Article

A Quantitative Model to Evaluate Serendipity in Hypertext

Yuri Kim¹, Bin Han¹, Jihyun Kim², Jisoo Song¹, Seoyeon Kang¹ and Seongbin Park¹,*

¹ Department of Computer Science, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Korea; fhrml1004@korea.ac.kr (Y.K.); hanbin97@korea.ac.kr (B.H.); y7116385@korea.ac.kr (J.S.); ange0827@korea.ac.kr (S.K.)
² Samsung SDS, 125, Olympic-ro, 35-gil, Songpa-gu, Seoul 05510, Korea; herb1110@korea.ac.kr
* Correspondence: hyperspace@korea.ac.kr

Abstract: Serendipity is the phenomenon of people making unexpected and beneficial discoveries. While research on the mechanism and effectiveness of serendipity in information acquisition has been actively conducted, little attempt has been made to quantify serendipity when it occurs. In this paper, we present a quantitative model that measures serendipity experienced by users in a hypertext environment. In order to propose an evaluation model that measures how probable users would experience serendipitous moments in the process of an active search, we define a serendipitous discovery as an unexpected discovery that can happen during a sidetracked search. The proposed model consists of three parts: (a) pre-encountering—how early the user falls into the sidetracked search in the process of an active search; (b) post-encountering—the degree of interests of the entire process from the active search to obtaining unexpected information; and (c) discovery—the degree of the unexpectedness of the information obtained from the discovery. We evaluated the proposed model against examples with different structures and the potential serendipity values computed indicated the difference between the spaces in a meaningful way.

Keywords: serendipity; information acquisition; evaluation model

1. Introduction

Serendipity has played an important role in scientific discoveries from penicillin to X-rays, and has been attributed to the final push of these historic discoveries [1]. Serendipity, in the context of information acquisition, has also emerged as a highly important area. In the flood of information, serendipity has increasingly been recognized as useful to provide information environments that facilitate new ways of supporting passive information acquisition. Researchers have been making efforts to understand the slippery concept of serendipity—for example, ref. [2] defined it as “revealing unexpected connections between information when browsing”, ref. [3] stated “it is an information behavior, which incorporates “passive search” and “passive attention”, and ref. [4] defines the phenomenon as the “unexpected discovery of useful or interesting information … during the search for some other information”. The subjective nature of serendipity resulted in different, inconsistent definitions, but also terminology being used to define serendipity in information space [5]—for example, information encountering [6,7], accidental information discovery [8], incidental information acquisition (IIA) [9,10]. This issue probably made the algorithm of serendipity difficult to study and access, thus there is also no consensus on evaluation metrics to measure serendipity.

In this paper, we present an evaluation model for measuring the degree of potential exposure to serendipity in the context of information acquisition. It is an extension of [11] in which different factors that can influence serendipitous discoveries were investigated. For the evaluation metrics, we formalize the definition in [4] and use the term IIA for referring serendipitous information encounter in an information space. The information space refers to the Web space where a lot of hyperlinked Web pages exist. We focused on
assessing the relative degree of a user being exposed to serendipitous events in the course of searching for specific information in a given information space. For the evaluation model, we analyzed the process of serendipitous events and identify the essential factors for triggering serendipity. For each factor, we devised a separate metric, then combined them to complete the serendipity metric. To verify our evaluation program, we evaluated information spaces that are similarly designed to the sample models suggested in the study [12] where authors conducted a user study on information spaces with different recommendation algorithms to evaluate which model provides the most serendipitous environment. We observed whether our potential serendipity evaluation program achieves a similar result with the user study result.

This paper is organized as follows. We review related literature in Section 2, which is dedicated to defining metrics that are used for serendipity evaluation. In Section 3, we propose our serendipity evaluation metrics and describe strategies to assess serendipity. Section 4 discusses the utilization of our novel model. Finally, we reports the result of our serendipity evaluation model in Section 5.

