

Rahman, M. Ataur and Wilson, Max L. (2015) Exploring opportunities to facilitate serendipity in search. In: 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 9-13 Aug 2015, Santiago, Chile.

Access from the University of Nottingham repository:

<http://eprints.nottingham.ac.uk/31315/1/sigir2015-serendipity.pdf>

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

- Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners.
- To the extent reasonable and practicable the material made available in Nottingham ePrints has been checked for eligibility before being made available.
- Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
- Quotations or similar reproductions must be sufficiently acknowledged.

Please see our full end user licence at:

http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Exploring Opportunities to Facilitate Serendipity in Search

M. Ataur Rahman
School of Computer Science
University of Nottingham, UK
shuvo.2007@gmail.com

Max L. Wilson
School of Computer Science
University of Nottingham, UK
max.wilson@nottingham.ac.uk

ABSTRACT

Serendipitously discovering new information can bring many benefits. Although we can design systems to highlight serendipitous information, serendipity cannot be easily orchestrated and is thus hard to study. In this paper, we deployed a working search engine that matched search results with Facebook ‘Like’ data, as a technology probe to examine naturally occurring serendipitous discoveries. Search logs and diary entries revealed the nature of these occasions in both leisure and work contexts. The findings support the use of the micro-serendipity model in search system design.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Relevance Feedback; H.5.2 [User Interfaces]: Screen Design

Keywords

Serendipity; design; web search

1. INTRODUCTION

Serendipity is a naturally occurring phenomenon in our daily life, and has been proven to be valuable in many domains, including: medicine [18], business [8], creative thinking [2], as well as simply discovering new songs, good movies, or even useful literature in a library. Fine & Deegan [12] proposed the definition of serendipity as “*the unique and contingent mix of insight coupled with chance*”, which sums prior definitions by saying that a successful serendipitous encounter requires both a prepared mind and an accidental occurrence, to create a new insight.

To deliberately support serendipitous discoveries, research has tried to deconstruct serendipity and identify design opportunities [1]. In this project, we suggest that user interests, gathered from social media profiles, can be used in web search to facilitate serendipity: 1) users might select partially relevant or less directly relevant search results, when seeing

information is related to other interests, and 2) users can identify valuable ways to connect new interesting information to existing ideas.

The subjective and individual nature of serendipity, however, makes it very difficult to perform a controlled user study to examine the impact of potentially serendipitous user interface designs [13]: 1) Serendipitous searches must be personally relevant, rather than from fictional tasks, 2) users cannot be tasked with being serendipitous, and 3) the discovery must also be personally valuable to make the connection. Consequently, rather than perform a controlled laboratory experiment, we chose to deploy a live system, named Feegli (Figure 1), which used real profile data from users’ social media profiles. We asked participants to use this as their primary search engine for 7 consecutive days, which allowed us to explore the potential of encountering serendipitous information in real contexts, and to investigate participants’ experiences. This paper, therefore, describes a lightweight naturalistic study of serendipity in web search happening in the wild, insights into its nature and frequency of it happening, and presents findings that support the notion of micro-serendipity as a relevant model for web search.

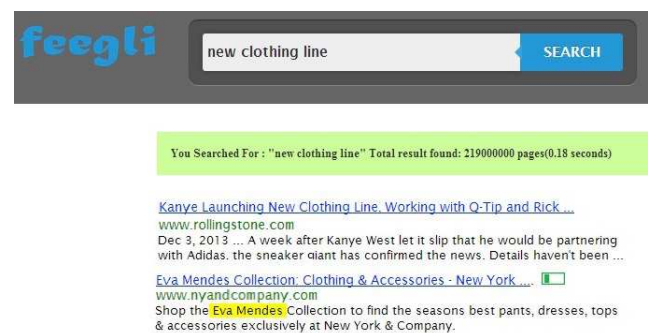


Figure 1: Feegli highlights interest topics, gathered from social media profiles, within search results

2. RELATED WORK

The definition of serendipity is both unclear and disputed [21, 15], with many researchers from various disciplines, having invested a great amount of time to study it [1, 5, 7, 9, 8, 16, 20]. Beyond simply encountering new information [11], McCay-Peet & Toms [17] describe serendipity as an event involving unexpectedness and a bi-sociation (mental connection). Conversely, Rubin et al. [20] described serendipity as happening when 1) someone is prepared to see cues, 2)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

SIGIR'15, August 09 - 13, 2015, Santiago, Chile.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3621-5/15/08 ...\$15.00.

