

A New Information Theory-Based Serendipitous Algorithm Design

Xiaosong Zhou¹, Zhan Xu¹, Xu Sun¹(✉), and Qingfeng Wang²

¹ Faculty of Science and Engineering,
University of Nottingham Ningbo China, Ningbo, China
Xu.sun@nottingham.edu.cn

² Business School, University of Nottingham Ningbo China, Ningbo, China

Abstract. The development of information technology has stimulated an increasing number of researchers to investigate how to provide serendipitous experience to users in the digital environment, especially in the fields of information research and recommendation systems. Although a number of achievements have been made in understanding the nature of serendipity in the context of information research, few of these achievements have been employed in the design of information systems. This paper proposes a new serendipitous recommendation algorithm based on previous empirical studies by taking into considerations of the three important elements of serendipity, namely “unexpectedness”, “insight” and “value”. We consider our design of the algorithm as an important attempt to bridge the research fruits between the two areas of information research and recommendation systems. By applying the designed algorithm to a game-based application in a real life experiment with target users, we have found that comparing to the conventional designed method; the proposed algorithm has successfully provided more possibilities to the participants to experience serendipitous encountering.

Keywords: Serendipity · Recommendation system · Information theory

1 Introduction

Serendipity is widely experienced in human history, it is defined as “an unexpected experience prompted by an individual’s valuable interaction with ideas, information, objects, or phenomena” [1]. So far studies relating to serendipity mainly focus on the following two directions: theoretical studies in the area of information research which aim to investigate the nature of serendipity [2–4], and the empirical studies with the purpose to develop applications or algorithms that provide users with serendipitous encountering especially in the digital environment [5–7].

One of the areas which try to employ serendipity applications is the design of recommender system. The overloaded information in the cyber space has made current users no longer satisfied by recommending them those “accurate” information, instead, users aims to be recommended with the information that are more serendipitous and interesting to them [8–10]. However, a rising concern identified in our reviewing of relevant studies is that those discoveries from information research regarding the nature

of serendipity do not receive sufficient attentions in the recommender system designs. This paper proposes a new algorithm to support serendipitous recommendation by applying recent research fruits on serendipity in the area of information research.

2 Problem and Research Question

Recommender system researchers often consider serendipity as “unexpected” and “useful” [11], and have designed recommendation algorithms through either content-based filtering [12] or collaborative filtering [13]. However, most of the recommendation algorithms mainly focus on providing “unexpectedness” to the users, and treated the “usefulness” as only a metric value to measure the effectiveness of their algorithms rather than considering it as a design clue [14].

As a comparison, serendipity in information research is often considered with three main characteristics: unexpectedness, insight and value [4]. “Unexpectedness” is considered as the encountered information should be unexpected or a surprise to the information actor, while “value” specifies that the encountered information should be considered as useful and beneficial to the information actor. These two understandings of “unexpectedness” and “value” consist with the current view of serendipity in designing recommender systems [11, 14]; however, the “insight” aspect tends to be neglected.

“Insight” is considered as an ability to find some clue in current environment, then “making connections” between the clue and one’s previous knowledge or experience, and finally shift the attention to the new discovered clue [15]. Some researchers have found such ability of “making connections” is actually a key facet in experiencing serendipity [4] and can be quite different among individuals and result in a range of serendipity encounterers from the super-encounterers to occasional-encounterers [16]. The connections can be made between different pieces of information, people and ideas [3]; therefore, to support or “trigger” connection-making in order to bring more possibilities of experiencing serendipity have always been considered as an important design clue for those information researchers [17, 18].

Based on the discussed issues, we then raise our research question: is it possible to combine the theoretical studies of serendipity in information research, especially the ignored aspect of “insight” or “making connection”, into the recommender system design?

Followed by our research question, we proposed a collaborative-filtering based algorithm by considering the theoretical discoveries of serendipity from the area of information research. Based on the discovery from information research that serendipity is often encountered in a relaxed and leisure personal state [1, 3], we then applied the algorithm into a game based application and conducted an empirical experiment.

3 Proposed Algorithm

There are two major concerns in providing serendipitous encountering in the recommendation system design: the first concern is how to balance “unexpectedness” and “useful”. As pointed out by [14], there should be “a most preferred distance” between

the two values, as the high level of unexpectedness may cause user’s dissatisfaction of the recommended information, while users may also lose interest to that information with a low unexpectedness. The second concern is how to combine “insight” into system design to stimulate the process of “making connections”.

