

## Towards Serendipity for Content-Based Recommender Systems

Nur Izyan Yasmin Saat<sup>#</sup>, Shahrul Azman Mohd Noah<sup>\*</sup>, Masnizah Mohd<sup>##</sup>

Centre of Artificial Intelligence, Faculty of Technology and Information Science, National University of Malaysia, Selangor, 43600, Malaysia  
E-mail: p88930@siswa.ukm.edu.my<sup>#</sup>, shahrul@ukm.edu.my<sup>\*</sup>, masnizah.mohd@ukm.edu.my<sup>##</sup>

**Abstract**— Recommender systems are intelligent applications build to predict the *rating* or *preference* that a user would give to an item. One of the fundamental recommendation methods in the content-based method that predict ratings by exploiting *attributes* about the users and items such as users' profile and textual content of items. A current issue faces by recommender systems based on this method is that the systems seem to recommend too similar items to what users have known. Thus, creating over-specialisation issues, in which a self-referential loop is created that leaves user in their own circle of finding and never get expose to new items. In order for these systems to be of significance used, it is important that not only relevant items been recommender, but the items must be also interesting and serendipitous. Having a serendipitous recommendation let users explore new items that they least expect. This has resulted in the issues of *serendipity* in recommender systems. However, it is difficult to define serendipity because in recommender system, there is no consensus definition for this term. Most of researchers define serendipity based on their research purposes. From the reviews, majority shows that unexpected as the important aspect in defining serendipity. Thus, in this paper, we aim to formally define the concept of serendipity in recommender systems based on the literature work done. We also reviewed few approaches that apply serendipity in the content-based methods in recommendation. Techniques that used Linked Open Data (LOD) approaches seems to be a good candidate to find relevant, unexpected and novel item in a large dataset.

**Keywords**— content-based recommendation; serendipity; recommender systems.

### I. INTRODUCTION

The Internet has given users the freedom to obtain and access various kinds of information at any time and at anywhere. According to the Malaysian Communication and Multimedia Commission report on 2017, about 24.5 million people use the internet in their daily life. Among those numbers, about 70% of people use it to stream or download movies and about 48.8% use it for online purchases [1].

Despite of having good technologies and wider network, the high traffic of internet usage resulted in the issues of information overload, whereby information is produce more than what people concern [2]. Recommender system technology is one of the solutions to overcome the problem of information overload. The technology has been applied to various applications such as information retrieval systems [3], on-line learning [4,5] and planning [6]. Recommender systems guide users based on their feedback, in which it can be obtained explicitly from ratings or implicitly through user actions on the web (item purchase/item view) or via users' past purchasing or interactivity experiences when using the system. However, one of the issues currently faced by users

of recommender system is the filter bubble problem, in which users been provided with a bunch of too *similar* recommendation that sometimes are irrelevant to them.

A self-referential loop is created, in which the user is only been provided with items in their range of interest alone. Thus, improving similarity by providing serendipity can be seen as one of the effective ways in handling this issue and hence may improve the recommendations and widen users' view [7]. Serendipity by definition is about something new, interesting and unexpected by users. A famous quote by Steve Jobs; "*people do not know what they want, until you show it to them*" might be linked to serendipity.

Serendipity is a new research problem in recommender systems and therefore receive little attention among researcher in this area. This is particularly because there has been no consensus for the definition of serendipity and the lack of datasets and metrics to support its evaluation. Therefore, realising the importance to establish the concept of serendipity, this paper aim to formally define the concept of serendipity in recommender systems in order to overcome the issue of over-specialisation.

Recommender system is a software tool and techniques to recommend item to user. It also has the effect of guiding users in a personalized way to interesting or useful items in a large space of possible options. The basic models of recommender systems work with two kinds of data: the user-item interactions, such as ratings or buying behaviour; and attributes about the users and items such as users' profile and textual content of items. Methods that use the former are referred as collaborative filtering methods, whereas methods that use the latter are referred as content-based recommender methods. Another basic type of recommendation currently adopted by systems named knowledge-based recommender systems use explicitly specified user requirements whereby external knowledge bases and constrains are used to create the recommendation. Some recommender systems combine the strengths of various types of recommendation methods to create hybrid systems.

The focused of this paper is on content-based recommender method. Systems based on this method can perform well when little or no rating is available from the users. In this case, the systems analyse the items' content that the user has been interacting with, such as items purchased online, movie streaming or any other activities captured and stored by the systems. Content-based recommendation systems use what user has liked in the past to recommend items as illustrated in Fig 1. The basic approach of this system is to match attributes of a user profile based on what the system has received and stored in their database along with the attributes of item content, in order to recommend to the user new similar and interesting items [7].