2. Related Work

There have been many efforts to analyze serendipity to identify influencing factors. It is necessary to clarify the essential nature of serendipity in order to evaluate it. In this section, we mainly summarize commonly identified triggering factors that have been considered in serendipity-related studies.

Ref. [13] claims that that the encouraging factor is having a tight time limit with space for creativity. They perceived a longer length of time from exposure to the serendipitous environment to recognize that it may prevent a user from taking action in the new direction. Ref. [14] has pointed out that timing is one factor that may influence serendipity. They enshrined the power of “being in the right place at the right time”. Ref. [15] identified that timing was critical for serendipitous experiences in the course of conducting their field research. To better understand the process of serendipitous discovery, they observed serendipitous moments encountered during their project. The project had a limited schedule time; therefore, good timing was critical to associate with serendipitous discoveries. Although serendipity is an unpredictable event, many researchers agree that when it is observed is considered as a crucial factor in making the observer make temporal choices.

Even though there is no concrete definition of serendipity, the unexpected is a term that is needed for defining serendipity. According to [16], the primarily cognitive account of serendipity is recognizing the potential value in the unexpected. For example, the discovery of Helicobacter pylori bacteria in the stomach was unexpectedly observed, such that Waren was initially not looking for it when he found the bacteria during their work as a pathologist. Another example is the serendipitous discovery of penicillin by Alexander Fleming. He also discovered penicillin by accidentally contaminating one of their staphylococcus culture plates. Ref. [17] claimed that serendipity is a process of discovery that frequently ends in a way unpredictable from the perspective of their origin.

Table 1 shows a summary of related work to our research.

We could identify the explicit factor triggering a serendipitous moment, but a conceptual difficulty remains in quantitatively measuring serendipity. There has been no consistent form of quantitative measurement for serendipity. The factors related to timing and the user’s perception of events make it challenging to make measurements. Therefore, in this study, we address the following questions:

- Q1. How can we interpret the serendipity factors in light of the information acquisition?
- Q2. How can we build a quantitative serendipity measurement system with the serendipity factors?

Our work aimed to analyze the conducive circumstances needed for serendipity to occur and propose metrics that quantitatively measure the essential circumstances. We formulated a pre-encounter metric for measuring the timing factor and a post-encounter
and discovery metric to measure the factors related to the user’s perception. We will explain each metric in the following section.

Table 1. A summary of related work.

| Reference | The Core Factor of Serendipity |
|-----------|-------------------------------|
| [13]      |                               |
| [14]      | Timing                        |
| [15]      |                               |
| [16]      | Unexpectedness                |
| [17]      |                               |

3. Methods

In this section, we formalize the serendipity algorithm to extract serendipity evaluation metrics. Then, we propose an evaluation strategy with the metrics. All terminologies used are summarized in the Appendix A at the end.

3.1. Formalization of Serendipity

The terminology related to this study requires us to set up a vocabulary to be used throughout the paper. A user is generally understood to mean a person who uses the Web for information acquisition. The term active search refers to “information-seeking behavior which looks for specific information”. Target information is the “specific information that a user tries to find during active search”. A sidetracked search refers to “information-seeking behavior which incorporates passive attention, that involves finding unsought information”. The term search path refers to “a finite sequence of information pieces which a user encounters on the process of active search”. Similarly, a sidetracked path is “a finite sequence of information pieces which a user encounters on the process of sidetracked search”.

Figure 1 shows the structure of a serendipitous discovery.

In our term, we rephrased [6]’s definition into “a user experiencing IIA during the active search process for specific target information”. We investigated the defining features of serendipity and identified the components that determine whether a discovery is serendipitous. We divided the underlying mechanism into three components: pre-encountering, post-encountering, and discovery.