The two concerns are addressed from the following perspective of “relevance” with two hypotheses:

- Hypothesis 1: Given the information that is highly relevant to a user’s personal profile, the information would also of a high potential value to the user;
- Hypothesis 2: A user will be unexpected to the information that is relevant to his profile while is not previous acknowledged or known by the user.

Consider a target user A, who is the user that will be provided with the recommended information, a user B who is highly relevant to user A and a user C who is highly relevant to user B while is not known by user A. The user A may experience serendipity by providing the information of user C, which is unexpected to him/her, and by providing the relationship between user B and user C, which may further cause interestingness or usefulness to user A. The following part of this section illustrates a detailed implementation of the algorithm.

1. Target user

Consider a table of a target user profile U_j with a category set $C = \{C_1, C_2, C_3...C_i...C_n\}$, where C_i represents the i -th category of the user profile. All the categories are arranged through the value of their weights in the user profile. The weight can either be a given weight by the dataset or calculated through clustering analysis [19]. In order to simplify the introduction of our proposed algorithm here, it is more convenient to set the weight for each C_i which is given by the dataset in the very beginning. The weight of C_i is larger than C_j ($i > j$) in C set:

$$w_c = \{w_{C_1}, w_{C_2}, \dots, w_{C_i}, \dots, w_{C_j}, \dots, w_{C_n} | w_{C_i} \geq w_{C_j}, i > j\} \tag{1}$$

For each category set C_i , consider $C_i = \{a_1, a_2, a_3... a_i ... a_n\}$, where a_i is the corresponded attribute to each vector C_i . In particular, for each a_i represents the dimension according to which a new user profile may be produced (i.e. author of literatures; musicians). The values for each a_i are also arranged by their weight in each vector C_i and can be calculated through semantic analysis such as the *tf*idf* weight (term-frequency times inverse document frequency) calculation [20]:

$$w(t, d) = \frac{tf_{t,d} \log\left(\frac{N}{df_t}\right)}{\sqrt{\sum_i (tf_{i,d})^2 \log\left(\frac{N}{df_i}\right)^2}} \tag{2}$$

Where $w(t,d)$ represents for the weight of a term t in a document d , and it is a function of the frequency of t in the document (tf_t,d), the number of documents that

contain the term (*dft*) and the number of documents in the collection (*N*). As a result, the weight for a category set C_i is determined by the weight of each attribute in the set:

$$w_{C_i} = \{w_{a_1}, w_{a_2}, \dots, w_{a_i}, \dots, w_{a_j}, \dots, w_{a_n} | w_{a_i} \geq w_{a_j}, i > j\} \tag{3}$$

2. *Screen the weight*

As been pre-defined that C_I with the largest weight in the C set and a_I with the largest weight in the C_i set. Set a threshold τ to eliminate the low weight value from the user profile U_I :

$$w_{C_i} = \{w_{a_1}, w_{a_2}, \dots, w_{a_i}, \dots, w_{a_j}, \dots, w_{a_n} | w_{a_i} \geq w_{a_j}, i > j\} \tag{4}$$

Similarly, set a threshold θ to eliminate the low weight value from the C_i set:

$$w_{C_i} = \{w_{C_i a_1}, w_{C_i a_2}, w_{C_i a_3}, \dots, w_{C_i a_i} | w_{C_i a_i} \geq \theta\} \tag{5}$$

3. *Generate a new user profile*

A new user profile U_{i+1} is produced according to each a_i in the C_i set. Here, the generation of the user profile arranges from the largest weight of $w_{C_i a_1}$ to the smallest weight of $w_{C_i a_i}$.

4. *Iteration and End condition*

Based on the weight arrangement in a user profile, it is intuitional that for an attribute a_i with a large weight, it is also with more possibility for the current user to have acknowledged about the information of a_i . In other words, the probability for a current user U_i to make connection with the next user profile U_{i+1} is proportional to the weight of the attribute in current user profile:

$$P(U_{i+1}|U_i) = \lambda w_{C_i} * w_{C_i a_i} \tag{6}$$

where λ is the proportionality coefficient of the probability to the relevant weight.

