Egyptian Computer Science Journal Vol. 43 No.2 May 2019 ISSN-1110-2586

A Literature Survey on Recommendation Systems for Scientific Articles

Basma Moukhtar, Akram Salah, Cherry Ahmed

Computer Science Department, Faculty of Computers and Information,

Cairo University, Cairo, Egypt

b.moukhtar@fci-cu.edu.eg, akram.salah@fci-cu.edu.eg, c.ahmed@fci-cu.edu.eg

Abstract

Researchers consume a lot of time searching for relevant resources for their research. This slows down the research process and reduces the researchers' productivity. Nowadays, Recommender Systems play an important role in facilitating to the users finding what they are looking for. This study discusses the use of Recommender Systems to enhance the research process. Relying on the fact that research topics in a specified field are related, we use semantics to represent their relationships in terms of their relevance to the field. The objective is to develop a framework for semantic-based Recommender System for scientific articles, integrating the concepts of Semantic web (Ontologies) and Recommender Systems (semantic-based). To fulfill this objective, we first mention the problems that face the research nowadays and then we show a survey of the existing recommender systems in different fields and in the field of Recommendation of scientific articles.

Keywords: *Recommender Systems, Semantics, Ontology***.**

1. Introduction

Scientific research suffers from many shortcomings all over the world. Researchers face a lot of problems and obstacles that may slow down their research. Starting from choosing the research point, methodology and resources to publishing and financial issues. [1] Categorized the obstacles that face the researchers to personal and family obstacles, social factors, technical skills, organizational and professional obstacles, and societal obstacles. Other articles also handled these issues like [2] , [3] and [4].

This shows that there are many problems and obstacles facing the research. A great obstacle that may face a researcher, is the information overload. Along with the search engines problems, reaching the relevant resources for the research becomes a great burden to the researcher.

The term "information overload" was mentioned the first time in 1964 by an American social scientist Bertram M. Gross in his book "The Managing of Organizations: the administrative Struggle" [5]. Gross defined information overload as follows:

"Information overload occurs when the amount of input to a system exceeds its processing capacity. Decision makers have fairly limited cognitive processing capacity. Consequently, when information overload occurs, it is likely that a reduction in decision quality will occur."

There exist many other definitions for information overload like that in [6] , [7], [8] and [9]. Our definition for information overload is:

"A huge and non-structured set of information, where relationships between its entries are missing. That makes it difficult to the decision maker (the researcher in our case) to choose relevant information and extract knowledge of the related concepts to his point of interest."

This may be the result of:

- Huge volumes of information being constantly created.
- A lack of clear structure of information entries and poor relationships between those entries.
- The simplicity of creating, duplicating and sharing of information online.
- The exponential increase in channels to receive information by; radio, television, print media, websites, e-mail, mobile telephony, RSS feeds, etc.
- High volumes of conflicting and contradictory inaccurate information.
- No simple methodologies for quickly processing, comparing and evaluating information sources.

Search Engines suffer from many problems. Some of the search engine problems were mentioned in [10], [11] and [12]. The most important of them is providing millions of non-structured search results whose relationships are not clear. The results are not categorized by topics or authors. Also the "keyword-based" nature of search engines leads to missing related concepts of the main search keyword.

Recommender systems are tools for filtering and sorting items and information. They help users find what they want quickly. Thus help in solving the information overload and the search engine problems. Recommender Systems approaches include content-based, collaborative and knowledge-based recommendation. One of the important approaches for recommender systems is the semantic-based approach. Semantic-based recommender systems classify users and items according to their domains and interests. They may build an ontology describing the entities and their relationships.

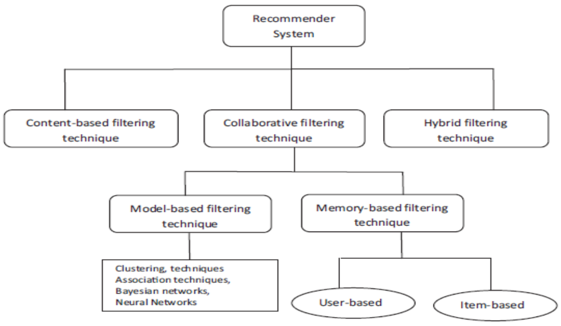
Our Research Objective is to make use of the recommender systems benefits to enhance the research process. This can be achieved by developing a framework for semantic-based Recommender System for scientific articles, integrating the concepts of Semantic web (Ontologies) and Recommender Systems (semantic-based). This is the main motivation for this survey. We made a survey of the existing recommender systems in different fields and in the field of Recommendation of scientific papers.

The rest of the paper is organized as follows: Section 2 shows the background. Section 3 shows the search queries and the criteria used in the survey. Section 4 is the literature review. Section 5 is the discussion and conclusion.

2. Background

2.1. Recommender Systems

Recommender systems are tools for filtering and sorting items and information. They use opinions of a community of users to help individuals in that community to more effectively identify content of interest from a potentially overwhelming set of choices. There is a huge diversity of algorithms and approaches that help creating personalized recommendations. Two of them became very popular: collaborative filtering and content-based filtering. They are used as a base of most modern recommender systems. There are some modern domain-specific recommender approaches such as Semantic-based approaches, Context-aware approaches, Social approaches, Cross-lingual approaches and domain-specific approaches [13]. The Goals of any recommender system are: Relevance, Novelty, Serendipity and diversity [14]. The main challenges that recommender systems face include: Cold-start problem, Trust, Scalability, Privacy, and others [13].



The basic approaches of recommender systems are shown in Figure 1 [15]:

Figure 1: Recommender Systems Basic Approaches [15]

2.2. Recommender Systems Traditional Approaches

2.2.1. Content-based Approach

Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his/her taste. Taste is based on how the user rated items. Generally, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem [13].

