.

[38] J. K. Tarus, Z. Niu, and G. Mustafa, “Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning,” *Arti6f. Intell. Rev.*, vol. 50, no. 1, pp. 21–48, 2018, doi: 10.1007/s10462-017-9539-5.

[39] M. A. Paredes-Valverde, M. A. Rodriguez-Garcia, A. Ruiz-Martinez, R. Valencia-Garcia, and G. Alor-Hernandez, “ONLI: An ontology-based system for querying DBpedia using natural language paradigm,” *Expert Syst. Appl.*, vol. 42, no. 12, pp. 5163–5176, 2015, doi: 10.1016/j.eswa.2015.02.034.

[40] M. Harrathi, N. Touzani, and R. Braham, “A hybrid knowledge-based approach for recommending massive learning activities,” in *Proc. 2017 IEEE/ACS 14th Int. Conf. Computer Systems and Applications (AICCSA)*, pp. 49–54, 2017, doi: 10.1109/AICCSA.2017.150.

[41] G. George, and A. M. Lal, “Review of ontology-based recommender systems in e-learning,” *Computers & Education*, vol. 142, p. 103642, 2019, doi: 10.1016/j.compedu.2019.103642.

[42] Xtreme1, “The 'ontology' in machine learning - multimodal data training,” *Medium*, Feb. 11, 2023. [online]. Available: https://medium.com/multisensory-data-training/the-ontology-in-machine-learning-e4ba59cd3fc4.

[43] O.C. Santos, and J.G. Boticario, “Practical guidelines for designing and evaluating educationally oriented recommendations,” *Computers & Education*, vol. 81, pp. 354–374, 2015, doi: 10.1016/j.compedu.2014.10.008.

[44] U.A. Bhatti, H. Tang, G. Wu, S. Marjan, and A. Hussain, “Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence,” *International Journal of Intelligent Systems*, vol. 2023, pp. 1–28, Feb. 2023, doi: 10.1155/2023/8342104.

[45] X. Wang, W. Lu, M. Ester, C. Wang, and C. Chen, "Social recommendation with strong and weak ties," *Proc. 25th ACM Int. Conf. Inf. Knowl. Manag.*, pp. 5–14, 2016, doi: 10.1145/2983323.2983701.

[46] Q. Wang et al., "Learning domain-independent representations via shared weight auto-encoder for transfer learning in recommender systems," *IEEE Access*, vol. 10, pp. 71961–71972, 2022, doi: 10.1109/ACCESS.2022.3188709.

[47] L. Ye, H. Xie, Y. Lin, and J.C.S. Lui, "Rewarding social recommendation in OSNs: Empirical evidences, modeling and optimization," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 9, pp. 4410–4424, 2022, doi: 10.1109/TKDE.2020.3038930.

[48] A. Farmaki, H. Olya, and B. Taheri, "Unpacking the complex interactions among customers in online fan pages," *Journal of Business Research*, vol. 125, pp. 164–176, 2021, doi: 10.1016/j.jbusres.2020.11.068.

[49] R. van den Berg, T.N. Kipf, and M. Welling, "Graph convolutional matrix completion," *arXiv preprint*, 2017. Available: http://arxiv.org/abs/1706.02263

[50] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," *Proc. 42nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 165–174, 2019, doi: 10.1145/3331184.3331267.

[51] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1267–1275, 2012, doi: 10.1145/2339530.2339728.

[52] M. O. Ayemowa, R. Ibrahim, and M. M. Khan, "Analysis of recommender system using generative artificial intelligence: A systematic literature review," *IEEE Access*, vol. 12, pp. 87742–87766, 2024, doi: 10.1109/ACCESS.2024.3416962.

[53] S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative AI," *Business & Information Systems Engineering*, vol. 66, no. 1, pp. 111–126, 2024, doi: 10.1007/s12599-023-00834-7.

[54] I.J. Goodfellow et al., "Generative adversarial networks," *arXiv preprint*, 2014. [Online]. Available: http://arxiv.org/abs/1406.2661.

[55] M. Gao et al., "Recommender systems based on generative adversarial networks: A problem-driven perspective," *Information Sciences*, vol. 546, pp. 1166–1185, 2021, doi: 10.1016/j.ins.2020.09.013.

[56] H. Chen, S. Wang, N. Jiang, Z. Li, N. Yan, and L. Shi, "Trust-aware generative adversarial network with recurrent neural network for recommender systems," *International Journal of Intelligent Systems*, vol. 36, no. 2, pp. 778–795, 2021, doi: 10.1002/int.22320.