The probability of making connections by target user U_I to i -th user can be further extended if only the generated user is always new to the prior generated ones:

$$P(U_i|U_1) = P(U_2|U_1) * P(U_3|U_2) * \dots * P(U_i|U_{i-1}) \tag{7}$$

The iteration to find the next user would not continue until it meets the following two end conditions:

- the generated user is no longer new to all the previous generated users;
- $P(U_i|U_1)$ comes to a threshold δ , where δ represents an appropriate threshold of the probability.

The reason to set the threshold δ is to ensure the effectiveness of the iteration process. This is because if $P(U_i|U_1)$ comes too large, the recommended information may fail to bring the target user with the sense of unexpectedness, as the recommendation may probably have been acknowledged by the user; however, if the value of $P(U_i|U_1)$ is too small, the recommended information may be too irrelevant to the target user and he/she may lose interest on it. Hence the setting of the threshold δ is a very important step for the iteration process and it needs to be further identified based on empirical studies in the future. Once the recommendation list is generated within the threshold δ , they can be recommended to the target user by selecting the item with the highest values of $P(U_i|U_1)$.

5. Recommendation

When the iteration is finished, the content with the largest weighted category in current candidate will be provided to the target user, in addition with the relevant information of the previous searched users that result in the current user.

6. An example of the proposed algorithm

An example of the proposed algorithm is provided in Fig. 1. Consider *Ann* as the target user (U_1) with different literature categories of $\{A, B, C\}$ in her personal library, whose weight is $\{0.5, 0.3, 0.2\}$ (Fig. 1a). The author names of the literatures are set as the attributes for each category and according to the *tf*idf* weight calculation, there are three values $\{a_1, a_2, a_3\}$ in category A with the weight $W^A = \{0.6, 0.3, 0.1\}$. Set $\lambda = 1$ for each probability of the current user to find the next user profile, the probability for *Ann* to find *al*'s profile (U_2) can be calculated according to Eq. (6):

$$P(U_2|U_1) = w_A * w_{A,a_1} = 0.5 * 0.6 = 0.3 \quad (8)$$

The profile of *al* is then produced as Fig. 1b. Likewise, among the four authors in the D category, author *dl* (U_3) weights largest and then produce *dl*'s profile (Fig. 1c):

$$P(U_3|U_2) = w_D * w_{D,d_1} = 0.4 * 0.5 = 0.2 \quad (9)$$

According to Eq. (7), the probability for *Ann* (U_1) to find *dl*'s profile (U_3) is:

$$P(U_3|U_1) = P(U_2|U_1) * P(U_3|U_2) = 0.3 * 0.2 = 0.06 \quad (10)$$

Set the threshold δ as 0.06, then the iteration of the algorithm stops and recommend literatures of category F in *dl*'s profile to *Ann*, in addition with the relevant information of *dl* and *al*. For example, the recommended information can be "these papers (category F) are most stored by *dl*, who had published papers (*d1, d2, d3, d4*) with *al* before".

7. Description of the Proposed Algorithm

The proposed algorithm is collaborative filtering based, hence it is more appropriate to those dataset whose content is generated by different users, according to which the next user's profile will be easier to produce for a current user.

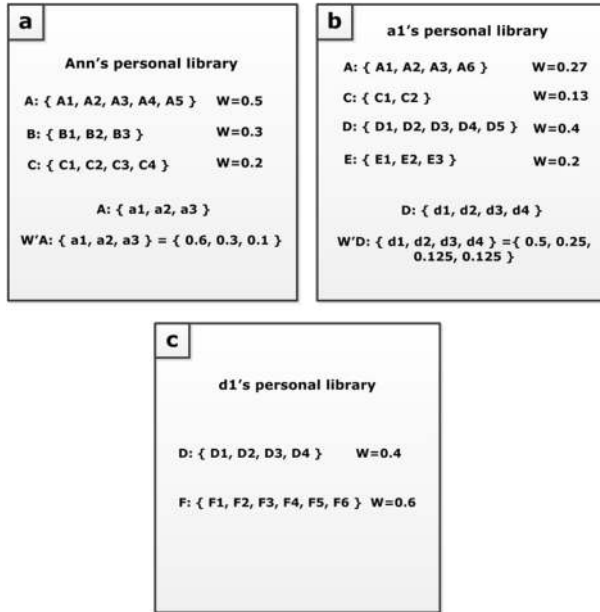


Fig. 1. An example of the proposed algorithm: (a) target user *Ann*'s personal library; (b) user *a1*'s personal library generated by *Ann*; (c) user *d1*'s personal library generated by *a1*