2.2.2. Collaborative-based Approach

The idea of collaborative filtering is finding users in a community that share appreciations. If two users have same or almost same rated items in common, then they have similar tastes. A user gets recommendations to those items that he/she hasn't rated before, but that were already positively rated by users in his/her neighborhood. The taste is considered to be constant or at least change slowly [13].

a) User-based approach:

The users perform the main role. If certain majority of the customers has the same taste, then they join into one group. Recommendations are given to user based on evaluation of items by other users from the same group, with whom he/she shares common preferences. If the item was positively rated by the community, it will be recommended to the user [13].

b) Item-based approach:

Referring to the fact that the taste of users remains constant or change very slightly, similar items build neighborhoods based on appreciations of users - Items with the same ratings are joined into one group [13].

c) Model-based approach:

Machine Learning and Data Mining methods are used in the context of predictive models. Examples of such model-based methods include decision trees, rule-based models, Bayesian methods and latent factor models [14].

2.2.3. Hybrid Approach

This approach combines different techniques of collaborative approaches and content-based approaches. Using hybrid approaches can avoid some limitations and problems of pure recommender systems, like the cold-start problem. The combination of approaches can proceed in different ways: 1) Separate implementation of algorithms and joining the results. 2) Utilize some rules of content-based filtering in collaborative approach. 3) Utilize some rules of collaborative filtering in content-based approach. 4) Create a unified recommender system that brings together both approaches (Asanov and Daniar 2011).

2.3. Recommender Systems Special Approaches

Special approaches of recommender systems include: Semantic-based Recommendation Systems, Context-aware Recommendation Systems, Social Recommendation Systems, and Cross-Domain Recommendation.

2.3.1. Semantic-based Recommender Systems

Those systems are based on a knowledgebase, usually defined as a concept diagram (like taxonomy) or ontology. Taxonomy plays an important role in semantic analysis. Classification of items and/or users concerning their domains and groups brings much efficiency in recommendation system [13].

2.3.2. Context-aware Recommender Systems

Context-based or context-aware recommender systems take various types of contextual information into account, while making recommendations. Such contextual information could include time and location [14].

2.3.3. Social Recommender Systems

Social recommender systems are based on network structures, social cues and tags, or a combination of these various network aspects. There are several kinds of social recommender systems: Structural Recommendation of Nodes and Links, Product and Content Recommendations with Social Influence, Trustworthy Recommender Systems and Social Tagging Feedback for Recommendations [14].

2.3.4. Cross-domain Recommender Systems

In cross-domain systems similarities of users are computed domain-dependent. An engine creates local neighborhoods for each user according to domains. Then, computed similarity values and finite set of nearest-neighbors are sent for overall similarities computation. Recommender system determines the overall similarity, creates overall neighborhoods and makes predictions and recommendations [13].

2.4. Evaluation of Recommender Systems

Evaluation of recommender systems is done by one or more of the following three ways: user studies, online evaluation and offline evaluation.

2.4.1. User Studies

User studies typically measure user satisfaction through explicit ratings. Users receive recommendations generated by different recommendation approaches, users rate the recommendations, and the approach with the highest average rating is considered most effective [16].

An important advantage of user studies is that they allow for the collection of information about the user interaction with the system. On the other hand, the active awareness of the user about the testing of the recommender system can often bias her choices and actions. It is also difficult and expensive to recruit large cohorts of users for evaluation purposes [14].

2.4.2. Online Evaluation

They measure the acceptance rates of recommendations in real-world recommender systems. Acceptance rates are typically measured by click-through rates (CTR), i.e., the ratio of clicked recommendations to displayed recommendations [16]. This approach is sometimes less susceptible to bias from the recruitment process, because the users are often directly using the system in the natural course of affairs [14].

The main disadvantage is that such systems cannot be realistically deployed unless a large number of users are already enrolled. Therefore, it is hard to use this method during the startup phase. Furthermore, such systems are usually not openly accessible, and they are only accessible to the owner of the specific commercial system at hand [14].

2.4.3. Offline Evaluation

An offline experiment is performed by using a pre-collected data set of users choosing or rating items. Using this data set we can try to simulate the behavior of users that interact with a recommendation system. They require no interaction with real users, and thus allow us to compare a wide range of candidate algorithms at a low cost. The downside of offline experiments is that they can answer a very narrow set of questions, typically questions about the prediction power of an algorithm. Thus we cannot directly measure the recommender's influence on user behavior in this setting [17].

The quality of a recommendation algorithm can be evaluated using different types of measurement which can be accuracy or coverage. Accuracy is the fraction of correct recommendations out of total possible recommendations while coverage measures the fraction of objects in the search space the system is able to provide recommendations for.

a) Accuracy Measurement

Statistical accuracy metrics: evaluate accuracy of a filtering technique by comparing the predicted ratings directly with the actual user rating. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Correlation metrics are used [15].

Decision support accuracy metrics: These metrics help users in selecting items that are of very high quality out of the available set of items. The metrics that are popularly used are Reversal rate, Weighted errors, Receiver Operating Characteristics (ROC) and Precision Recall Curve (PRC), Precision, Recall and F-measure [15].

b) Coverage Measurement

The percentage of items and users that a recommender system can provide predictions. Prediction may be practically impossible to make if no users or few users rated an item. [15]

3. Search Queries and Criteria

As mentioned in the abstract that our objective is to develop a framework for semantic-based Recommender System for scientific articles, integrating the concepts of Semantic web (Ontologies) and Recommender Systems (semantic-based). This survey is a step in our research. To identify relevant literature for our survey, we made a literature search on Google Scholar in time range from 2000 to 2018. Our search was directed to two main directions:

First: the recommendation systems in different fields.

Second: the recommendation systems for scientific articles.