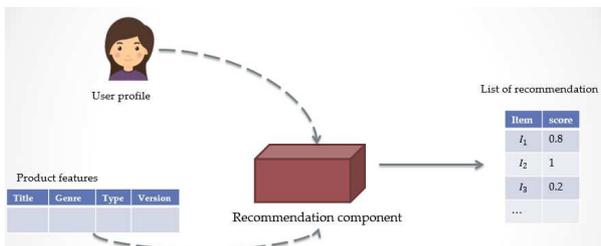


Fig 1: Content-based recommendation (figure adapted from Jannach et al. [8])

### A. Features of Content-Based Recommender System

Recommender system has a huge database consisting of items information and their respective features, hence forming item profile. Whereas a user profile is build up from the information provided that combine the item profile information along with the user preferences. Items not yet seen or experienced by the users but matched with the users' profile will be recommended to the users as illustrated in Fig. 2.

According to Aggarwal [9], basic level of content-based rely on two sources of data. First, a description of variety of items in the terms of content-centric such as product descriptions provided by manufacturer or any related context that consist of text descriptions. Second, a user profile which is generated from actions or information given by users about various items.

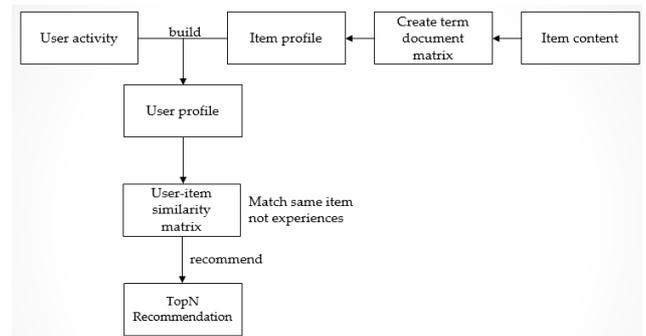


Fig 2: Basic Process of Content-Based RS

Content-based recommender systems use metadata from a set of user,  $U_a$  and set of item  $I$  (represented as keywords) to extract existing items such as item features in order to make item prediction to a target user [9,10]. For a user, such gender, age, occupation and location can be used as a metadata to build a user profile, while for item it depends on the type of item but for certain cases such as movies, the metadata can be variety such as genre, main actors and directors. From the metadata provided, the similarity ( $U_a, I$ ) can be measured based on the keyword found. Various similarity measures can be used. One example of such measures is the *Dice* coefficient as follows

$$Similarity(U_a, I) = \frac{2 * |category(U_a) \cap category(I)|}{|category(U_a) + category(I)|} \quad (1)$$

According to Di Noia & Ostuni [2], content-based recommendation has two main approaches: heuristic-based and model-based. The heuristic based are rooted from the Information Retrieval (IR) and Information Filtering (IF) fields. Items are recommended to a user, based on the comparison made up between its content and user profile. A vector matrix is build up from the analysis of the items liked by the user. A Vector Space Model (VSM) is the most common model to represent the user-item model, whereby items and users are represented as weighted vector such as using the *tf-idf* weighting scheme. The similarities between items and user profile vectors can be computed for example using the cosine similarity measure and the most similar items according to certain threshold will be recommended to the users.

The model-based approaches use Machine Learning (ML) techniques to learn a model of user preferences by analysing the content characteristics from the ratings of items. A regression or classification model is learnt from a collection of items of past ratings that are available. The set will be divided to training and test set that contain feature vectors labelled of with ratings. This learnt user model can be trained to estimate the unknown ratings.

### B. Advantages and Disadvantages of Content-Based Recommender Systems

Some of the advantages of having content based recommender systems, is that it can build user preferences that do not depend on many users, perhaps it can generate

recommendations even there exist only a single user [8]. They act independently and do not exploit ratings from a group of user to build up a user profile. Instead of using ratings, the content-based method can work well using a genre or other item content that are available to process the recommendation. Furthermore, the system is also 'friendly' to new items due to its capability to recommend items that are not yet rated by users.

However, there are few shortcomings in the process of recommendation. First, even it can handle new items, it still faces a problem of cold start user. This happens when a user is new to the Web or recommender system and does not have any history of ratings. Thus, it gives difficulties to the system to understand what the user like in order to build a user profile. Optimum number of rating is needed for a recommender to process recommendation. If the new user only rates a small number of items, hence the system cannot predict accurately what actually the user likes [7,13].

Second, content-based method faces with the issue of limited content analysis. When the provided information is not enough, the system cannot provide suitable suggestion or recommendation to users due to its inability to discriminate items that the user like from dislike. A book recommendation system for example may need to know the authors and genre of books, and sometimes, other additional knowledge [7]. The additional knowledge can be obtained from external resources such as the Wikipedia.

Third, is the issue of over-specialisation where it happens when the system only recommends too similar items to the users [8]. In this situation, the systems only recommend items that have high similarity score, and therefore recommendations are limited towards the same items that the users have rated [7,17]. As an example, a comedy film lover would not be exposed to other genres of film such as horror, because they do not have any experience in it.