**Pre-encountering** While a user is performing an active search, the user should encounter a sidetracked path that leads to IIA;

**Post-encountering** After a user encounters sidetracked path, the user should start a sidetracked search;

**Discovery** The user should complete the sidetracked search until the user encounters IIA.
When a user tries to find certain information on the Web, the user might have to browse some of the Web pages that will eventually lead to the target information. However, other information might catch the user’s attention during the search process for the target information. This moment is considered a pre-encountering part which indicates that the user is experiencing a focus shift. After the focus shift, the post-encountering happens, if the user starts to browse other information in a new direction, instead of continuing the original search for target information. Finally, as a result of the exploration, if the user discovers useful or interesting information that the user was not originally looking for, and it is considered as a discovery.

Figure 2 shows a flow diagram of how a user may experience a serendipity.

![Flow diagram](image)

**Figure 2.** A flow diagram of how a user may experience a serendipity.

### 3.2. Evaluation Metrics

We propose an evaluation metric that measures serendipity in the context of information acquisition. To this end, we propose a component metrics that measures each component of the serendipity mechanism, such as pre-encountering, post-encountering and discovery. Then, we propose a serendipity metric that measures serendipity as a whole.
3.2.1. Pre-Encountering

The first metric reflects how probable it is that a user would explore a sidetracked path. We could not identify discernable patterns on users such that when they decide to explore sidetracked paths. However, we identified a reduction factor in user behavior that prevents users from exploring sidetracked paths. Most serendipitous events occurred in the early stages in the scope of activities, which means ‘timing’ plays a crucial role. We also identified that a user is more likely to feel fatigued or exhausted for longer search time. Accordingly, the probability of a user exploring a sidetracked path will decrease as it appears later in the search path. The metric has a single parameter [11]:

\[
pre(n) = \frac{1}{n}
\]

where \( n \) indicates the number of data that the user browsed until they experience pre-encountering.

For example, if the user found another Web page that triggers a focus shift after 4th browsing of Web pages (as illustrated in Figure 3), then \( n \) is calculated as 4. Therefore, the pre-encountering value is calculated as \( pre(4) = \frac{1}{4} \).

![Figure 3. An illustration of pre-encountering calculation process with simple example.](image)

3.2.2. Post-Encountering

The second metric reflects the likelihood that the user will continue to perform the sidetracked search until the user encounters IIA. Similarly to the first metric, we aimed to speculate on users’ behavior. We identified that if the information enables a user to have a pleasant experience in the search process, the user is more likely to stay and continue the exploration [18]. To measure this metric, we calculated the average interest value of the information that constitutes the sidetracked path. The metric has two parameters:

\[
post(m, p^{int}) = \text{Avg}(p^{int}) = \frac{\sum_{i=1}^{m} p^{int}_i}{m},
\]

where \( p^{int}_i \) is the interest value of the \( i \)-th information that the user encountered during the sidetracked search, and \( m \) indicates the total number of pages that the user discovered during the sidetracked search. For example, we suppose the sidetracked path and the interest value for each page constituting the sidetracked path exist as Figure 4.
Figure 4. An illustration of post-encountering calculation process with a simple example.

In this example, we calculate the post-encountering value as simply averaging the interest values: 
\[ \text{post}(6, \{0.65, 0.51, 0.18, 0.24, 0.47, 0.71\}) = 0.65 + 0.51 + 0.18 + 0.24 + 0.47 + 0.71 = 0.46. \]

For the precise evaluation, users’ own personal interest should be evaluated along with each datum’s popularity. However, it is not easy to personalize the interest value of each user. Therefore, we only evaluate the popularity of the page for the evaluation model. To measure the degree of interest of each datum, we apply the PageRank algorithm that gives an approximation of a page’s importance or quality [19]. In the algorithm, it is assumed that there is a “random surfer” who is given a Web page at random and keeps clicking on links without going back until the surfer gets bored of it. Then, the surfer starts the same behavior on another random page. PageRank measures how probable the random surfer visits a page. PageRank not only counts citations or backlinks to a given page but it is also normalized by the number of links on a page. PageRank is calculated using a simple iterative algorithm, and we iterate four times for measuring the interest value for each page. The PageRank of pageA is given as follows [19]:

\[
PR(A) = \frac{PR(T_1)}{C(T_1)} + ... + \frac{PR(T_n)}{C(T_n)}
\]

where \( T_1, \ldots, T_n \) are pages that establish links towards page A, and \( C(T_i) \) is the number of links going out of pages \( T_i \).

