From X-Rays to Silly Putty via Uranus: Serendipity and its Role in Web Search

Paul André
Electronics and Computer Science
University of Southampton, UK
pa2@ecs.soton.ac.uk

Jaime Teevan, Susan T. Dumais
Microsoft Research
Redmond, WA 98052 USA
{teevan,sdumais}@microsoft.com

ABSTRACT

The act of encountering information unexpectedly has long been identified as valuable, both as a joy in itself and as part of task-focused problem solving. There has been a concern that highly accurate search engines and targeted personalization may reduce opportunities for serendipity on the Web. We examine whether there is the potential for serendipitous encounters during Web search, and whether improving search relevance through personalization reduces this potential. By studying Web search query logs and the results people judge relevant and interesting, we find many of the queries people perform return interesting (potentially serendipitous) results that are not directly relevant. Rather than harming serendipity, personalization appears to identify interesting results in addition to relevant ones.

Author Keywords

Serendipity, Web search, personalization, partially relevant results.

ACM Classification Keywords

H5.2: Information interfaces and presentation: User Interfaces. H3.3. Information storage and retrieval: Information search and retrieval.

INTRODUCTION

Serendipity, or the act of unexpectedly encountering something fortunate, is widely regarded as valuable in the processes of science, technology, art, and 'daily life' [15]. From discoveries and inventions (the chance penetration of black paper by X-Rays, the unsuccessful rubber substitute Silly Putty finding new life in a toy store, the planet Uranus being identified during a search for comets), to fun diversions (stumbling across adjacent entries in an encyclopedia), to proposed metrics for recommender systems [5], serendipity is seen to play an important role in both work and pleasure.

But how has the Web affected serendipity? Have highly targeted and directed search engines made discovery efficient but dull [7], or is it now even easier to stumble across something brilliant [6]? While Web sites such as BoingBoing, the most popular blog, are constantly updated with eclectic content and provide opportunities for 'unexpected encounters', we focus on Web search to understand the extent to which serendipity on the Web is possible in a task-focused context. Web search is one of the most common internet activities [8]. But concern has been expressed that ever-improving search engines, as well as the use of personalization to display exactly what the user is looking for, will interfere with serendipitous encounters [7].

By its nature, serendipity is hard to study. In this paper rather than trying to induce or identify serendipity, we conducted a study to explore the potential for serendipitous encounters. Specifically we examine:

1. Whether there is the potential for serendipitous encounters during Web search; and
2. Whether the ability to better target the user's interests through personalization reduces this potential.

We describe related work, our methodology results, and conclude with recommendations for future work.

RELATED WORK

Defining Serendipity

Modern definitions of serendipity in the literature cover a broad range of occurrences. Table 1 summarizes a number of definitions that have been explored by researchers, broken down along two axes: what activity was engaged in at the time of the serendipitous encounter, and what type of information was encountered.

		Type of info. found	
		Relevant to goal	**Not relevant to goal**
Info. activity	**Directed browsing**	Serendipitous Info. Encountering [4]	Info. Encountering [3]
	Non-directed browsing		Serendipitous Info. Retrieval [14], Opportunistic Browsing [1]
	None		Involuntary Browsing [1]

Table 1. Definitions of serendipity, broken down along two axes: what information activity was engaged in at the time of encounter, and what type of information was found.

Our research expands on what is known about serendipity in goal-directed browsing and search (top row of Table 1) by focusing on serendipity in Web search and the influence that personalization has on the potential for serendipity.

Value of Serendipity

Various values of serendipitous encounters are apparent in the definitions: reinforcing an existing problem or solution or taking it in a new direction [4], rejection or confirmation of ideas [14], identifying information relevant to a latent goal [1], or just finding information of interest [3].

Spink et al. [11] found that *partially-relevant* search results, identified as "*containing multiple concepts, [or] on target but too narrow,*" play an important role in a user's information seeking process and problem definition. These values are analogous to many definitions of serendipity, and suggest that there is value in partially-relevant results. In education the 'zone-of-learnability' [16] refers to texts that provide optimal learning because they are related to, but just distant enough, from what the student already knows. Since search results are inherently related to the query in some way, there may be a similar zone for serendipity.