The over-specialization has been drawn by the issue of filter bubble as the systems are unable to recommend different items from the user profile [22,23]. Review in this area shows that only a few researchers have focused on this problem in the content-based recommendation system. In order to fix this issue and to subsequently improve the list of recommendation, Lops et al. [7] suggested the application of serendipity in recommendation.

Having a serendipity item to recommend have a good and bad consequences, since some people may like something that surprise them and vice-versa. Although serendipity has been claimed to improve users' experience, dealing with serendipity is a complicated task. Having serendipity will reduce the value of accuracy in recommendation but increase the value of novelty throughout the recommendation. According to a survey in [15], there is no consensus definition on serendipity. The researcher proposed their own definition and reconcile it with their own research. In this paper, we propose to address serendipity by clarifying the definition for serendipity in content-based recommender system.

## II. MATERIAL AND METHOD

When it comes to information searching, Toms [16] explained that there exist three kinds of it. First, people seek information about a well-defined object, in which they find

something they want or information they want to know more about it. Second, people seek an information about object that cannot be fully described, but that will be recognized at first sight. People intend to find something they have a clue with but cannot described what are the things that they really meant. Lastly, people acquire information in an accidental, incidental or serendipitous manner. Sometimes, people love to receive information or item that appear accidentally without being search. For the third kind of information searching, it shows a concern on how serendipity may affect the system and user as the implementation of serendipity-inducing strategy for a content-based, could be adapt to the real world situation. For instance, as a person go shopping or visiting a museum who, while walking around seeking nothing in particular, would find something completely new that she/he has never expected to find.

### A. Definition of Serendipity

Serendipity is a paradoxical concept or term that is hard to explain [17]. The term serendipity is a symbol of valuable, hard to identify, unexpected and only happen at a first sight. According to an online dictionary, serendipity is the fact of finding interesting or valuable things by chance [18]. A serendipity is defined as an art of making unsought finding. This term has been famous since 1754, through a letter written by Horace Walpole to Sir Horace Man that tell the stories of the ancient tale of "*The Three Princes of Serendip*", a story on how the princes making unexpected finding without searching for it and how they used their intelligence to handle something that are not to be find [19].

Serendipity is widely in many fields not just in computer science but also in medical and engineering too. Serendipity expose people to new items or objects that is interesting and novel. Serendipity can be seen as an inverse of similarity, instead of suggesting or recommending item that user is familiar with, it try to enhance the list of recommendation by suggesting new items to them.

According to Maksai et al. [20], serendipitous items are not about interesting alone but must also be useful to users. It is a quality that have both unexpected and useful values. Adamopoulos & Tuzhilin [21] define serendipity as something that related to the concept of unexpected and receive positive response when users look at it. Akiyama et al. [22] defined serendipity as unexpected item, while Iaquina et al. [23] claims that, serendipity is about relevant, novel and unexpected to the user. Serendipity cannot happen if the user already knows what they have been suggested with. De Gemmis et al. [24] on the other hand, proposed serendipity as relevance and unexpectedness, and Jenders et al. [1] define it as unexpected and interesting to the user. Finally, Maccatrozzo et al. [14] propose serendipity as making a pleasant and relevant discovery that was unexpected. Summary of these definitions is as illustrated in Table 1. As can be seen, researchers have proposed different but related definitions for the concept of serendipity. These definitions usually are influenced by the field and domain they are working with. Thus, it is important to clearly define the concept of serendipity in the field of recommender systems.

TABLE I.  
DEFINITIONS OF SERENDIPITY

References	Definitions
Maksai et al. [20]	Interesting and useful.
Adamopoulos and Tuzhilin [21]	Unexpected with positive responses.
Akiyama et al. [22]	Unexpected.
Iaquinta et al. [23]	Relevant, novel and unexpected
De Gemmis et al. [24]	Relevant and unexpected
Jenders et al. [25]	Unexpected and interesting
Maccatrozzo et al. [14]	Pleasant, relevant and unexpected

### B. Features of Serendipity in Content-Based Recommender System

As mentioned in section II-A, the process of recommendation serendipitous item is almost similar with the conventional one. The process can be differentiating by their algorithm and process. Serendipity recommender is a vice-versa to the original content-based. If those systems are focussing on recommending similar item, meanwhile these recent approach tried to recommend some surprising item and novel to the user but still relevant to their profile. As example, if a user is a fan of adventure movie that have actor A, they might also like other genre that starring the same actor. In this system, the features or attributes of an item that have similar features and match the user preferences will be used to find other items using any content learner model such as nearest neighbours, Naïve Bayes or semantic analysis to find any hidden or relation that has been ignored in the system for a serendipitous purpose.