3.2.3. Discovery

The third metric reflects the core concept of serendipity: “unexpectedness” — the metric measures how dissimilar the IIA information to the original target information. The measurement is based on the taxonomic distance between the target and IIA information. The metric has two parameters [11]:

\[
dis(p_{TI}, p_{IIA}) = Tax\_dist(p_{TI}, p_{IIA})
\]

where \( p_{TI} \) indicates the page containing the target information, and \( p_{IIA} \) is the page with IIA information. The taxonomic distance is evaluated on the information taxonomy, where information space is hierarchically classified.

The information taxonomy consists of distinct pages for each level, and with each level down in classification, pages are split into more and more specific pages. The taxonomic distance is measured by counting the minimum number of jumps between two pages over the information taxonomy. Figure 5 illustrates the calculation process of the discovery value with a basic sample. In the given information taxonomy, suppose a user’s target information was on pageA. In the course of searching the pageA, if the user finds the pageB as a result of a sidetracked search, we calculate the taxonomic distance as explained step by step with arrows: 
\[ \text{dis}(\text{pageA}, \text{pageB}) = Tax\_dist(\text{pageA}, \text{pageB}) = 6. \]
3.2.4. Serendipity

The serendipity metric reflects how serendipitous the sidetracked search is. The metric consists of the component metrics introduced in Sections 3.2.1–3.2.3. The established formula is:

\[
\text{Serendipity}(p, p') = \text{pre}(n) \times \text{post}(m, p') \times \text{dis}(p_{TI}, p_{IIA}),
\]

that is a simple multiplication of the normalized values obtained from pre-encountering, post-encountering and discovery metrics. The normalization is applied to each value before the multiplication to combine the values from different component metrics. We eliminate the measurement units by applying feature scaling, which re-scales the range of data in [0, 1].

3.3. Evaluation Strategies

In this section, we propose an evaluation model that measures the potential serendipity value of information space with metrics introduced in Sections 3.2.1–3.2.4.

There are three phases for serendipity evaluation: input, process and output (see Figure 6).

3.3.1. Input

Input refers to the input value that is needed to provide for the evaluation. It involves the information space that is used in the assessment. The information space needs to be provided in Turtle (Terse RDF triple language) format. Turtle is a specific grammar of RDF that uses triples to represent information. RDF is a framework written in XML that describes resources on the Web. It is generally used for describing information about Web pages (e.g., contents, data information), content for search engines, or properties for shopping items. The triples include a subject, predicate and object [Figure 7a].

![Figure 7a](image)

(a) A triple defined in Turtle.

![Figure 7b](image)

(b) Page relationship defined in a triple.

Figure 7. The basic structure of triple relation.

Each part of the triple is separated with whitespace and terminated by “.” after each triple (e.g., :Engineering :SubField :Computer_Science). According to the Turtle syntax,
the relationship of pages can be defined as shown in Figure 7b. A hyperlink connecting two pages is represented with a predicate, and the pages are represented with a subject and an object, respectively. The process of transforming information taxonomy into a Turtle file is as follows:

1. Specify distinct keywords for each page;
2. Identify which page belongs to which level over the information taxonomy;
3. In a Turtle file, define a region for each level of the information taxonomy, from the 1-level to n-level (where n is the lowest level of the taxonomy) using # (the octothorpe) symbol;
4. For each level, insert what belongs to page triples.

With the keywords representing each page, which are used to express pages in the triple, it is needed to define the information taxonomy, including the pages’ levels and their relations [Figure 8].