For the first direction, the used search queries are: “Recommendation Systems”, “Semantic-based Recommendation/Recommender Systems”, “Topic Modeling”, “Context-based Recommendation/Recommender Systems”, “Social-based Recommendation/Recommender Systems” and “Cross-Domain Recommendation/Recommender Systems”.

We classified the results according to the main approach used in the recommendation regardless of the time and place it was published (20 papers). The main approaches covered are Semantic-based Recommendation, Topic Modeling, Multi Criteria Recommendation, Graph-based Recommendation, Context-based Recommendation, Social-based Recommendation and Cross-Domain Recommendation.

For the second direction, we first read a survey paper about recommendation of scientific papers to know the different approaches used for this purpose as well as the evaluation methods and the limitations of the current research [16]. Then we began to search for individual papers that make recommendation for scientific papers. We first organized the results sequentially by years. Then in the discussion and conclusion section, we grouped them by approach also.

The queries used for the second direction are: “Recommendation/Recommender Systems for Research”, “Recommendation/Recommender Systems for Scientific Research”, “Recommendation/Recommender Systems for Scientific Papers” and “Recommendation/Recommender Systems for Scholarity Papers”.

The main categories of our literature review is shown in Figure 2:

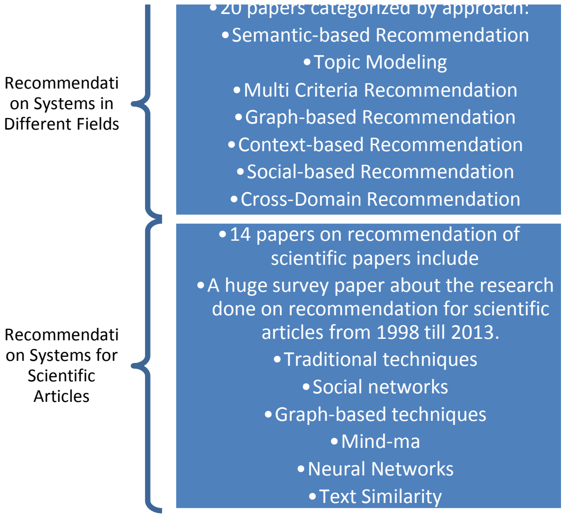


Figure 2: The Main Directions of the Literature Review

4. Literature Review

As mentioned in the previous section, this literature review categorizes the references into two main directions: First: Recommendation Systems in miscellaneous fields and Second: Recommendation Systems for scientific articles. The two directions are shown in the following subsections 4.1 and 4.2.

5. Recommender Systems in Miscellaneous Fields

At first we gathered general papers about different recommender systems in various fields. Then we categorized the papers according to the approach used for recommendation. The main approaches are: semantic-based recommender systems, topic modeling and text classification, multi criteria recommender systems, graph-based recommender systems, context-based recommender systems, social-based recommender systems and cross-domain recommender systems. This can be shown in Figure 3:

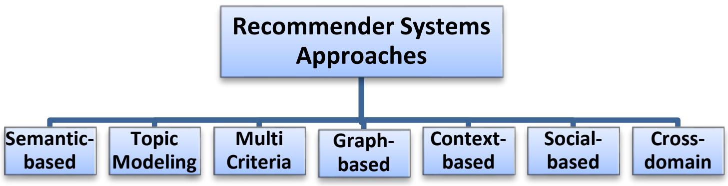


Figure 3: General Recommender Systems Main Approaches

4.1.1 Semantic-based Recommender Systems

[18] The framework uses a traditional method based on TF-IDF, and several ontology-based methods to recommend new articles to the user. The paper concludes with the evaluation of the different methods, which shows that the new ontology-based method that they propose in this paper performs better (w.r.t. accuracy, precision, and recall) than the

traditional method and, with the exception of one measure (recall), also better than the other considered ontology-based approaches.

[19] Propose the use of Wikipedia as ontology to solve the problems of using traditional ontologies for the text analysis in text-based recommendation systems. A full system model that unifies semantic-based analysis with a collaborative via content recommendation system is presented. Basically, the proposed semantic analysis model relies on the ESA model. A document gets annotated with a vector of weighted Wikipedia concepts (Articles) taking into consideration the top N concepts. Finally, concepts hierarchy-based approach is applied. It is used to re-weight the concepts according to the hierarchical structure. Collaborative-via-Content hybrid recommendation technique is used. The results proved that the concepts hierarchy based technique gives better results than applying the spread activation alone and is more robust. Also the algorithm increases the recommendation accuracy in cold start cases.

[20] Introduces how to use semantic technologies to improve the weighted Slope One scheme as an easy way to build collaborative filtering and it often gives better performance in usability, realizability, and efficiency. They build a movie recommender system which uses both traditional datasets and Linked Data. They made offline experiment and compare their algorithm with five other algorithms but unfortunately didn’t obtain good results.

[21] uses semantics in the recommendation of economic articles. Decision makers need economical information to drive their decisions. The Company Actualis SARL is specialized in the production and distribution of a press review about French regional economic actors. To reduce the overload of useless information, the company is moving towards a customized review for each customer. This paper presents a new type of recommendation based on the semantic description of both articles and user profile.

[22] Is about an application of the semantic-based technique in E-Recruitment. The motivation is that both users and recruiters suffer from the un-relevant information they receive from the recruitment websites. This research constructs a semantic vocabulary of the domain from the job offers corpus and initializes a profile for each user based on his Curriculum Vitae. A comparison between the actual recommendations service and this system is held. Recall, Precision, F-measure, False Acceptance and False Reject metrics are used. All give better results with the applied algorithm.

[23] This research used ontologies along with information inferred from social networks to enhance the accuracy of a tourism recommendation system. They present a semantic social recommender system employing two ontologies. First: a user interest ontology built from the user answers to direct questions and from his data and behavior from his Facebook profile. Second: Tunisian Medical Tourism ontology. Their system improved the quality of recommendation for Tunisian tourism domain.