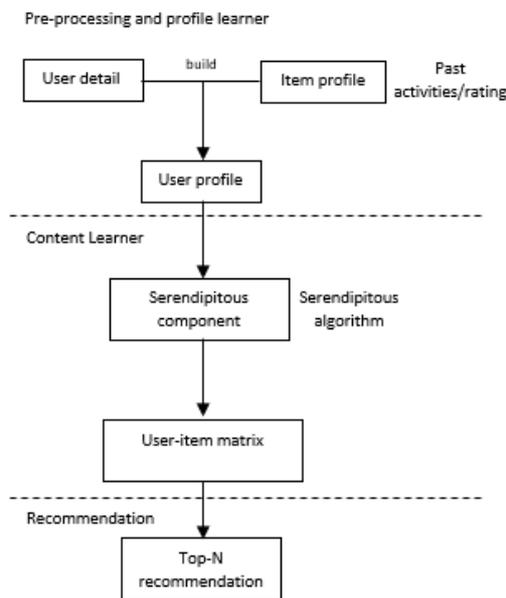


Fig 3: Basic Process of Serendipity in Content-Based RS

Serendipity in content-based recommender system use other related metadata that been processed in the phase of content learner by extracting the item features in order to do a prediction for the target user. For such a movie recommender, other features such as actor, director or genre could be used to find other related movies. There exist many metric approach that can be used to predict the serendipitous

item. One of the approach are from Maccatrozzo et al. [14], they introduce the usage of LOD paths as follow:

$$\{T_1, P_1, T_2, P_2, \dots, T_1, P_1, T_{i+1}\} \quad (2)$$

An LOD path is an ordered set of types and properties that connect two types of  $T_1$  and  $T_{i+1}$ . The cosine similarity measures were used to compare the vector of different lengths.  $l$  is representing the length of pattern. To extract those pattern, the system need a link between item and LOD dataset. For example, LOD path allow them to link the documentary of “Reggie Yates’s Extreme South Africa” to the show “The Sky at Night”. This happen from the link extract, starting from the word “extreme”, which is associated to the musical band Extreme. This band is influenced by the “Queen band” that have a member names Brian May, who was the guest for the show of The Sky at Night. A connection is needed for a relation to exist. Thus from the existing relation, the serendipitous could be generate and can be consider as relevant to the user as the item are still connected either by the keyword used or the genre related.

### C. Concept of Serendipity in Recommender System

In recent years, the concept of serendipity has received much attention in the field of recommender systems, due to the issue that users were recommended with too *similar* items to what he/she has likes in the past. Thus the recommended items are similar to those the users already know and not surprisingly novel and interesting.

Serendipity can be simply defined as one experience in receiving items recommendation that are unexpected and surprise [26]. Toms [27] suggested four strategies to induce serendipity in recommender systems, which were further detailed by De Gemmis et al. [24], as follows:

- Role of chance or blind luck – strategies are implemented based on a random information node generator that provides random suggestions.
- Pasteur Principle (“chance favours the prepared mind”) – implemented using user profile. Different context of information about user preferences can be applied.
- Anomalies and exception – using distances measures in order to identify dissimilar items based on those items from user past information.
- Reasoning by analogy – using abstraction mechanism that allow system to discover the applicability of an existing schema to perform a new situation.

A serendipitous recommendation gives users new experiences to deal with an unknown item that are not likely been discovered that triggers their positive response and motivates them to accept the recommendation [14]. In recommender systems, serendipitous items are items that are interesting, unexpected and novel to the user [23].

Some of the researcher had been applied these strategies to induce serendipity in their research. The ITeM Recommender (ITR) [23] implement the third strategy in their research and new entry points were provided to the user for the items in the system. In this cases, system provides an item that less similar to the user profile. In another approach, Random Walk with Restart and knowledge infusion (RWR-KI) system, they implement the Pasteur principle strategy for

their approach [24]. Using this strategy, they believed the capability of an algorithm could be improving to produce serendipitous recommendation by enhancing the knowledge infusion process that provides the system with memory of world facts and the linguistic competencies. Furthermore, they also believe that this approach could contribute to build the prepared mind. From the analysis review, both of the system might still recommend similar item to the user. Such research by De Gemmis et al. [24], they may provide accurate item but still obvious to the user, in which the item recommended to them are something that user already knows. Both systems might be improved if they have a specific definition of serendipity in recommender system.

Recommending serendipitous items is about suggesting new relevant item that possibly gives satisfaction to users and trigger the positive response in handling it. Most recommender systems suggest similar items to what the users have experienced or liked in the past. Instead of giving the same item to the user, serendipity helps user to explore new item that do not even cross in their mind or any idea about it.

As explained in section II about the features of content-based recommendation, some alteration or modification of techniques have been built in order to recommend serendipitous items. The classic VSM and ML techniques were used widely to learn user model.