3.3.2. Process

Process refers to the serendipity value evaluation process according to the given input data. First, the program finds all the existing search paths for every page residing in the information space. In Section 3.1, we defined the search path as “a finite sequence of information pieces which a user encounters during the process of searching specific target information”. The program visits every page on the information space, assuming that the page is the target information. For each visit, the program finds search paths that satisfy:

1. The search path is a sequence of pages \( P = (p_1, p_2, \ldots, p_n) \), such that \( p_i \) is adjacent to \( p_{i+1} \) for \( 1 \leq i < n \);
2. Two pages are adjacent when they are linked;
3. The search path always ends with \( P_{TI} (p_n = P_{TI}) \);
4. All the pages constituting the search path \( p_1, p_2, \ldots, p_n \) are linked to common domain page (DP) with \( P_{TI} \).

![Image](http://example.com)

Figure 8. An example of information taxonomy defined in Turtle.

With this definition, the program finds all the existing search paths from a source page to the \( P_{TI} \). The source page refers to the page that the user browses first during the search process. The possible source pages are restricted as the pages linked to the shared DP with
the $P_{TI}$. The $DP$ lies on one level higher in the information taxonomy. For example, if the $P_{TI}$ had four levels, then the $DP$ could be found as three-level. The program finds search paths from each source page to the $P_{TI}$ for all the possible source pages.

After the program finds the search paths, it finds all the sidetracked paths derived from each search path. Similarly to the search path, we define a sidetracked path as:

1. The sidetracked path is a sequence of pages $p' = p'_1, p'_2, \ldots, p'_m$, such that $p'_i$ is adjacent to $p'_{i+1}$ for $1 \leq i < m$;
2. The pages are adjacent when they are connected with a hyperlink;
3. The sidetracked path is derived from a search path, such that $p'_1 \subset p_1, p_2, \ldots, p_{n-1}$ (the page $p_n$ is excluded, since the user tends to quit the search process when they found the purposed information);
4. All the pages constituting a sidetracked path except for the last page are linked to common $DP$ with $P_{TI}$;
5. The sidetracked path always ends with the page $P_{IIA}$ ($p'_m = P_{IIA}$), where $P_{IIA}$ is the page linked to different $DP$ ($DP'$) with $P_{TI}$.

Figure 9 shows both a search path and a sidetracked path.

![Figure 9. A search path and a sidetracked path.](image)

For all (search path, sidetracked path) pairs on the information space, the serendipity value is measured with a metric proposed in Section 3.2.4. The program aggregates the measured values to compute the whole information space’s potential serendipity value as a final step.

3.3.3. Output

This section describes the output value produced by the serendipity evaluation model. After the evaluation process is completed, the program returns three values: number of search paths, number of sidetracked paths, and the potential serendipity value of the information space given as input data. There is no unit for the potential serendipity value. This means the result can only be used to compare the relative degree of serendipity among information taxonomies.

3.4. Use of Result

The outcome of the potential serendipity evaluation program is a console output containing the potential serendipity value of the given information space. Since the program proceeds with the evaluation on information space that is organized in the Turtle file, users can easily modify the structure of information space. Program users can exclude or include a hyperlink between pages by simply removing or adding a triple to the Turtle file. We recommend that the program users test differently structured information space by modifying it however they want and observe the impact of the changes in serendipity.
Experimental Verification

Ref. [12] conducted research that compared different recommendation algorithms to identify the most serendipitous recommendation algorithm. They evaluated equivalency-based algorithms (EQ) and diversity-based algorithms (DV) by a user study. The EQ was designed to display the most similar items to the user’s specified item. This algorithm allowed showing multiple items for the same category. On the contrary, the DV was designed to display only one item per category. Users were asked to rate the two algorithms along five dimensions: (1) unexpectedness; (2) interest; (3) novelty; (4) diversity; and (5) commonality among the results. The results of user ratings are shown in Table 2. The average of five measurements of the DV was higher than the EQ, implying that the DV is considered more serendipitous compared to the EQ.