DOI: http://dx.doi.org/10.1145/2766462.2767783

an unexpected event occurs, and 3) a connection leads to a fortuitous outcome.

Makri & Blandford [16] focused on the stages of connection, realisation, and reflection that lead people to retrospectively consider an event serendipitous. They also highlight that most serendipitous events lead to a *macro*-serendipity tangible benefit, but that outcomes may also simply be an advancement of knowledge or understanding. Bogers & Bjorneborn further focused on these as *micro*-serendipity events, which were valuable even without clear tangible outcomes [5].

Andre et al [2] found that serendipity happens in web search, where interesting and both partially relevant or completely irrelevant search results have the potential for serendipity. Letizia [14], a browser agent, analysed web histories to recommend potentially interesting items to the user, but aimed to recommend interesting pages rather than highlight serendipity during normal web search. Similarly, a system called Max [7] sent recommendations to users via email after analysing two months of browsing data. Evaluations revealed that of 100 recommendations, 7 were found to be valuable by users. Mitsikeru [4] did try to highlight serendipitous results during search, but was not studied.

3. RESEARCH APPROACH

As serendipity cannot be simulated within a lab study, we chose, like others [2], to use investigative methods to explore naturally occurring examples of serendipity. Unlike prior work, however, we built a custom working search engine, Feegli, which uses real social media data to highlight results from normal searches that might also relate to one of their interests. As found with the design of PIM tasks [10], the individual nature of serendipity requires the use of real personal data. This personal nature of serendipity, however, precludes the creation of a comparable control group. Feegli, as a technology probe, however, allowed us to study naturally occurring serendipitous occasions, using event logs and insights from a longitudinal diary study.

3.1 Feegli: a serendipitous search engine

Feegli was built using the Google Custom Search REST API, and extracted interest data, as JSON files, from the Facebook Graph API v1.0, as soon as they signed up using their Facebook account. Further, participants added additional interests within a settings page for the Feegli system. When users issued a query, the titles and snippets of the search results were matched against the recorded Like data, as exact strings and whole words. Partial matches were not highlighted to avoid matches like ‘cha**IR**’. Unlike typical personalisation systems, we did not re-rank results based on these interests. Instead, we highlighted results, leveraging the Restorff Isolation Effect [22] to provide a secondary notion of potential relevance to ranking.

Several indicators of interest-relevance were considered during the design phase: 1) interest matched, 2) intensity of match, and 3) source of interest. After discussing several alternatives for such indicators during focus groups, the last of these was not included in the design: social media icons were presumed to be for sharing results on social networks, and listing the source network provided less value than including more results before the fold.

Matched interests were highlighted in place, with a yellow background as seen in Figure 1, rather than being listed separately. For intensity of match, we chose to indicate

number of matched interests, rather than estimating from the snippets an *amount* of relevance. Consequently, as shown in Figure 2, we designed a glyph (like [3]) to show up to three matches with increasing intensity.

Each query and all presented search results were logged by the system, including whether each result was highlighted as being related to an interest. Consequently, we could tell from the logs every potentially serendipitous result shown, for which query and which interest, and where in the results list it was presented. Further, each clicked result was logged by the system. We also wanted to gather at-the-time judgements about results, but with minimal interruption to their normal search behavior. Consequently, each time, but only when, a highlighted link was clicked on by a participant, Feegli asked for the primary reason, as per Figure 3. A popup was not shown if they did not click on a visible highlight, as we did not want to artificially drive click through towards such results. Further, we wanted to reduce the impact of the popups on normal web search behaviour.