#### A. Item Representation.

In details, items are also called attributes or properties that are represented by a set of features. This set of features are known as the textual features which represent the item description that are extracted from unstructured data such as product descriptions, news articles and also Webpages. According to Lops et al. [7], the textual features creates a number of complication when learning a user profiles due to natural language ambiguity, in which the issues of polysemy and synonymy can derived to different meaning of an item.

Thus, a semantic analysis is seen as the best method to encounter the problem faced. One of the key idea is to adopt the knowledge bases such lexicon or ontologies [4,32] as explicit output that can be used to annotate other items related and represent the profiles in order to provide a semantic interpretation of user preferences.

Several serendipity recommender systems have been developed recently using semantic analysis approach in various fields of applications such as movies, news and e-commerce. In the area of serendipity in content-based recommender systems, one of the earliest are ITR (Item Recommender) [23] a system that are capable of providing recommendations for items in several domains (e.g. movies, books) provided that descriptions of items available as a text documents. ITR integrates the knowledge bases from the WordNet lexical ontology. Items are represented according to a synset-based vector space model called, bag-of-synsets (BOS) the extension of bag-of-words (BOW).

A movie recommender [24] is a system that proposed strategy enriches a graph-based recommendation algorithm. This system used the background knowledge such as WordNet and Wikipedia that allows the system to deeply understand the items it deals with.

SIRUP [14] is a system for the recommendation of BBC TV programmes. This system is inspired by the Silvia's curiosity theory, based on the fundamental theory of Berlyne. The novelty of the items in this system are calculated with cosine similarities between items, using Linked Open Data (LOD) paths. This LOD path were calculated by performing the semantic enrichment of the title of the TV programme with DBpedia concepts.

Another features such as tags [3], can also be used as a basic keyword-based approach for item representation especially in conventional content-based recommender systems. These features are yet to be explored in identifying serendipitous items for content-based recommendations. As mentioned in [7], we believe that tags can be useful for item representation since tags can also express users' opinions and emotions that can be represented as the degree of users' satisfaction.

However, in this paper, we planned to propose a serendipity recommender system based on SIRUP model that uses LOD as the explicit sources to find other relevant items that are novel and unexpected to users. This is based on the fact that research has shown positive results of using LOD to find other relevant items that are ignored in the dataset. Furthermore, LOD is an open source that are freely available online and globally accessible to everyone. One notable example of LOD is DBpedia [34].

#### B. Learning User Profiles

In traditional content-based recommender system, machine learning (ML) techniques have been widely used for the task of inducing user profiles. Learning user profiles are important so that system can learn what item is interesting to the user or not with respect to the user preferences [7].

The learning user profiles can be cast as a binary text categorization task, in which the document need to be classified as interesting or not with the respect of user preferences that represent the categories of  $C = \{c_+, c_-\}$ , where  $c_+$  represent the positives class (user-likes) and  $c_-$  as the negative class (user-dislikes)[7].

In ITR [23], a Naïve Bayes text classifier was adapted to filter a like and dislike item by the user. A scholarly paper recommendation system [28] uses neighbouring cluster in order to find dissimilarity user. From the recent research, most of the researcher uses questionnaire to build the user profile [14,22,23]. Literally, the questionnaire was used to collect user's actual impressions about serendipity item and to like and dislike categories were filtered based on user preferences.

#### C. Approaches of Serendipity in Recommender System

According to Kotkov et al. [15] serendipity oriented recommender system can be classified based on the data they used or the architecture of the recommendation. The approaches can be divided into three categories, which are:

- Re-ranking algorithm – enhance the systems by re-rank the output of the list of recommendation, by assigning low ranks to obvious suggestions
- Serendipity-oriented modification – enhance the systems by modifying the accuracy-oriented system and adapt it for a serendipity purpose. For example, the

research of k-furthest neighbour approach [29] is a modified of k-nearest neighbour algorithm. Therefore, instead of suggesting items liked by users similar to a target user, a k-furthest neighbour algorithm recommends items disliked by users dissimilar to a target user.

- Novel algorithm – an algorithm that are not based on the accuracy-oriented recommender system; it can be diverse and develop using different techniques.

Based on the survey made in Kotkov et al. [15], few research on serendipity for content-based recommender system are classified as novel algorithm and only the research by Iaquina et al. [23] is classified under the modifying algorithm as they upgrade an existing recommendation algorithm using the hybrid technique to recommend serendipitous items. Early attempts of serendipity recommendation are illustrated in Iaquina et al. [23], which introduced a hybrid recommender in order to improve over-specialization.

Another research presented in Akiyama et al. [22] proposed a method using distance and general unexpectedness of users. In which, they collect data from users through questionnaire concerning TV programmes based on three categories, “recognized programs”, “serendipitous programs” and “not recognized programs”. In this research, they need to improve the accuracy of the distances used in order to find where the serendipitous programs are laid about.