[24] The motivation of this research was that health-related videos are very popular on YouTube, but their quality is always a matter of concern. One approach to enhancing the quality of online videos is to provide additional educational health content, such as websites, to support health consumers. This study investigates the feasibility of building a content-based recommender system that links health consumers to reputable health educational websites from MedlinePlus for a given health video from YouTube. The relevance of the recommended health websites from MedlinePlus to the videos was measured using information retrieval metrics such as the normalized discounted

cumulative gain and precision at K. Their results demonstrate the feasibility of using a semantic content-based recommender system to enrich YouTube health videos.

4.1.2 Topic Modeling and Text Classification

[25] proposes a simple feature extraction algorithm that can achieve high document classification accuracy in the context of development-centric topics. Given a focused topic and a training set of candidate authoritative pages on the topic, extract an appropriate feature set of textual terms that can be used in conjunction with any standard classifier (Bayes or SVM) to determine if any document is related to the topic or not. The time taken by feature extraction is trivial in comparison to the classification time. In comparison to WebKB, Bayes give better results but SVM had bad accuracy

[26] The objective of this research is to support clinical decision-making, such as recommending the need for a certain medical test while avoiding intrusive tests or medical costs. They represented reports according to their topic distributions as this is more compact than bag-of-words representation and can be processed faster than raw text in subsequent automated processes. And, finally an aggregate topic classifier was built where reports are classified based on a single discriminative topic that is determined from the training dataset. The results were good on small datasets.

[27] uses topic modelling to recommend tags for tweets. They propose a novel method for unsupervised and content-based hashtag recommendation for tweets. This is to easily organize and categorize tweets. The approach relies on Latent Dirichlet Allocation (LDA) to model the underlying topic assignment of language classified tweets. The advantage of this approach is the use of a topic distribution to recommend general hashtags.

4.1.3 Multi-criteria Recommender Systems

[28] emphasizes the importance of multi criteria recommendation systems as it gives more accurate results than single criteria systems. They deal with the problem of choosing the criteria (attributes) as a decision making problem. Propose an objective weight determination method called CCSD method, which is referred to as correlation coefficient (CC) and standard deviation (SD) integrated approach for determining the weights of attributes. Evaluation is done on a movie system. Their approach proves the idea that multi criteria recommendation gives more accurate results than mono criteria.

[29] proposes an Item-based Multi-Criteria Collaborative Filtering (IMCCF) algorithm that integrates the items' semantic information and multi-criteria ratings of items to lessen known limitations of the item-based CF techniques (sparsity and cold start). The input is a raw matrix of user-item MC ratings, which consists of multi-criteria ratings of M users on N items and hierarchical tree structured item taxonomy. The item taxonomy, given by the domain experts, has a set of main items' categories where items should belong to as leaf nodes. They first compute the item-based similarity based on the multi-criteria ratings in the input matrix. And then, they compute the item-based semantic similarity. They combine both similarities to make prediction. The evaluation is done on the Yahoo! Movies dataset. The results were compared with two other algorithms. Their algorithm gives better results on solving both the sparsity and the cold start problem.

4.1.4 Graph-based recommender Systems

[30] Twitter messages are only displayed by time recency. In this paper, they propose to re-rank tweets in user's timeline, by constructing a user profile based on user's previous tweets and measuring the relevance between a tweet and user interest. The user interest

profile is represented as concepts from Wikipedia. They make use of Explicit Semantic Analysis algorithm to extract related concepts from tweets, and then expand user's profile by random walk on Wikipedia concept graph, utilizing the inter-links between Wikipedia articles. The experiments show that this model is effective and efficient to recommend tweets to users.

[31] They present an efficient semantic recommendation method that helps users filter the Twitter stream for interesting content. The foundation of this method is a knowledge graph (KG) that can represent all user topics of interest as a variety of concepts, objects, events, persons, entities, locations and the relations between them. Their method uses the KG and graph theory algorithms not yet applied in social network analysis in order to construct user interest profiles by retrieving semantic information from tweets. Next, it produces ranked tweet recommendations. In addition, they use the KG to calculate interest similarity between users, and they present a followee recommender based on the same underlying principles.

4.1.5 Context-based Recommender Systems

[32] studies the relationships between geo-location information published by users at different times. This geo-location information was used to model user's interest and behavior in order to enhance prediction of user locations. Furthermore, semantic features such as topics of interest and location category were extracted from this information in order to overcome sparsity of data. Several experiments on real twitter dataset showed that the proposed context-based prediction model which applies machine learning techniques outperformed traditional probabilistic location prediction model that only rely on words extracted from tweets associated with specific locations.

4.1.6 Social-based Recommender Systems

[33] Social network systems, like last.fm, play a significant role in Web 2.0, containing large amounts of multimedia-enriched data that are enhanced both by explicit user-provided annotations and implicit aggregated feedback describing the personal preferences of each user. They investigate the role of these additional relationships in developing a track recommendation system. Taking into account both the social annotation and friendships inherent in the social graph established among users, items and tags, they created a collaborative recommendation system that effectively adapts to the personal information needs of each user. They performed a series of comparison experiments between the Random Walk with Restarts model and a user-based collaborative filtering method. The results show that the graph model system benefits from the additional information embedded in social knowledge. In addition, the graph model outperforms the standard collaborative filtering method.

[34] introduces a tag and social-based recommender system. Most of the current tag-based systems do not emphasize using only the common tags on common items to measure user similarity, which is the approach they used in this recommender system. Their user similarity metric not only takes into account the interaction of users with items (in terms of tagging items), but also incorporates the social interactions of users (in terms of friendship and membership). A recommendation for a target user is based on both user and item similarities.