Table 2. User ratings of recommendations from the two algorithms across the five measurement categories.

| Algorithm | Unexpectedness | Successfulness | Novelty | Diversity | Commonality | Average |
|-----------|----------------|----------------|---------|-----------|-------------|---------|
| DV        | 3.550          | 2.904          | 3.896   | 4.229     | 3.229       | 3.227   |
| EQ        | 2.091          | 3.614          | 2.559   | 2.457     | 4.500       | 2.855   |

Based on the study by [12], we built two different information spaces that are based on different link establishment algorithms. One algorithm focuses on the relevance between pages, while the other algorithm focuses on distinctiveness between pages. In this experiment, we will refer to the information space with the relevance-focused algorithm as “RV” and the information space with the distinctiveness-focused algorithm as “DT”. The information spaces are designed as follows:

1. We collected the data of copyright-expired books to compose the pages on information spaces. In total, there are five domains (e.g., “Adventure”, “Fantasy”, “Horror”, “Opera”, and “Travel”) and 75 pages under them (see Figure 10);
2. For each page, we generated a word set with an extractive summarization method [20];
3. Each page is linked to the top four relevant pages, where the degree of relevance between pages is measured as follows:
   We consider the two pages are more relevant as they have more common elements between their word sets. In this regard, we count the common terms between the pages’ word sets and consider the number as the degree of relevance between the pages. In this example, we supposed there exist pageA and pageB, and their word sets are $W_{\text{SET}}(\text{pageA})$ and $W_{\text{SET}}(\text{pageB})$. Therefore, in this example, the degree of relevance between pageA and pageB is 3:
   $W_{\text{SET}}(\text{pageA}) = \{\text{quote}, \text{point}, \text{throw}, \text{radical}, \text{lane}, \text{drama}, \text{premium}\}$
   $W_{\text{SET}}(\text{pageB}) = \{\text{ribbon}, \text{race}, \text{radical}, \text{drama}, \text{general}, \text{item}, \text{quote}\}$
   $\text{DegreeOfRelavence}(\text{pageA}, \text{pageB}) = W_{\text{SET}}(\text{pageA}) \cap W_{\text{SET}}(\text{pageB}) = 3$
4. For the RV, co-domain pages are selected in preference if there exist the same degrees of relevant pages.
   For the DT, cross-domain pages are selected in preference if there exist the same degrees of relevant pages.

We evaluated the two differently linked information spaces by our potential serendipity evaluation program. The results are shown in Table 3.

The DT has a potential serendipity value more than twice that of RV. The ratio of measured values is not exactly the same as that of the user study result conducted by [12] but it shows a similar pattern. The small margin of error can be considered allowable in this situation since the concept of serendipity is subject in nature. This verifies that our program provides seemingly trustful serendipity measurement on information space.
4. Discussion

In this section, we discuss ways to graft the serendipity on to the information system that seems promising. We suggest three directions: (1) system design that enriches an education system; (2) Web banner advertisement network that can effectively attract traffic; and (3) a search engine that reduces the “filter bubble” problem.

According to the reviewed literature, we discovered that engaging serendipity with the course recommendation system enhances recommendation quality. Most course recommendation systems use a collaborative-filtering-based algorithm, which focuses on users’ behavior. This suggests an overly narrow set of suggestions, making students have fewer opportunities to engage with new materials in new fields. It is desired to foster learning environments to be serendipitous in an educational context, giving students a chance to engage with new materials by chance. This might broaden the students’ view of their study. It was identified that learning new materials helps students to foster their further interest and feel satisfied. We suggest engaging with unplanned suggestions as well as structured curricula. We expect that the serendipitous course recommendation system will foster academic growth by delivering various choices to engage students in the learning process.

We consider that collaborating advertisement networks with the concept of serendipity can also bring positive effect. A cost-per-click is a standard metric for banner advertisements; the advertiser sends money to the advertisement provider. It is essential to not only expose banners to users but also make users click and browse the banners. In many cases, users’ click tags are monitored in real time, and the banners are displayed that best match users’ interests. This way of digital advertising tends to show repeated or similar contents to the users. Most people ignore the boring banner advertisements—as only a small percentage of the people are engaged. Google reported “the average click-through rate for banners has fallen to 0.06%” [21]. Confirming the research of [22], recommending serendipitous items in online content hosting services grabbed users’ attention and encouraged them to be immersed more. Ref. [23] also identified that the success of online marketing could be achieved by creating exciting experiences for users.