Jenders et al. [25] proposed an unexpectedness model of topic combinations in articles and a traditional cosine-based similarity model that recommends serendipitous news article. The research only focus on the dissimilarity of the latent topic. They do not consider the relevant and unexpectedness of the recommendation. While Sugiyama & Kan [28], proposed a recommendation that recommend paper based on individual interest.

A research by De Gemmis et al. [24], proposed a graph-based recommendation algorithm that used background knowledge; which are WordNet and Wikipedia. The additional knowledge helps them to introduce non-obvious recommendations to users without hurting the accuracy. However, the system does not make use of the structured knowledge and does not personalize the serendipity approach. They do not take the user interest into consideration. Another recent work by Maccatrozzo et al. [14], used Sylvia’s curiosity theory that based on fundamental theory of Berlyne. They used Linked Open Data (LOD) approach to suggest serendipitous TV programmes. LOD was used to find others related item that maybe ignored in the system. The user profile is build up based on the questionnaire they provide to the target user. However, the system does not take into account the full semantic enrichment as such other textual metadata, such as genre.

Summary of these approaches is as illustrated in Table 2. From the table, can be seen that there is variety of approaches that have been proposed by the past researchers for serendipity recommendation. Each of the approaches have their own pros and cons in recommending serendipitous items to user.

TABLE II.  
SERENDIPITY APPROACHES IN CONTENT-BASED RECOMMENDER SYSTEM

References	Approaches
Iaquina et al. [23]	Hybrid based
Akiyama et al. [22]	Distance metric
Jenders et al. [25]	Unexpectedness model of topic combination
Sugiyama & Kan [28]	Model individual profile
De Gemmis et al. [24]	Graph based with additional knowledge based from Wikipedia and WordNet.
Maccatrozzo et al. [14]	SIRUP model, using Sylvia’s curiosity theory based on fundamental theory of Berlyne.

### III. RESULT AND DISCUSSION

Serendipity is a difficult concept that involve emotional dimension. Every each of the user has different emotion and its’ depend on what they had received. Thus, it is a challenge to define serendipity in RS, to decide an item that referring to it and to questioned why an item is called serendipity, as the discovery of serendipity is rarely found [31].

From the recent researches, there is still no proper tools to evaluate serendipitous items. Most of the researchers measure the effectiveness of the system using precision for the system accuracy. They believe that the lower the accuracy, the higher the serendipitous item might be. Furthermore, De Gemmis et al. [24] uses emotion detection using Noldus FaceReader™ to measure the perception of users towards the serendipity item. Others evaluation approach are based on questionnaire provided to the user to indicate how serendipitous items might be to the user.

According to Lops et al. [7], an effective serendipity measurement need to be move beyond the conventional accuracy metrics in order to handle a level of emotional response that associate with serendipity. A new user centric directions to evaluate new emerging aspect in recommender system are required because there exist no essential tools to capture serendipity item. Furthermore, from the research infer that the adoption of strategies for realizing operational serendipity is one of the effective way to extend the capabilities of content-based recommender systems in order to overcome the overspecialization problem, by providing user with surprising suggestions.

### IV. CONCLUSION

To date there has been no formally accepted accurate definition for serendipity in recommendation system, and there has been no method based on actual data regarding user actual impression of selected items have been devised [22]. Thus based on our review, we opted the definition given by Kotkov et al. [15] saying that serendipity as unexpected, relevant and novel. Within the context of content-based recommender systems, we define serendipity as:

- *relevant* – items that are still relate or have some similarities with the user profile;
- *unexpected* – a positive response from users; and
- *novel* – referring as new items to users.

Kotkov et al. [15] further illustrate the framework of serendipity through an Euler diagram as illustrated in Fig 4.

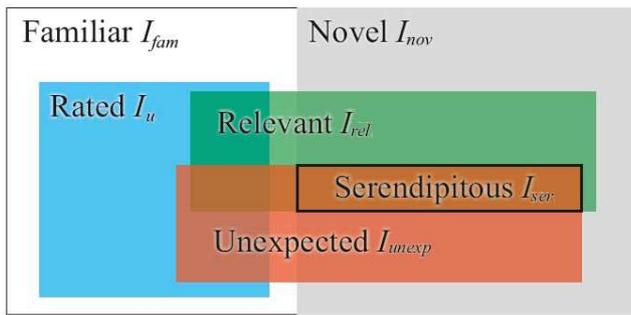


Fig 4: Euler Diagram of items from a user's point of view at a given moment of time (figure adapted from Kotkov et al. [15])

At a particular time, user's point of view of an item can be perceived as a set of intersections that consist of familiar, rated, novel, relevant, serendipitous and unexpected. According to Kotkov et al. [15], assumed that a recommender system contains items,  $I$  and a user,  $U$  that is familiar with items  $I_{fam}$ .  $I_{nov}$  denoted the items that user never heard or found in past,  $I_u$  denoted the rated item by user,  $I_{rel}$  refer to an item that user found interesting,  $I_{unexp}$  represents an item that differ from the items user rated and can be either novel or familiar to the user and  $I_{ser}$  is the inverse of familiarity. Referring to the diagram, they conclude that the task of serendipity-oriented RS is to suggest user with serendipitous items, novel items that user might like and unexpected,  $I_{ser} = I_{nov} + I_{rel} + I_{unexp}$ .