[35] proposes a social recommender system that combing preference similarity (based on collaborative filtering), reputation-based trust and social relations between users. Preference Similarity is computed using the rating data. Reputation-based Trust is also

computed using the rating data. System computes users' expertise and reputation according to the difference between rating score and the real one. While the Social Relation is computed using the relationship data. System computes the relationship of two customers according to the coincidence their social network share. Using the real data from Epinions.com, they compared their framework with other five systems to evaluate its performance. The experimental results show that the new recommender systems is quite promising in terms of mean absolute error (MAE), prediction precision, and recommendation precision compared to the traditional ones.

4.1.7 Cross-Domain Recommender Systems

[36] addresses the cold-start problem in Recommendation Systems. They deal with such a cold-start situation exploiting cross-domain recommendation techniques, i.e., they suggest items to a user in one target domain by using ratings of other users in a, completely disjoint, auxiliary domain. They present three rating prediction models that make use of information about how users tag items in an auxiliary domain, and how these tags correlate with the ratings to improve the rating prediction task in a different target domain. They show that the proposed techniques can effectively deal with the considered cold-start situation, given that the tags used in the two domains overlap.

[37] evaluate various CF methods enhanced with user personality traits and cross-domain ratings. Their empirical results on 22,289 Facebook user profiles with preferences for items in several domains –movies, TV shows, music and books– show that incorporating additional ratings from other domains improves recommendation accuracy, and that it is better to enrich user models with both cross-domain rating and personality trait information.

It's clear that our focus is about the semantic-based recommendation as this should be applied in our research. The resources approach categorization can be shown in the pie chart in Figure 4.

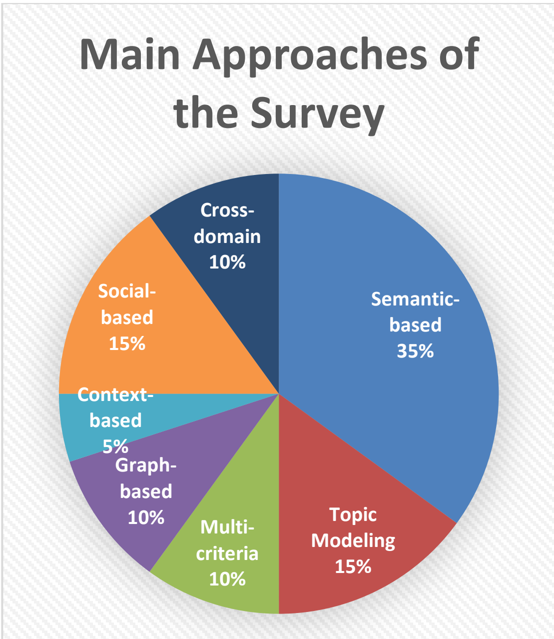


Figure 4: Main Approaches of the Survey

Another view of the resources can be by the year of publication. Our time ranges from 2000 to 2017. As shown in the graph in Figure 5, most of the work is done in 2013.

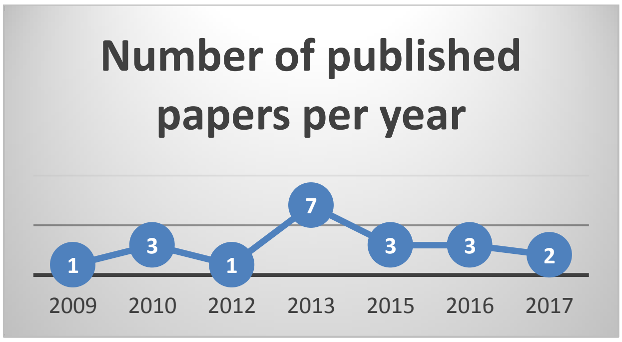


Figure 5: Resources shown by the years of publication

Table 1 shows what approaches are published in which years.

Table 1: The Resources Viewed by the Year of Publication.

Year	Number of papers	Approach
2009	1	Social-based (1)
2010	3	Semantic-based (2), Topic Modeling (1)
2012	1	Graph-based (1)
2013	7	Semantic-based (2), Topic Modeling (2), Multi-criteria (1), Social-based (1), cross-domain (1)
2015	3	Semantic-based (2), Cross-domain (1)
2016	3	Multi-criteria (1), social-based (1), context-based (1)
2017	2	Semantic-based (1), Graph-based (1)

4.2 Recommendation Systems for Scientific Articles

This subsection shows the second direction of our literature review, which is the recommendation systems for scientific articles.

[38] Digital books can significantly enhance the reading experience, providing many functions not available in printed books. They study a particular augmentation of digital books that provides readers with customized recommendations. they systematically explore the application of spreading activation over text and citation data to generate useful recommendations. Their findings reveal that for the tasks performed in their corpus, spreading activation over text is more useful than citation data.

[39] uses a graph-based recommender system that naturally combines the content-based and collaborative approaches. A Hopfield net algorithm was used to exploit high-

degree book-book, user-user and book-user associations. It was found that the system gained improvement with respect to both precision and recall by combining content-based and collaborative approaches.

[40] addresses the problem of document recommendation in a digital library, where the documents in question are networked by citations and are associated with other entities by various relations. Due to the sparsity of a single graph and noise in graph construction, they propose a new method for combining multiple graphs to measure document similarities. A new recommendation framework is developed using semi-supervised learning on graphs. In addition, they address the scalability issue and propose an incremental algorithm. The new incremental method significantly improves the efficiency by calculating the embedding for new incoming documents only.