[57] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special lecture on IE*, vol. 2, no. 1, pp. 1-18, 2015.

[58] K. Haruna, M. A. Ismail, A. Qazi, H. A. Kakudi, M. Hassan, S.A. Muaz, and H. Chiroma, "Research paper recommender system based on public contextual metadata," *Scientometrics*, vol. 125, no. 1, pp. 101–114, 2020, doi: 10.1007/s11192-020-03642-y.

[59] M. Kartheek, and G.P. Sajeev, "Building semantic based recommender system using knowledge graph embedding," *2021 Sixth International Conference on Image Information Processing (ICIIP)*, pp. 25–29, 2021, doi: 10.1109/ICIIP53038.2021.9702632.

[60] L.J. Chew, S.C. Haw, S. Subramaniam, and K.W. Ng, "A hybrid ontology-based recommender system utilizing data enrichment and SVD approaches," *Journal of System and Management Sciences*, vol. 12, no. 5, pp. 139–154, 2022, doi: 10.33168/JSMS.2022.0509.

[61] W. Liu, and Q. Li, "Collaborative filtering recommender algorithm based on ontology and singular value decomposition," *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pp. 134–137, 2019, doi: 10.1109/IHMSC.2019.10127.

[62] A. Chaudhuri, M. Sarma, and D. Samanta, "SHARE: Designing multiple criteria-based personalized research paper recommendation system," *Information Sciences*, vol. 617, pp. 41–64, 2022, doi: 10.1016/j.ins.2022.09.064.

[63] J. Zhang, and L. Zhu, "Citation recommendation using semantic representation of cited papers' relations and content," *Expert Systems with Applications*, vol. 187, p. 115826, 2022, doi: 10.1016/j.eswa.2021.115826.

[64] I. Pinedo, M. Larrañaga, and A. Arruarte, "ArZiGo: A recommendation system for scientific articles," *Information Systems*, vol. 122, pp. 102367, 2024, doi: 10.1016/j.is.2024.102367.

[65] X. Xiao, J. Xu, J. Huang, C. Zhang, and X. Chen, "TCRec: A novel paper recommendation method based on ternary coauthor interaction," *Knowledge-Based Systems*, vol. 280, pp. 111065, 2023, doi: 10.1016/j.knosys.2023.111065.

[66] S.J. Gharibi, K. BagheriFard, H. Parvin, S. Nejatian, and S.H. Yaghoubyan, "Ontology-based recommender system: a deep learning approach," *Journal of Supercomputing*, vol. 80, no. 9, pp. 12102–12122, 2024, doi: 10.1007/s11227-023-05874-0.

[67] P. Bahrani, B. Minaei-Bidgoli, H. Parvin, M. Mirzarezaee, and A. Keshavarz, "A hybrid semantic recommender system enriched with an imputation method," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 15985–16018, 2024, doi: 10.1007/s11042-023-15258-4.

[68] M. Nura, and Z. Adamu Hamisu, "Author-centric scientific paper recommender system to improve content-based filtering approach," *International Journal of Software Engineering and Computer Systems*, vol. 10, no. 1, pp. 40–49, 2024, doi: 10.15282/ijsecs.10.1.2024.4.0122.

[69] Murrell, and Zhai, "Is this the ChatGPT moment for recommendation systems? | Shaped Blog." [Online]. Available: https://www.shaped.ai/blog/is-this-the-chatgpt-moment-for-recommendation-systems

BIOGRAPHIES OF AUTHORS

	**Sir-Yuean Lim** graduated with a degree in International Relations from the University of Nottingham, is now in his final year of computer science with data science degree in Multimedia University. His research focuses on recommendation systems and AI applications. He can be contacted via email: limsiryuaen000@gmail.com.
	**Noramiza Hashim** is a lecturer and researcher at the Multimedia University in data science and multimedia technology. She obtained her PhD in Information Technology under a joint program between Multimedia University (MMU), Malaysia, and Université de La Rochelle, France. Her research interests lie in the fields of image and video processing, computer vision, and machine learning. She has published in various peer-reviewed journals and conferences, reviewed several academic publications, and served on the program committee for academic conferences. She can be contacted via email noramiza.hashim@mmu.edu.my.
	**Lanh Le Thanh** is a lecture and Vice Dean at the Faculty of Information Technology, Dong Nai Technology University, Bien Hoa City, Vietnam. His research interests include Optimization System, Electrical Electronic Engineering. The Intelligence of Things, Intelligent Power systems, and Optoelectronics, Photonics and Optical System. He can be contacted via email lethanhlanh@dntu.edu.vn.