Figure 2: Interest glyph shows number of interests

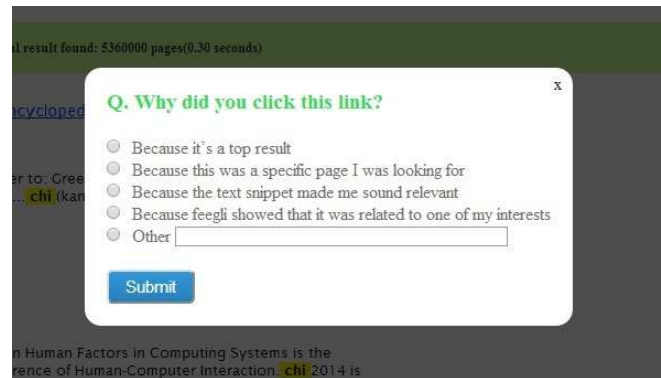


Figure 3: Popup when a highlighted link was clicked

3.2 Participants - Daily Diary Entries

Study participants were recruited primarily from the University of Nottingham, UK. In total, 14 active facebook users, with ‘Like’ data, were recruited. 9 were male, 5 were female. 12 already had an undergraduate degree. The study was approved by the school’s ethics board and participants were entered into a draw to win \$100.

To learn about the potentially serendipitous events experienced by our participants, we asked them to fill in a daily structured diary entry. The diary began with an open-ended request to describe any serendipitous experiences they had with Feegli during that day. Then, to support participants in recalling events, participants were presented with illustrated examples of their queries from that day. If a user clicked on the highlighted serendipitous results, they were asked to briefly explain the reason for clicking on it, and whether their information goal was met or not. Alternatively, when a user *did not* click on a highlighted result, they were asked about whether they had seen it, considered clicking on it,

and so on. The daily diary was manually constructed by the experimenter to include the daily examples, and sent to the participants as a word document via email. Diary entries were returned in the same way.

4. RESULTS

Each participant used Feegli as their primary search engine between 25th April to 4th May, 2014. During these 11 days, participants made 506 queries in total (avg. 46 queries per day). Feegli returned 5380 search results during the study. Of these, Feegli highlighted 445 (8%) to be potentially serendipitous. Of these, 57 were clicked on by participants. From the immediate popup feedback, 35.1% were chosen because the text snippet made them seem relevant and 26.6% because the link was the top result. 17.5% of results were clicked because the user was specifically looking for the link. 22.8%, however, were clicked because Feegli highlighted them.

Commonly there are three types of queries people make in search engines: Navigational (to get to a certain website), Transactional (e.g. to buy, download, or retrieve something) and Informational (to resolve an information need) [6]. 84% of the queries were classified as informational; navigational and transactional counted for 8% each. Of the 445 highlighted results, 96% were shown for informational queries, and 4% for navigational queries. Using proportion of queries as an expected ratio, a chi-squared analysis showed that highlights were significantly more likely to be shown for informational queries ($\chi^2(2) = 49.826, p < 0.0001$). Of the 57 highlighted results clicked on by users, 98% were for informational queries. Using proportion of highlights shown as an estimated ratio, a chi-squared analysis showed that this was not significantly more than would be expected. All of the occasions where a user indicated that they clicked on the highlighted result because of the potentially serendipitous content were from these informational queries.

We also analysed the likely work or leisure focus of the queries. 52% were related to work and 40% of the queries are related to leisure search; 8% could not be easily classified. 64% of the highlighted, potentially serendipitous, results were shown for work queries, while only 34% were for social. Using proportion of classifiable queries submitted as an estimated ratio, a chi-squared analysis showed that significantly more highlights were shown for work-oriented queries than expected ($\chi^2(1) = 11.352, p < 0.001$). Of the 57 actually clicked, 60% were for work queries, while the remaining 40% were for leisure queries. Using proportion of highlights shown as an estimated ratio, a chi-squared analysis showed that highlights were not significantly more likely to be clicked in either work or leisure searches. The clicks that users said were because of the potentially serendipitous highlight were present in both groups. Perhaps surprisingly, these results indicate that there is opportunity and reason to support serendipitous discoveries in focused work tasks too, as well as when more casually searching.

4.1 Diary Entries

Diary entries confirmed that users did experience serendipity while using Feegli, primarily for informational needs, but also for both leisure and work searches.