[41] explores the problem of personalization in a specific kind of social systems known as collaborative tagging systems. The systems of this kind assembled a large volume of user-contributed items, such as Web bookmarks in Delicious, pictures in Flickr, and bibliographic references in CiteULike. However, by the nature of these systems, they lack any kind of centrally provided description, metadata or hierarchical categorization as in more traditional Web systems (i.e., online stores, Web directories, library catalogs). Each contributed item may include user-contributed tags and comments instead. The primary goal of this work was to run a reliable comparison of all combinations of the experimental approaches using standard n-fold-based evaluation approach. The results obtained after two phases of evaluation that the enhancements are beneficial. Incorporating the number of raters into the algorithms leads to an improvement of precision, while tag-based BM25 similarity measure, an alternative to Pearson correlation for calculating the similarity between users and their neighbors, increases the coverage of the recommendation process.

[42] This paper develops an algorithm to recommend scientific articles to users of an online community. It combines the merits of traditional collaborative filtering (for old articles) and probabilistic topic modeling (for new articles), providing an interpretable latent structure for users and items.

[43] targets the problem of result diversification in citation-based bibliographic search. It surveys a set of techniques which aim to find a set of papers with satisfactory quality and diversity. It enhances these algorithms with a direction-awareness functionality to allow the users to reach either old, well-cited, well-known research papers or recent, less-known ones. It also proposes a set of novel techniques for a better diversification of the results.

[16] is a huge survey paper about recommendation systems for scientific articles. It surveys all the related papers from 1998 till 2013 (217 papers). It classifies the papers according to the used evaluation methods, the recommendation class and the shortcomings in the research. They considered seven classes in the field of research-paper recommender systems: Stereotyping, Content-based Filtering, Collaborative Filtering, Co-Occurrence, Graph-based, Global Relevance and hybrid systems. The paper describes each class and mentions the papers that use it.

[44] Proposes a framework of faceted recommendation for scientific articles (abbreviated as FeRoSA) which apart from ensuring quality retrieval of scientific articles for a query paper, also efficiently arranges the recommended papers into different facets (categories). Providing users with an interface which enables the filtering of recommendations across multiple facets can increase users' control over how the

recommendation system behaves. FeRoSA groups the recommendations into four naturally observed facets, namely, Background, Alternative Approaches, Methods and Comparison.

[45] extracts only author information to build relations between articles, i.e., common author relations. Then, these relations and researchers' historical preferences are used together to build a heterogeneous graph for article ranking. They define features to find relevant target researchers who have author-based search patterns by analyzing information on common author relations existing in a researcher's historical preferences. They conduct relevant experiments using a real-world dataset CiteULike to evaluate the impacts of the defined features and the performance of the proposed method.

[46] While user-modeling and recommender systems successfully utilize items like emails, news, social tags, and movies, they widely neglect mind-maps as a source for user modeling. However, millions of mind-mapping users could benefit from user-modeling applications such as recommender systems. The objective of this study is to develop an effective user-modeling approach based on mind maps. To achieve this objective, they integrate a research-paper recommender system in a mind-mapping and reference-management software Docear. The recommender system builds user models based on the users' mind maps, and recommends research papers based on the user models. The findings show that user modeling based on mind maps is a promising research field, and that developers of mind-mapping applications should integrate recommender systems into their applications.

[47] presents a collaborative approach for research paper recommender system. In addition to mining the hidden associations between a target paper and its references, in this paper, they also consider the hidden associations between the target paper's citations. A candidate paper is qualified for consideration if and only if it cited any of the target paper's references and there exist another paper which cited both the candidate and the target papers simultaneously. This strictness in qualifying a candidate paper helps in enhancing the overall performance of the approach and the ability to return relevant and useful recommendations at the top of the recommendation list. The approach gives better results than collaborative filtering in terms of Precision and Recall measures,

[48] They address the lack of social recommendation approaches in social bookmarking websites for scholarly papers. They propose three implicit social networks that exploit data from the users' publication list and bookmarked papers in the social bookmarking websites. Network 1: Readership Implicit Social Network, connects users to the authors of the papers that they have bookmarked. Network 2: Co-readership Implicit Social Network, connects users who bookmark (and presumably read) papers written by the same authors. Network 3: Tag-Based Implicit Social Network, connects users if they use the same tags to annotate their bookmarked papers.

[49] addresses the problem that most of the recommendation approaches which are based on text embedding have utilized bag-of-words technique. While proposed deep learning methods for capturing semantic meanings in the text, have been proved to be effective in various natural language processing (NLP) applications. In this paper, they present a content-based TR (Tag Recommendation) method that adopts deep recurrent neural networks to encode titles and abstracts of scientific articles into semantic vectors for enhancing the recommendation task. They made a comparison with multiple baseline methods in text-based multi-label classification like Naïve Bayes (NB), Support Vector

Machines (SVM) and Latent Dirichlet Allocation (LDA). The overall findings show that the proposed model is effective in representing scientific articles for tag recommendation.

[50] outlines a hybrid technique called the IDSP technique for finding similar papers based on a (Seed Basket) SB of research papers. They considered the Literature Review is a task of three steps. The three steps are: (1) building a reading list of research papers; (2) finding similar papers based on a set of papers; and (3) shortlisting papers from the final reading list for inclusion in a manuscript based on article type. The technique takes multiple seed papers for formulating recommendations, thereby overcoming the gap in earlier studies where similar papers were found for an input paper. The evaluation results indicated that the students' group found the recommended papers to be more useful than the staff group.

To conclude, the research done for the recommendation systems of scientific articles made it clear that first: no much work is done in this field, second: very few use semantics in the recommendation. Also as mentioned by [16] the work in the field of scientific articles recommendation is not continuous; the authors don't continue their research. That's why we are interested in the field. However, in the years 2016, 2017 and 2018, there is more attention done to this kind of recommendation which emphasizes that the research in this area is promising. This can be shown in Figure 6.