The proposed algorithm relates with serendipity from the following three aspects:

- *Unexpectedness*: by setting the value of probability. In an identified threshold δ , the unexpectedness of the information to a target user is inversely related to the magnitude of probability. The smaller probability for a target user to find another user, the more unexpectedness he/she receives from the provided information of the current candidate.
- *Insight*: by providing the information of the searched clues which demonstrates the relationship between the provided user (recommendation source) and the target user. As aforementioned that the ability to connect the new clue with previous knowledge/experience is a key element in the occurrence of serendipity, and thus there is a necessity for the designers to provide the design clues can contribute to a customer's noticeability or attention to connect the provided information with his/her personal profile. In the provided example of Fig. 1, such insight is provided by showing the relationship between *d1* and the target user, who had published paper together before.
- *Value*: by generating the next user's file according to the weight arrangement of the attributes; those with larger weights are considered as priorities. This is because the larger weight the attribute is, the more possibility it may have to satisfy the target user's need/concern, and finally brings more potential value to the user.



Fig. 2. Different stages of the designed sketch application: (a) Memorised picture; (b) Participant’s sketching; (c) Retrieving; (d) Sketching result and game score; (e) provided picture information

4 Empirical Study

A problem that the developed algorithm confronted is how to evaluate it successfully in a real life environment. According to the information research, studying serendipity in a controlled experiment always has negatively influences on the participants [21, 22]; in addition, serendipity is such a subjective phenomenon that it is tightly closed to the participant’s own experience or knowledge [4, 15]. A hint to address the problem may rise from Shute’s [23] stealth assessment theory where the assessments or inferences of conceptions or models that is elusive to humans is embedded into new computer-based technologies such as games. In the centre of Shute’s theory is the Evident-Centred Design (ECD), where a player’s abilities and understandings, especially those that cannot be directly observed by researchers (e.g. critical thinking, problem solving) is reflected through the embedded tasks or situations in the design, such as the interaction

processes of the game. Serendipity is exactly such a phenomenon that cannot be observed directly by the researchers; however, during the process of game-playing, participants would naturally produce sequences of actions while performing the designed tasks and hence provides us with possible evidences to access the encounter of serendipity. In addition, there is also evidence from the information research that serendipity is often experienced by those participants who are in a relaxed and leisure state [1, 3], and playing games can bring participant to such a relaxed state comparing with other activities. Based on the above discussion, we have then employed the algorithm into a game-based application and have conducted an empirical experiment to investigate whether our proposed algorithm could provide serendipitous encountering to researchers. The study is described in details below:

4.1 Participant

28 PhD students (14 males and 14 females) from different disciplines are invited to the study. They were asked to conduct a drawing game on a mobile application which was developed by the research group.

4.2 Game-Based Application

The developed game is an android-based drawing game, which involves the following stages:

- Memorising and sketching

Each participant was given a picture in the very beginning for observation. Participant was then asked to layout the colour features of the picture based on the memory. A time clock is set during this stage where the maximum observation time is 30 s and the maximum sketch time for each participant is 120 s.

- Retrieving

When a participant finishes sketching, a group of 30 images is displayed to the participant for retrieving whether or not his/her drawing picture was contained in the provided pictures. If the picture is contained in the group, he/she may click on the picture to pick it out. Or the participant only needs to click “Next” button.

- Sketching result

Participant’s final sketching result is provided after retrieving. A winning game means the participants has successfully retrieved the drawing picture, and then he/she will be given a game score based on the observation time and sketching time. Otherwise, the participant will be noticed he/she has failed in the sketching.

- Providing picture information

The last stage of the application provides participants the related information of the picture, in regardless of whether or not the participant has made a successful sketching.

4.3 Embedded Algorithm and Comparison

– *Embed proposed algorithm into the developed application*

The next step is to embed the proposed algorithm into the application. As all the participants are PhD researchers, the algorithm is designed based on three assumptions:

- Assumption 1: For each PhD student, their supervisor’s information is a large weight attribute in their personal profile.
- Assumption 2: For each PhD’s supervisor, the co-author from their publications is a large weight attribute in the supervisor’s profile.
- Assumption 3: For each co-author’s personal profile, their working institution is a large weight attribute.