Facilitating or Inducing Serendipity

Toms [14] manipulated the purpose with which users approached a digital newspaper: goal-directed, or no pre-defined goal, with two methods of access: keyword search, or suggested articles. Participants spoke of the value of chance encounters: "*If you focus on your interests, then your interests are going to stay what they are.*" In a study designed to induce serendipity [3], participants with a common coursework task were given a new search task that was reverse engineered, so that one coursework relevant result appeared. Nine out of ten participants noted the course-relevant result, but none diverted from their search task to view it, highlighting the difficulty of measuring serendipity in search interactions.

Collaborative filtering systems identify interesting content by matching individuals with other similar individuals. Herlocker [5] suggests measures like novelty and serendipity should be used to assess the quality of recommendations. Collaborative filtering systems have promoted novelty and serendipity by helping users uncover less popular [9] and more diverse items [17]. Though participants are often able to *talk* about past experiences of "chance encountering" [2,4], it is hard to identify, induce or study serendipity. To combat this in our research, rather than try to create serendipitous situations or identify instances of serendipity, we conducted a study that allowed us to understand the potential for serendipity, which we hope we can then capture in future work.

METHODOLOGY

This section describes the methodology and data sources we used to explore whether there is the potential for people to encounter serendipitous results in Web search, and how that potential is affected by personalization. We conducted a controlled study to look at which search results people rated interesting and relevant, and used large-scale log analyses to complement the explicit judgments. In the study we asked participants to rate search results on two dimensions:

1. *Relevance* on three levels: Relevant, Partially, or Not. This judgement relates to whether the result answered the information goal the participant had in mind.

2. *Interestingness* on three levels: Interesting, Partially, or Not. This judgement relates to whether the result was of interest to the participant. That is, whether they would be tempted to click the result because it answers this query, another task-related query, or just for general interest.

We hypothesize that search results that are interesting but not highly relevant indicate a potential for serendipity – as indicated in the top row of Table 1 (encountering something related or unrelated to the goal of directed browsing or search), and as suggested by Spink's findings of the value of partially-relevant results. It is this particular aspect of serendipity (interesting but not directly related results) that we focus on in this research.

To gather these judgements, participants were asked to generate queries based either on previously issued queries or an existing information need, and to write a short description of their intent. For each query, 25 results were presented with the usual title, snippet and URL. Next to each result were two sets of rating buttons – one set for relevance, one for interestingness. Subjects could rate results in any order, view the associated website, and change their ratings up until pressing 'save'. The 25 results selected for evaluation were the odd numbered results from the top 50 returned by Live Search, shown in random order. While searchers typically only look at the top few results, we collected judgments that spanned the top 50 because interesting and relevant results are often missed because they are ranked low [12], and we hypothesized 'partially-relevant' (but interesting and potentially serendipitous) results [11] would likely appear lower down the list.

Besides collecting relevance judgements, we obtained additional information about each result by examining search logs from Live Search. The log data were from a one month sample (Jul 18 – Aug 15, 2008), which coincided with the time during which explicit judgments were collected. We used the logs to measure the popularity of queries, the number and diversity of results for each query, and user interactions with search results. We hypothesized that serendipity was more likely to occur in diverse result sets, and that diverse content would be reflected in diverse clicks, as measured by click entropy. Low click entropy means that a small number of results were clicked for the query (if entropy = 0 then the same result was clicked every time the query was issued). High click entropy means many different results were clicked for the query. The study was conducted on the participants' own computers by installing a browser toolbar. The toolbar enabled us to collect information relating to how personally relevant each search result was to the participant who issued the query. We measured: 1) how similar the text of each result was to the text of desktop content including Web pages,

documents and email (a *content*-based similarity measure), and 2) how similar each URL was to pages in the participants' browsing history and favourites (a *behaviour*-based similarity measure).

These two measures (content match and user interaction history) have been commonly used in previous work [10,13] to personalize search result ranking, and have been shown to be correlated with explicit relevance judgements. In this paper we explore how they relate not only to how relevant a result is, but also how interesting it is.

Participants. Thirty-six people participated in the study, and evaluated a total of 92 queries (2300 judgements). Participants were all employees of Microsoft.

Interviews. Three interviews were conducted with participants who rated a high number of results *interesting* and *not relevant*, since we hypothesize that these results are potentially serendipitous. This allowed us to further explore the nature of the particular results and the query.