In the same context, we predict the serendipity can play a role in search engines. Search engines are designed to show the most relevant results based on a computed likelihood that certain information relates to the query (Dabrowska, 2015), inadvertently placing users in a filter bubble. Users are exposed to overly personalized information, leading to a lack of pleasure and surprise. To effectively serve information in the Big Data world, it is needed to
balance out the personalized search results with unexpected but interesting search results. The best example of breaking the filter bubble with serendipity is YouTube. YouTube promoted the exploration of novelty and serendipity instead of lowering the diversity of consumed content [24]. The exceptional YouTube algorithm enabled developing YouTube into a successful platform. We recommend that personalized websites be changed in the direction of serendipity to confine Web users in their platforms.

In addition to the application possibilities discussed in this section, we believe that there exist plenty of others that have not yet been found. We account for the importance of serendipity in the context of information space and encourage that serendipity be continuously and actively studied.

5. Conclusions

In this paper, we proposed an evaluation model that can measure the potential of a serendipitous discovery in hypertext. The model reflects three aspects that can contribute to a serendipitous discovery. These include pre-encountering, post-encountering and a discovery. The novel part of the model is a quantitative formula which can be used against a hyperlinked information space and the system becomes publicly available. It occupies around 15 Mb of a disk space and runs on Java-8 or later versions. This can be used in developing a serendipitous information space without adding manpower to the verification process. We are currently investigating ways by which educational materials can be provided with the aid of the proposed system. The idea is that one can build an initial information space with the learning materials and our system can be used to compute the potential serendipity value of the space. Then, it could be possible to change the structure of the information space in order to maintain a certain degree of serendipity.

Author Contributions: Writing—original draft, Y.K.; Writing—review and editing, B.H., J.K., J.S., S.K. and S.P., Y.K. is the first author and S.P. is the corresponding author, respectively. Other authors contributed equally to the preparation of this manuscript and J.K. participated in this work before graduating from Korea University. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available data sets were analyzed in this study. The data can be found here: https://github.com/KimYuri94/SerendipityEvaluationModel, accessed on 9 June 2021.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Following Terminologies Are Used in This Manuscript

Terms and meanings are summarized in Table A1.

| Term                                      | Meaning                                                                 | First Appearance   |
|-------------------------------------------|-------------------------------------------------------------------------|--------------------|
| Active search                              | information-seeking behavior which looks for specific information        | page 1, Abstract   |
| Sidetracked search                         | information-seeking behavior which incorporates passive attention        | page 1, Abstract   |
| Incidental information acquisition (IIA)  | serendipitous information discovery in information space                 | page 1, Section 1  |
| Passive search                             | purposeless information-seeking behavior which does not look for any information in particular | page 1, Section 1  |
Table A1. Cont.

| Term              | Meaning                                                                 | First Appearance          |
|-------------------|-------------------------------------------------------------------------|----------------------------|
| Passive attention | during information search process, a user is open to unsought information | page 1, Section 1          |
| Passive information acquisition | finding unsought information while looking for other information | page 1, Section 1          |
| Information space | Web space, where a collection of Web pages lies                          | page 2, Section 1          |
| Target information | specific information that a user tries to find                          | page 3, Section 3.1        |
| Search path       | a finite sequence of information pieces which a user encounters during the process of active search | page 3, Section 3.1        |
| Sidetracked path  | a finite sequence of information pieces which a user encounters during the process of sidetracked search | page 3, Section 3.1        |
| Focus shift       | putting original target information out of mind, when they are not in active search mode | page 5, Section 3.2.1      |