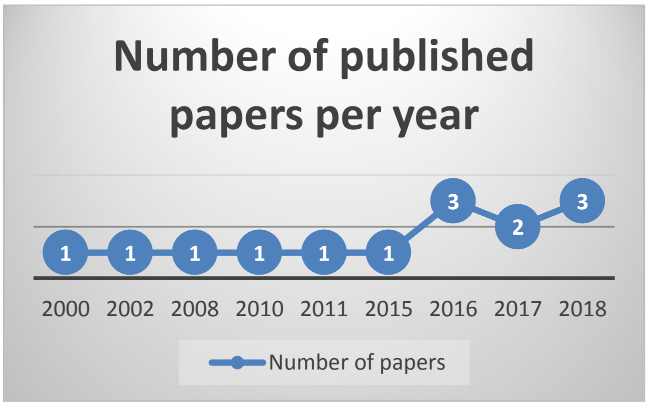


Figure 6: The Number of Papers Published in each Year for Scientific Recommendation

6. Discussion and Conclusion

The previous section introduces a variety of recommender systems. Some of which combine the semantics concepts with the traditional recommendation techniques [18] to [24]. Others use topic modeling to classify data to facilitate data processing [25] to [27]. And others proved that the multi-criteria rating of items gives better recommendations [28]& [29]. Some Graph-based Recommender Systems are also shown [30] & [31]. Context-based and Social-based Systems are shown in [32] to [34]. Finally Cross-domain recommendation samples are mentioned [36] & [37].All these research emphasize the importance of recommender systems in general for facilitating information retrieval.

Also, some research for recommendation systems of scientific articles are shown. One of which is a huge survey paper about the research done on recommendation for scientific articles from 1998 till 2013 [16]. Some use traditional techniques [38], [42], [43], [44], [45], [47] & [50], some use social networks [41], [48] & [49], some use graph-based techniques

[39] & [40] and others use mind-maps [46]. But none of these researches use semantics for the scientific recommendation.

Although, it was noticed that using semantics in terms of building an ontology for the concepts of a certain field, enhances the recommendation than using the traditional techniques only. It is proved to enhance the recommendation in other fields like news [18], document classification [19], movies recommendation [20] , [24] and [29] ,e-recruitment [22], economics [21] and Tourism [23]. But none make use of this enhancements to the favor of researchers.

The aim of this study is to make semantic-based recommendation for scientific articles. That is; to build an ontology for a specific field (e.g. software engineering), and give recommendation based on this ontology along with the user preferences. The output can be like a tree representing the hierarchy of researches related to this research interest. This is supposed to enhance recommendation for researchers and so increasing the research productivity.

References

[1] N. A. Algadheeb and M. A. Almeqren, "Obstacles To Scientific Research In Light Of A Number Of Variables," *Journal of International Education Research,* vol. 10, no. The Clute Institute, 2014.

[2] A. Sawyerr, "African universities and the challenge of research capacity development," *Journal of Higher Education in Africa/Revue de l'enseignement supérieur en Afrique,* no. JSTOR, pp. 213-242, 2004.

[3] M. Al Ataibi, "Causes of vulnerability in the implementation of scientific research among students in Jordanian universities," *Interdisciplinary Journal of Contemporary Research in Business,* vol. 2, pp. 143-164, 2010.

[4] S. A. Alghanim and R. M. Alhamali, "Research productivity among faculty members at medical and health schools in Saudi Arabia. Prevalence, obstacles, and associated factors.," *Saudi medical journal,* vol. 32, pp. 1297-1303, 2011.

[5] B. M. Gross, The managing of organizations: The administrative struggle, New York: Free Press of Glencoe, 1964.

[6] M. Silic, A. Back and D. Silic, "Atos-Towards Zero Email Company," in *ECIS*, 2015.

[7] P. Persson, "Attention manipulation and information overload," *Behavioural Public Policy*, vol. 2, no. Cambridge University Press, pp. 78-106, 2018.

[8] K. M. G. Hoq, "Information Overload: Causes, Consequences and Remedies-A Study," *Philosophy and Progress,* vol. 55, no. 2305-6851, pp. 49-68, 2016.

[9] P. G. Roetzel, "Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development," *Business Research,* no. Springer, pp. 1-44, 2018.

[10] M. R. Henzinger, R. Motwani and C. Silverstein, "Challenges in Web Search Engines,"

in *ACM SIGIR Forum*, ACM, 2002.

[11] D. Lewandowski, "Problems with the use of web search engines to find results in foreign languages," *Online information review,* vol. 32, no. Emerald Group Publishing Limited, pp. 668-672, 2008.

[12] T. Diamond and J. Liang, "Keyword-based search engine results using enhanced query strategies ; Reichhold, Jonathan ; Koperski, Krzysztof". Google Patents 2014.

[13] Asanov and D. , "Algorithms and methods in recommender systems," Berlin Institute of Technology, Berlin, Germany, 2011.

[14] C. C. Aggrawal, Recommender Systems: The Textbook, Springer International Publishing Switzerland, 2016.

[15] F. O. Isinkaye, Y. O. Folajimi and B. A. Ojokoh, "Recommendation systems: principles, methods and evaluation," vol. 16, no. 3, 2015.

[16] J. Beel, B. Gipp, S. Langer and C. Breitinger, "paper recommender systems: a literature survey," vol. 17, no. 1432-5012, 2016.

[17] M. de Gemmis, P. Lops, C. Musto, F. Narducci and G. Semeraro, Recommender Systems Handbook, Boston: Springer, 2015.

[18] W. IJntema, F. Goossen and F. Hogenboom, "Ontology-based news recommendation," 2010.

[19] A. Elgohary, H. Nomir, I. Sabek, M. Samir, M. Badawy and N. A. Yousri, "Wiki-rec: A semantic-based recommendation system using wikipedia as an ontology," in *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference*, 2010.