RESULTS

We examine the results organized by our initial questions.

Is There Potential To Encounter Serendipitous Results?

We examine explicit judgements of search result relevance and interestingness to measure the extent to which current search algorithms present searchers with the opportunity for serendipitous behaviour.

Table 2 shows the total number of the explicit judgements gathered, broken down into the three levels of Relevance and Interestingness. On average there are 5.4 very relevant results per query and 4.1 very interesting results per query.

Count of Judgements	Very	Partially	Not	
Relevance	494	599	1207	*(2300)*
Interestingness	381	574	1345	*(2300)*

Table 2. Count of results from judgement study, broken into Relevance and Interestingness.

Table 3 expands Table 2 to show the relationship between relevance and interestingness judgements. While there is a relationship between relevance and interestingness, there are also interesting differences. We highlight the area most likely to include serendipitous results - those that were judged partially or very interesting, but not very relevant to the query. Twenty-one percent of all results (an average of 5.25 per query) fall into this category.

All three participants who were interviewed stated that the title, snippet or URL made a difference in assessing interestingness, but that they would click on 'non-quality' results (e.g., random blogs) if they seemed interesting. One participant categorised the interesting items as 'one step away' from what he was looking for with his original search intent. Looking back over the judgements, another participant said he would now change some answers to be relevant or interesting that he had not considered so at the time, highlighting the variability even in self-consistency.

Interesting results appeared to generally be ones that participants would like to explore if they had more time, though some were only of a transient interest.

Count of Judgements	Very Interesting	Partially Interesting	Not Interesting
Very Relevant	288	174	32
Partially Relevant	77	318	204
Not Relevant	16	82	1109
	21% of results (Potentially Serendipitous)		

Table 3. Count of individual results according to judgment category. Highlighted area is the area most likely to contain serendipitous results.

Types of Queries That Are Serendipitous

To see if we could determine which types of queries had more potential for serendipitous results, we characterized each query using several features. Some of the features we explored appear to not be related to serendipity (e.g., query length, number of results returned, popularity), and some were promising but would need a larger sample to achieve statistical significance (e.g., work related vs. not, navigational vs. informational, contain person's name vs. not). Click entropy was found to be significant.

We calculated the click entropy for the 26 (of 92) queries that had ten or more clicks in our one month sample of log data. Table 4 shows the correlation between a query's click entropy and the number of results for that query that were judged either interesting, potentially serendipitous, neither interesting nor relevant, or not interesting.

Correlation with Click Entropy	Coefficient (R)	p-value
Interesting	0.36	$p<.10$
Potentially Serendipitous	0.43	$p<.05$
Not Relevant/Not Interesting	-0.52	$p<.01$
Not Interesting	-0.51	$p<.01$

Table 4. Correlation between click entropy for a query, and the number of four types of result.

The positive correlation between entropy and the number of interesting (and potentially serendipitous) results suggests that people may have clicked varied results not just because they could not find what they wanted, but because they considered more things interesting or were more willing to go off at a tangent. We also see a negative correlation between entropy and the number of not relevant and not interesting results, further supporting the idea that queries with high click entropy are more likely to include interesting and potentially serendipitous results.

In summary, we find that there is the potential for serendipitous encounters during Web search – more than a fifth of all search results were judged interesting but not highly relevant to the search task. Further, there are some characteristics of queries and search interactions that can be used to identify queries with more potential for serendipity.

Does Personalization Affect Serendipity?

The analyses above looked at potential for serendipity in general search results; here we focus on how personalized search could affect serendipity.

As part of our judgement study we also collected information relating to how similar each result was to the content in the participant's desktop index (a *content* score) and previously visited or bookmarked sites and domains (a *behaviour* score). These two personalized scores were combined (as described in [13]) to compute a personal relevance score for each search result.

Table 5 shows the average of these personal scores, broken down by the rated relevance and interestingness of the result. Table 5 shows that the personal score is related to relevance (right column), as previous research has shown, and it *also* shows that the personal score is related to interestingness (bottom row). This is not simply due to the relationship between relevance and interestingness ratings. Indeed, Table 5 highlights that very interesting results (first column) have higher personal scores ranging from .13 to .10 than very relevant results (first row) with personal scores ranging from .13 to .07.