References
1. Shackle, S. Science and Serendipity: Famous Accidental Discoveries. Available online: https://newhumanist.org.uk/articles/4852/science-and-serendipity-famous-accidental-discoveries (accessed on 11 July 2021).
2. Bates, M.J. The design of browsing and berry-picking techniques for the online search interface. Online Rev. 1989, 13, 407–424. [CrossRef]
3. Wilson, T. Models in Information Behaviour Research. J. Doc. 1999, 55. [CrossRef]
4. Erdelez, S Information encountering. Inf. Today 2005, 25, 179–184.
5. Erdelez, S.; Makri, S. Information encountering re-encountered: A conceptual re-examination of serendipity in the context of information acquisition. J. Doc. 2020, 76, 731–751. [CrossRef]
6. Erdelez, S. Information Encountering: An Exploration beyond Information Seeking. Ph.D. Thesis, Syracuse University, Syracuse, NY, USA, 1995.
7. Erdelez, S. Investigation of Information Encountering in the Controlled Research Environment. Inf. Process. Manag. 2004, 40, 1013–1025. [CrossRef]
8. Makri, T.M.R.S. Accidental Information Discovery: Cultivating Serendipity in the Digital Age; Chandos Publishing: Cambridge, MA, USA, 2016.
9. Williamson, K. Discovered by chance: The role of incidental information acquisition in an ecological model of information use. Libr. Inf. Sci. Res. 1998, 20, 23–40. [CrossRef]
10. Jannica, H. Fast surfing for availability or deep diving into quality—Motivation and information seeking among middle and high school students. Inf. Res. Int. Electron. J. 2006, 11, n4.
11. Kang, C.G.; Park, H.; Park, S. Towards a Model for Serendipitous Discoveries. In Information Science and Applications; Springer: Singapore, 2020; pp. 613–619.
12. Pardos, Z.A.; Jiang, W. Combating the Filter Bubble: Designing for Serendipity in a University Course Recommendation System. arXiv 2019, arXiv:1907.01591.
13. Mendonça, S.; Cunha, M.; Clegg, S. On serendipity and organizing. Eur. Manag. J. 2010, 28, 319–330. [CrossRef]
14. Fine, G.A.; Deegan, J.G. Three principles of serendip: Insight, chance, and discovery in qualitative research. Int. J. Qual. Stud. Educ. 1996, 9, 434–447. [CrossRef]
15. Simard, M.; Laberge, D. From a Methodology Exercise to the Discovery of a Crisis: Serendipity in Field Research. Proj. Manag. J. 2015, 46, 47–59. [CrossRef]
16. Copeland, S. On serendipity in science: Discovery at the intersection of chance and wisdom. Synthese 2019, 196. [CrossRef]
17. Kantorovich, A. Scientific Discovery: Logic and Tinkering; State University of New York Press: New York, NY, USA, 1993.
18. Zheng, D. The 15 Second Rule: 3 Reasons Why Users Leave a Website. 2020. Available online: https://www.crazyegg.com/blog/why-users-leave-a-website/ (accessed on 22 January 2021)
19. Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. Comput. Netw. ISDN Syst. 1998, 30, 107–117.
20. Gupta, V.; Lehag, G. A Survey of Text Summarization Extractive Techniques. J. Emerg. Technol. Web Intell. 2010, 2. [CrossRef]
21. Brenner, M. Banner Ads Have 99 Problems Furthermore, A Click Ain’t One. 2020. Available online: https://marketinginsidergroup.com/content-marketing/banners-99-problems/ (accessed on 14 February 2021).
22. Lu, H.P.; Cheng, Y.H. Sustainability in Online Video Hosting Services: The Effects of Serendipity and Flow Experience on Prolonged Usage Time. *Sustainability* 2020, 12, 1271. [CrossRef]
23. Hoffman, D.; Novak, T. A New Marketing Paradigm for Electronic Commerce. *Inf. Soc.* 1998, 13, 43–54.
24. IANS. How People End Up ‘Confined’ on YouTube. 2020. Available online: https://www.thehindu.com/sci-tech/technology/internet/how-people-end-up-confined-on-youtube/article31424163.ece (accessed on 14 February 2021).