Therefore, each PhD student supposed to be provided with the information of their supervisor’s co-author’s institution. Figure 3 shows the design of the study including how the proposed algorithm is embedded into the game-based application and the sketch game process. Based on each participant’s information, we start our study by providing them with the pictures which show the institution badge (Fig. 2a). Each participant was then asked to draw the picture out within 120 s (Fig. 2b). After retrieving (Fig. 2c) and showing the result of the sketch (Fig. 2d), the serendipitous information to the pictures was provided to the participant (Fig. 2e). The given information related to the picture includes two levels: (1) the introduction of the institution; (2) the publications of both the participant’s supervisor and the co-author, as is shown in Fig. 5a.

– *A comparison of the proposed algorithm*

As a comparison, each participant was also given the pictures that without the serendipitous information from our proposed algorithm (Fig. 4). Two cover pictures from the “Nature” website (www.nature.com) were selected to the participant, together with the description of the picture on the website (Fig. 5b). We consider such provision as the conventional way to introduce the relationship between the picture and the information (pic-and-info). As a result, each participant should draw two pictures with our proposed algorithm and two with the conventional way of “pic-and-info”.

4.4 Evaluation

The traditional measurement of serendipity in the area of recommender systems is often based upon the conventional perception on serendipity, where it is considered with the two main characters of “unexpected” and “useful” [11, 14, 24]. However, information research on serendipity has found that an important characteristic of serendipity is the element of time. The considered serendipity at a certain time may be changed with time [3, 4]. Therefore, [17] argued that apart from “unexpectedness” and “useful”, “interesting” and “relevant” should also be taken into considerations as new measurements of serendipity, this is because their studies have shown that users may keep or follow up the information that is “relevant” or “interest” to them and lead to serendipity at a

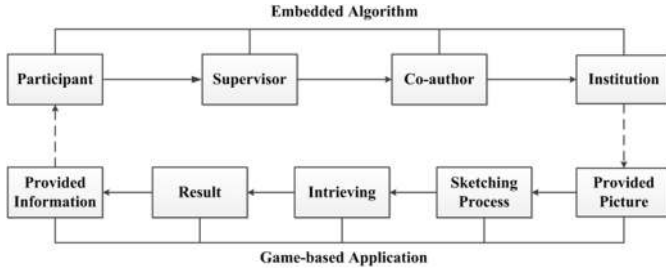


Fig. 3. Process of the study and the embedded proposed algorithm

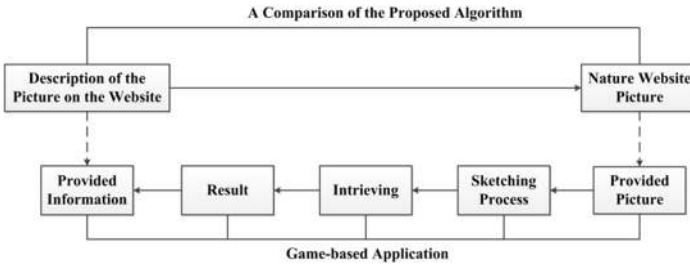


Fig. 4. A comparison of the proposed algorithm

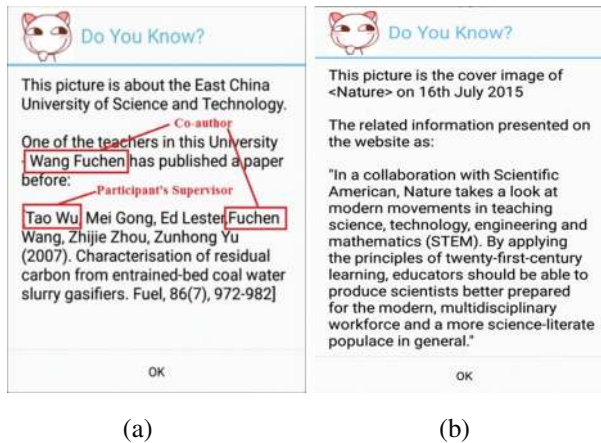


Fig. 5. Provided information: (a) designed algorithm; (b) information from the nature website

different time. They consider such serendipity as “pseudo-serendipity” which refers to “encounters experienced by users that have the potential of being serendipity in that users intended to do something in the future with those encounters” [17, 25].

In this paper, we tend to argue that both “pseudo-serendipity” and “serendipity” would happen in recommendation systems. This is because in some cases, whether or

not the recommended information is “useful” or “beneficial” to the participant needs to be further identified, and such identification may probably start from “interesting” or “relevant” [17].