To reflect with the definition of content-based recommender systems, familiar items  $I_{fam}$ , represent of user past activities that consist of rated items  $I_u$  and relevant items  $I_{rel}$ . Information of user past activities on items can be used to build the user profile. Furthermore, familiar items,  $I_{fam}$  can be referred to as items that users are familiar with. In this regards, for movies and books recommendation  $I_{fam}$  may refer to items with the same genre. We consider relevant items  $I_{rel}$ , as items that are related to the target user which are similar with the user profile. For example, in movie recommendation, users may love to receive recommendation of movies that comes from the same director or movies involving similar actors.

In the research by Kotkov et al. [33], they found that most types of serendipity components help to broaden user preferences. For now, the research only available for serendipitous recommendation in collaborative filtering. They distributed questionnaire to real users to address the meaning of serendipity. According to their results, different variation of unexpectedness and different kind of serendipity bring different effects on preference broadening and user satisfaction. Thus, in our research, we planned to use these dataset, to validate the serendipity definition.

In the context of content-based approach, we define serendipity as items that are not familiar to users, but relevant and novel to them and nevertheless least expected. The items can be recommended either from popular or non-popular categories, but it need to be different and bring surprise elements to the user. Based on Fig. 4, in some rare

cases, popular items can also be serendipitous and novel to users.

In this paper, we have formally defined the concept of serendipity in recommender systems in order to overcome the issue of over-specialisation with special emphasis on content-based recommendation. Our future works include exploring new approaches in integrating serendipity in content-based recommender systems based on the definition of serendipity given in this study. We particularly interested on looking into the role of LOD in supporting serendipitous recommendations. The LOD has shown to be beneficial in information retrieval [32], but its capability to support serendipitous recommendation is yet to be experimented. Serendipitous items prioritize the novelty and unexpected value in recommending an item to the user.

#### ACKNOWLEDGMENT

This research was funded under a research grant, DCP-2017-007/4.

#### REFERENCES

- [1] Malaysian Communication and Multimedia Commission. (2017) Internet Users Survey 2017. [Online]. Available: <https://www.mcmc.gov.my/skmmgovmy/files/17/17b084f3-cdc1-4b42-ab15-2b92aa6c0170/files/assets/basic-html/page-1.html>
- [2] T. Di Noia and V. C. Ostuni, "Recommender Systems and Linked Open Data," in *Reasoning Web. Web Logic Rules. Lecture Notes in Computer Science*, 2015, vol. 9203, p. 88–113.
- [3] A. S. Ghabayen and S. A. M. Noah, "Using Tags for Measuring the Semantic Similarity of Users to Enhance Collaborative Filtering Recommender Systems," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 6, pp. 2063–2070, 2017.
- [4] S. Shishhechi, S. Y. Banihashem, N. A. Mat Zin, and S. A. M. Noah, "Ontological approach in knowledge based recommender system to develop the quality of e-learning system," *Aust. J. Basic Appl. Sci.*, vol. 6, no. 2, pp. 115–123, 2012.
- [5] S. Shishhechi, S. Y. Banihashem, N. A. Mat Zin, and S. A. M. Noah, "Learning Content Recommendation for Visual Basic . Net Programming Language based on Ontology," *J. Comput. Sci.*, vol. 7, no. 2, pp. 188–196, 2011.
- [6] R. Ramli, S. A. M. Noah, and M. M. Yusof, "Ontological-based model for human resource decision support system (HRDSS)," in *9th Confed. Int. Workshops and Posters on On the Move to Meaningful Internet Systems, OTM 2010*, 2010, pp. 585–594.
- [7] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based Recommender Systems: State of the Art and Trends," in *Recommender Systems Handbook*, New York: Springer, 2011.
- [8] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender systems: an introduction*, New York: Cambridge University Press, 2011.
- [9] C. C. Aggarwal, *Recommender Systems*, Switzerland: Springer, 2016.
- [10] A. Felfernig, M. Jeran, G. Ninaus, F. Reinfrank, S. Reiterer, and M. Stettinger, *Basic Approaches in Recommendation Systems. Recommendation System in Software Engineering*, Berlin, Heidelberg: Springer, 2014.
- [11] P. Adamopoulos and A. Tuzhilin, "On unexpectedness in recommender systems: Or how to expect the unexpected," in *CEUR Workshop Proc.*, 2011, pp. 11–18.
- [12] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: a Survey of the State of the Art and Possible Extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [13] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, UK: Penguin Groups, 2011.
- [14] V. Maccatrozzo, M. Terstall, L. Aroyo, and G. Schreiber, "SIRUP: Serendipity in Recommendations via User Perceptions," in *Proc. 22nd Int. Conf. Intell. User Interfaces - IUI '17*, 2017, pp. 35–44.
- [15] D. Kotkov, S. Wang, and J. Veijalainen, "A survey of serendipity in recommender systems," *Knowledge-Based Syst.*, vol. 111, pp. 180–192, Nov. 2016.