[20] R. Yang, W. Hu and Y. Qu, "Using Semantic Technology to Improve Recommender Systems Based on Slope One," in *Semantic Web and Web Science*, New York, 2013.

[21] D. Werner, C. Cruz and C. Nicolle, "Ontology-based recommender system of economic articles," 2013.

[22] O. Chenni, Y. Bouda, H. Benachour and Z. Chahniz, "A Content-Based Recommendation Approach Using Semantic User Profile in E-recruitment," in *International Conference on Theory and Practice of Natural Computing*, Charm, 2015.

[23] M. Frikha, M. Mhiri and F. Gargouri, "A semantic social recommender system using ontologies based approach for Tunisian tourism," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 4, pp. 90-106, 2015.

[24] C. L. S. Bocanegra, J. L. S. Ramos, C. Rizo, A. Civit and L. Fernandez-Luque, "HealthRecSys: A semantic content-based recommender system to complement health videos," *BMC medical informatics and decision making*, vol. 17, no. BioMed Central, p. 63, 2017.

[25] R. Power, C. Jay, K. K. Trishank and S. Lakshminarayanan, "Document Classification for Focused Topics," 2010.

[26] S. Efsun, K. Yadav and H.-A. C. , *Topic Modeling Based Classification of Clinical Reports*, ACL (student research workshop), 2013.

[27] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen and R. Van de Walle, "Using Topic Models for Twitter Hashtag Recommendation," in *Proceedings of the 22nd International*

Conference on World Wide Web, ACM, 2013.

[28] H. Ferdaous, B. Frikh and B. Ouhbi, "Multi-criteria recommender systems based on multi-attribute decision making," in *Conference on Information Integration and Web-based Applications & Services*, 2013.

[29] Q. Shambour, M. Hourani and S. Fraihat, "An Item-based Multi-Criteria Collaborative Filtering Algorithm for Personalized Recommender Systems," *International Journal of Advanced Computer Science and Applications 7.8*, pp. 274-279, 2016.

[30] C. Lu, W. Lam and Y. Zhang, "Twitter user modeling and tweets recommendation based on wikipedia concept graph," 2012.

[31] D. P. Karidi, Y. Stavrakas and Y. Vassiliou, "Tweet and followee personalized recommendations based on knowledge graphs," 2017.

[32] A. Galal and A. El-Korany, "Enabling Semantic User Context to Enhance Twitter Location Prediction," in *ICAART (1)*, 2016.

[33] I. Konstas, V. Stathopoulos and J. M. Jose, "On Social Networks and Collaborative Recommendation," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2009.

[34] S. Naseri, "A Tag And Social Network Based Recommender," *Masters Theses, Computer Science Department, Ryerson University*, 2013.

[35] Y. Shen, T. Lv, X. Chen and Y. Wang, "A Collaborative Filtering Based Social Recommender System for E-Commerce," *International Journal of Simulation: Systems, Science and Technology*, vol. 17, pp. 91-96, 2016.

[36] M. Enrich, M. Braunhofer and F. Ricci, "Cold-Start Management with Cross-Domain collaborative filtering and tags," in *International Conference on Electronic Commerce and Web Technologies*, Springer, 2013.

[37] I. Fernández-Tobías and I. Cantador, "On the Use of Cross-Domain User Preferences and Personality Traits in Collaborative Filtering," in *International Conference on User Modeling, Adaptation, and Personalization*, Springer, 2015.

[38] A. Woodruff, J. Pitkow, E. H. Chi and S. K. Card, "Enhancing a digital book with a reading recommender," ACM, 2000.

[39] Z. Huang, W. Chung, T.-H. Ong and H. Chen, "A graph-based recommender system for digital library," 2002.

[40] D. Zhou, S. Zhu, K. Yu, X. Song, B. L. Tseng, H. Zha and C. L. Giles, "Learning multiple graphs for document recommendations," ACM, 2008.

[41] D. Parra-Santander and P. Brusilovsky, "Improving Collaborative Filtering in Social Tagging Systems for the recommendation of scientific articles," in *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, 2010.

[42] W. Chong and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," 2011.

[43] O. Küçüktunç, E. Saule, K. Kaya and Ü. V. Çatalyürek, "Diversifying citation recommendations," vol. 5, no. 4, 2015.

[44] T. Chakraborty, A. Krishna, M. Singh, N. Ganguly, P. Goyal and A. Mukherjee, "Ferosa: A faceted recommendation system for scientific articles," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2016.

[45] F. Xia, H. Liu, I. Lee and L. Cao, "Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences.," *IEEE Transactions on Big Data,* vol. 2, no. 2, pp. 101-112, 2016.

[46] J. Beel, "Towards effective research-paper recommender systems and user modeling based on mind maps," *arXiv preprint arXiv:1703.09109*, 2017.

[47] K. Haruna and M. A. Ismail, "A collaborative approach for research paper recommender system ; Damiasih, Damiasih ; Sutopo, Joko ; Herawan, Tutut," *PloS one,* vol. 12, no. Public Library of Science, 2017.

[48] S. Alotaibi and J. Vassileva, "Implicit Social Networks for Social Recommendation of Scholarly Papers," in *Highlighting the Importance of Big Data Management and Analysis for Various Applications*, Springer, 2018.

[49] H. A. M. Hassan, G. Sansonetti, F. Gasparetti and A. Micarelli, "Semantic-based Tag Recommendation in Scientific bookmarking systems," in *Proceedings of the 12th ACM Conference on Recommender Systems*, ACM, 2018.

[50] A. Sesagiri Raamkumar, S. Foo and N. Pang, "Can I have more of these please? Assisting researchers in finding similar research papers from a seed basket of papers," *The Electronic Library,* no. Emerald Publishing Limited, 2018.