Personal Score	**Very Interesting**	**Partially Interesting**	**Not Interesting**	(Total)
Very Relevant	.13	.08	.07	.11
Partially Relevant	.11	.08	.07	.08*
Not Relevant	.10	.06	.06	.06*
(Total)	.13	.08*	.06*	

Table 5. The personalized score, indicating how personally relevant a result is, broken down by Relevance and Interestingness judgements.
(* indicates total is significantly different (independent measures t-test) from column or row preceding it, at p<0.001 or greater).

As we mentioned previously, the personal score is made up of two components: content and behaviour. Though space constraints preclude more detail, it is interesting to note that past browsing behaviour in particular is important in identifying interesting results.

CONCLUSIONS AND FUTURE WORK

We have examined the potential for serendipity in Web search, and the effect of personalization on that potential. We find that there *does* exist potential for serendipity, and that certain query features, notably click entropy, are correlated with serendipitous queries. We also find that personalization scores correlate with both relevance and also with interestingness, suggesting that information about personal interests and behaviour may be used to support serendipity.

In future work we intend to examine to what extent this potential is realized in *actual* search interactions. We also intend to examine how to present such serendipitous results. For many goal-directed tasks it may not be appropriate to show them mid-task, but rather to save them for a more opportune time. Our long-term goal is to investigate how a system may harness the creative potential of serendipity, and how to measure that.

REFERENCES

1. De Bruijn, O. & Spence, R. A new framework for theory-based interaction design applied to serendipitous information retrieval, *ACM TOCHI 15,*1 (2008), 1--38.

2. Erdelez, S. Information encountering: It's more than just bumping into information, *Bulletin of the American Society for Information Science 25,*3(1999) 25--29.

3. Erdelez, S. Investigation of information encountering in the controlled research environment, *Information Processing and Management 40,*6 (2004), 1013--1025.

4. Foster, A. & Ford, N. Serendipity and information seeking: An empirical study, *J. Documentation 59,*3 (2003) 321--340.

5. Herlocker, J. L., Konstan, J. A., Terveen, L. G. & Riedl, J. T. Evaluating collaborative filtering recommender systems, *ACM TOIS 22*, 1 (2004), 5--53.

6. Johnson, S. B. (2006), Can we please kill this meme now. Retrieved 14 Sep 2008: http://tinyurl.com/sbjohnson

7. McKeen, W. (2006), The endangered joy of serendipity, St. Petersburg Times Retrieved Sep 2008 http://tinyurl.com/mckeen

8. Ranie, L. & Shermak, J. (2005), Pew Internet and American Life Project: Data memo on search engine use. Retrieved 14 Sep 2008: http://pewinternet.org/pdfs/PIP_SearchData_1105.pdf

9. Sarwar, B., Karypis, G., Konstan, J. & Reidl, J. Item-based collaborative filtering recommendation algorithms, *Proc WWW '01*, 285--295.

10. Shen, X., Tan, B. & Zhai, C. Implicit user modeling for personalized search, *Proc CIKM'05* 824—831.

11. Spink, A., Greisdorf, H. & Bateman, J. From highly relevant to not relevant: Examining different regions of relevance, *IP&M 34*, 5 (1998), 599--621.

12. Teevan, J., Dumais, S.T. & Horvitz, E. Beyond the commons: on the value of personalizing Web search. *Proc PIA'05,* 84--92.

13. Teevan, J., Dumais, S.T. & Horvitz, E. Personalizing search via automated analysis of interests and activities, *Proc SIGIR '05,* 449--456.

14. Toms, E. Serendipitous information retrieval, *Proc of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries,* 2000.

15. Van Andel, P. Anatomy of the Unsought Finding. Serendipity: Origin, history, domains, traditions, appearances, patterns and programmability, *The British Journal for the Philosophy of Science45*, 2 (1994), 631--648.

16. Wolfe, M., Schreiner, M., Rehder, B., Laham, D., Foltz, P., Kintsch, W. & Landauer, T. Learning from text: Matching readers and texts by Latent Semantic Analysis, *Discourse Processes 25*, (1998) 309--336.

17. Ziegler, C., McNee, S.M., Konstan, J.A. & Lausen, G. Improving recommendation lists through topic diversification, *Proc. WWW '05* 22—32.