- [16] E. G. Toms, "Understanding and facilitating the browsing of electronic text," *Int. J. Hum. Comput. Stud.*, vol. 52, no. 3, pp. 423-452, 2000.
- [17] A. E. Foster and N. Ford, "Serendipity and Information Seeking: an Empirical Study," *Journal of Documentation*, vol. 59, no. 3, pp.321-340, 2003.
- [18] (2018) Cambridge Dictionary. [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/serendipity>.
- [19] P. V. Anel, "Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability" *The British Journal for the Philosophy of Science*, vol. 45, no. 2, pp. 631-648, 1994.
- [20] A. Maksai, F. Garcin, and B. Faltings, "Predicting Online Performance of News Recommender Systems Through Richer Evaluation Metrics," in *Proc. 9th ACM Conf. Recomm. Syst. - RecSys '15*, 2015, pp. 179-186.
- [21] P. Adamopoulos and A. Tuzhilin, "On over-specialization and concentration bias of recommendations," in *Proc. 8th ACM Conf. Recomm. Syst. - RecSys '14*, 2014 pp. 153-160.
- [22] T. Akiyama, K. Obara, and M. Tanizaki, "Proposal and evaluation of serendipitous recommendation method using general unexpectedness," in *Workshop on Practical Use of Recommender Systems, Algorithms, and Technologies*, 2010, vol. 676, pp. 3-10.
- [23] L. Iaquinta, M. De Gemmis, P. Lops, G. Semeraro, M. Filannino and P. Molino, "Introducing serendipity in a content-based recommender system," in *Proceedings of the 2008<sup>th</sup> 8<sup>th</sup> International Conference on Hybrid Intelligence Systems, HIS'08*, 2008, pp. 168-173.
- [24] M. De Gemmis, P. Lops, G. Semeraro, and C. Musto, "An investigation on the serendipity problem in recommender systems," *Inf. Process. Manag.*, vol. 51, no. 5, pp. 695-717, Sept. 2015.
- [25] M. Jenders, T. Lindhauer, G. Kasneci, R. Krestel, and F. Naumann, *A serendipity model for news recommendation*, Lecture Notes in Computer Science. Cham: Springer, 2015, vol. 9324.
- [26] S. M. McNee, J. Riedl, and J. a Konstan, "Being accurate is not enough: how accuracy metrics have hurt recommender systems," in *CHI'06 Ext. Abstr. Hum. factors Comput. Syst.*, 2006, pp. 1097-1101.
- [27] E. G. Toms, "Serendipitous Information Retrieval," in *DELOS, Workshop: Information Seeking, Searching and Querying in Digital Libraries*, 2000, pp. 17-20.
- [28] K. Sugiyama and M.-Y. Kan, "Towards higher relevance and serendipity in scholarly paper recommendation," in *ACM SIGWEB Newsletter*, 2015, pp. 1-16.
- [29] A. Said, B. Fields, B. J. Jain, and S. Albayrak, "User-centric Evaluation of a K-furthest Neighbor Collaborative Filtering Recommender Algorithm," in *Proc. 2013 Conf. Comput. Support. Coop. Work*, 2013, pp. 1399-1408.
- [30] P. André, M. C. Schraefel, J. Teevan, and S. T. Dumais, "Discovery is never by chance: Designing for (Un)Serendipity," in *Proceeding seventh ACM Conf. Creat. Cogn. - C&C '09*, 2009, pp. 305-314.
- [31] Y. I. A. Mohd Khalid, S. A. M. Noah, "Semantic text-based image retrieval with multi-modality ontology and DBpedia," *The Electronic Library*, vol. 35, no. 6, pp. 1191-1214, 2017.
- [32] F. Ramli, S. A. M. Noah, "Building an event ontology for historical domain to support semantic document retrieval." *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, p. 1154-1160, 2016.
- [33] D. Kotkov, J. A. Konstan, Q. Zhao and J. Veijalainen, "Investigating Serendipity in Recommender System Based on Real User Feedback," in *SAC '18 Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 1341-1350.
- [34] (2018) The ontotext webpage on Linked Open Data. [Online] Available: <https://ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>