Therefore, the evaluation on serendipity in our empirical study is also identified from the four dimensions of “unexpected”, “interesting”, “relevant” and “beneficial”. After a participant finished sketching all the pictures, he/she was given a questionnaire with the four dimensions, and with each dimension a Likert scale from one represents “not at all” to five represents “extremely”. Participant needs to rate in the questionnaire based on their experience of the whole sketching process from the four dimensions.

In addition, a 15 min post-interview is carried out right after each participant finished their sketching. The interview explored participant’s subjective experience and the further reason for their ratings of the four dimensions.

4.5 Results

1. Questionnaire

In total, 20 effective questionnaires were picked out from the 28 participants, as the other eight participants were too concentrated in the gameplay and failed to read the related information of the picture. These questionnaires were the feedbacks of 40 pictures of the conventional way of “pic-and-info” and the other 40 pictures based on our designed algorithm.

Only the marks of four or five are considered to be effective values on the corresponding dimension, which is shown in Fig. 6. According to the four identified dimensions of unexpected, interested, related and beneficial, it is obvious that comparing with the conventional way of “pic-and-info”, our designed algorithm is more possible to result in participant’s serendipitous encountering.

2. Interview

During the interviews, most participants reported their senses of serendipity relating to the serendipitous algorithm designed pictures from the following two perspectives:

- All the participants reported that they had experienced “unexpectedness” because of the relationship between the picture and the provided information:

I’ve never thought the picture is related to my supervisor! I’ve just taken it as a drawing game..... The information in the end really surprised me and I really think this is a very good design to provide me with the information in such a context! (Participant 3)

In addition, 12 out of 20 participants reported another level of unexpectedness existing in the content of the information, as the provided information was previously unacknowledged to them:

I never know that my supervisor had published such a paper with him (the co-author) before..... I’m interested about it and will check the details of the paper later. (Participant 10)

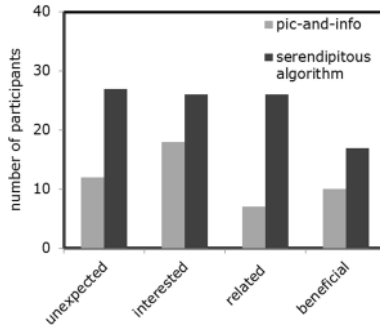


Fig. 6. Questionnaire result

- The result of the sketching game:

Over eight participants expressed their requirements to add an external link of the presented information (e.g. published paper of ...). One participant even asked us to send him the detailed information after the study.

By contrast, most participants have reported a less interest in the conventional “pic-and-info”, this also reflects the important role of “relatedness” played in the design of the algorithms. As a result, the feedbacks from the participants have demonstrated that our proposed algorithm can effectively support the design strategies for serendipity.

5 Conclusion and Future Work

In this paper, we have presented a new serendipitous recommendation algorithm by combining the theory of serendipity in information research. In particular, our proposal extended the design of such serendipitous recommendation by including two other vital aspects in serendipity, namely, “insight” and “value”.

We also performed an empirical experiment with target users by employing the proposed algorithm to a game-based application. The result demonstrates that comparing with the conventional design of “pic-and-info”, our algorithm has effectively encouraged our participants to experience serendipitous encountering. However, the study is limited by the small sample number of participants, so our future work will aim to explore the algorithm through more datasets, and to investigate the appropriate thresholds (e.g. τ , θ , δ) which have been set in current algorithms. We will also compare our proposed algorithm with other existing algorithms so as to better evaluate and optimize the algorithm in different situations.

Acknowledgments. We acknowledge the financial support from a NSFC grant with code 71401085.

References

1. McCay-Peet, L., Toms, E.G.: Investigating serendipity: how it unfolds and what may influence it. *J. Assoc. Inf. Sci. Technol.* **66**(7), 1463–1476 (2015)
2. Erdelez, S.: Investigation of information encountering in the controlled research environment. *Inf. Process. Manage.* **40**(6), 1013–1025 (2004)
3. Sun, X., Sharples, S., Makri, S.: A user-centred mobile diary study approach to understanding serendipity in information research. *Inf. Res.* **16**(3) (2011)
4. Makri, S., Blandford, A.: Coming across information serendipitously – Part 1. *J. Documentation* **68**(5), 684–705 (2012)
5. Iaquina, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., Molino, P.: Introducing serendipity in a content-based recommender system. In: *Proceedings of the 8th International Conference on Hybrid Intelligent Systems, HIS 2008* (2008)
6. Yamaba, H., Tanoue, M., Takatsuka, K., Okazaki, N., Tomita, S.: On a Serendipity-oriented recommender system based on folksonomy and its evaluation. *Procedia Comput. Sci.* **22**, 276–284 (2013)
7. Makri, S., Blandford, A., Woods, M., Sharples, S., Maxwell, D.: Making my own luck: serendipity strategies and how to support them in digital information environments. *J. Assoc. Inf. Sci. Technol.* **65**(11), 2179–2194 (2014)
8. Lu, Q., Chen, T., Zhang, W., Yang, D., Yu, Y.: Serendipitous personalized ranking for Top-N recommendation. In: *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 258–265, IEEE Computer Society (2012)
9. Sun, T., Zhang, M., Mei, Q.: Unexpected relevance: an empirical study of serendipity in retweets. In: *ICWSM*, pp. 592–601 (2013)
10. de Gemmis, M., Lops, P., Semeraro, G., Musto, C.: An investigation on the serendipity problem in recommender systems. *Inf. Process. Manage.* **51**(5), 695–717 (2015)
11. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: *Proceedings of the fourth ACM Conference on Recommender systems* (2010)
12. Jenders, M., Lindhauer, T., Kasneci, G., Krestel, R., Naumann, F.: A serendipity model for news recommendation. In: Hölldobler, S., Kröttsch, M., Peñaloza, R., Rudolph, S. (eds.) *KI 2015. LNCS (LNAI)*, vol. 9324, pp. 111–123. Springer, Cham (2015). doi:[10.1007/978-3-319-24489-1_9](https://doi.org/10.1007/978-3-319-24489-1_9)
13. Oku, K., Hattori, F.: Fusion-based recommender system for improving serendipity. In: *Proceedings of the Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011)*, 5th ACM International Conference on Recommender Systems (2011)
14. Adamopoulos, P., Tuzhilin, A.: On unexpectedness in recommender systems: or how to better expect the unexpected. *ACM Trans. Intell. Syst. Technol.* **5**(4), 1–32 (2014)
15. Rubin, V.L., Burkell, J., Quan-Haase, A.: Facets of serendipity in everyday chance encounters: a grounded theory approach to blog analysis. *Inf. Res.* **16**(3) (2011)
16. Erdelez, S.: Information encountering: a conceptual framework for accidental information discovery. In: *Proceedings of an International Conference on Information Seeking in Context*. Taylor Graham Publishing (1997)
17. Pontis, S., Kefalidou, G., Blandford, A., Forth, J., Makri, S., Sharples, S., Woods, M.: Academics’ responses to encountered information: context matters. *J. Assoc. Inf. Sci. Technol.* **67**(8), 1883–1903 (2016)
18. Kefalidou, G., Sharples, S.: Encouraging serendipity in research: designing technologies to support connection-making. *Int. J. Hum. Comput Stud.* **89**, 1–23 (2016)

19. Rohlf, F.J.: NTSYS-pc numerical taxonomy and multivariate analysis system, version 2.0. *Appl. Biostatistics*, 23 (1998)
20. Pazzani, Michael J., Billsus, D.: Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *The Adaptive Web*. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007). doi:[10.1007/978-3-540-72079-9_10](https://doi.org/10.1007/978-3-540-72079-9_10)
21. McCay-Peet, L., Toms, E.G.: Uses and gratifications: measuring the dimensions of serendipity in digital environments. *Inf. Res.* **16**(3) (2011)
22. Bogers, T., Rasmussen, R.R., Jensen, L.S.B.: Measuring serendipity in the lab: the effects of priming and monitoring. In: *Proceedings of the iConference 2013* (2013)
23. Shute, V.J.: Stealth assessment in computer-based games to support learning. *Comput. Games Instr.* **55**(2), 503–524 (2011)
24. Murakami, T., Mori, K., Orihara, R.: Metrics for evaluating the serendipity of recommendation lists. In: Satoh, K., Inokuchi, A., Nagao, K., Kawamura, T. (eds.) *JSAI 2007*. LNCS (LNAI), vol. 4914, pp. 40–46. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-78197-4_5](https://doi.org/10.1007/978-3-540-78197-4_5)
25. André, P., Schraefel, M.C.: Designing for (un)serendipity - computing and chance. *Biochem. Soc.* **31**(6), 19–22 (2009)