4.1.1 Macro- vs Micro-serendipity

Following a recommendation, Feegli helped P3 to stumble upon unexpected link and later helped to find serendipitous

results, for example: *“I clicked it because the green icon suggested that it had information I needed. After looking at closer look I found out the paper has a lot details about designing a mobile communication system and it had information on different generation of systems, which helped me. I found serendipitous information [about] a comparison between different modulation techniques and it led to me change my initial plan and gave me a series of new ideas about designing the system.”* (P3)

The intensity glyph helped P3 to experience serendipitous information that was not among the top results; rather it was positioned seventh. This indicates that Feegli helped users to examine more results below the top, most frequently clicked search results. Perhaps interestingly, however, the user saw the glyph as being information they *needed*, rather than being related to an interest. In this case, the result led to a serious change in a work product, consistent with the macro-view of serendipity.

Highlighted interest keywords also helped to draw attention, as P5 said *“Well, the search engine showed me that the link is of my interest and that’s why I clicked it. I was randomly searching and looking for if there is a new brand in clothing. It was the fifth link which highlighted Eva Mendes in the text, I was curious to know more and later found out Eva Mendes has a clothing line. It is difficult for me to answer whether it met my information need or not however it was nice to know that she has a clothing line. I would call this serendipitous [...] I am not sure if it was valuable to me or not but it definitely satisfied my curiosity.”* (P5)

P5’s experience is a good example of a leisure search, but also of micro-serendipity as the user was not sure if the new information was fulfilling the original information goal but it was serendipitous. This user may also have been more open to serendipitous recommendations, given the undirected nature of the query.

4.1.2 Serendipitous Discovery vs. Directed Search

Highly directed search may reduce serendipity, as when a user is focused on finding specific information, serendipitous recommendations were ignored: *“I did not click any of those because I did not look that far down; I found my target website within first few top results.”* (P8)

Users also reported either looking at the type of link or reading the snippet to understand what type of information the link can offer: *“I did not click because I was looking for the information regarding old Robocop TV series and I went to see at first IMDB link to check it and I found it from there.”* (P11). Instead, P11 utilized the highlight when goal was vague in nature: *“... first link was a Wikipedia link and of second and third, only 3rd link was highlighted so I thought it would may be important. I was looking for alternative models for kuznet curve and after a while I found it...”*

5. DISCUSSION

Participants in our study described serendipitous encounters for both work and leisure related searches. Perhaps surprisingly, work related queries drove serendipitous encounters more than for leisure search, despite both Facebook and Feegli data showing the ‘interests’ of the users to be more hobby and leisure related. It is possible that there is less chance for leisure searches to only partially overlap with leisure interests, where as work overlapping with leisure interests are more serendipitous. One observation from the data,

is that people who inserted more data in the profile, and had more Facebook Likes, received more serendipitous results than others. Future work may wish to examine optimal levels of interest modeling to support serendipity.

5.1 Models and Theories of Serendipity

Our study data loosely supports both main perspectives on serendipity: the unexpected encounter [17], and the purposeful awareness of cues [20]. In both cases, however, there was element of ‘surprise’ in the process and also there was presence of ‘insight’ where user made the connection [16]. More notably, however, it is evident that some of the serendipitous events helped our participants in some less outcome-oriented ways, but still described as valuable to them. These findings further support the focus on occasions of micro-serendipity [5], which were notably present in participants’ diary entries.

5.2 Recommendations

Several key recommendations for designing systems can be drawn from our exploratory investigation: 1) evidence suggests that the model of microserendipity is worth following in design, 2) rich interest profiles are likely to highlight more potentially serendipitous results, 3) capturing work ‘interests’ may be equally as important as everyday interests, and 4) serendipitous design may be better if focused towards informational queries.

Future work may wish to focus on the influence of different design decisions on these encounters, or by integrating other sources of interest data. Unfortunately, we were not able to compare a range of possible user interface designs in our study (e.g. like [19]). Systems like Google, however, have a notable potential to examine this issue by combining topical content from +1 data, rather than specific pages, with search results. Likewise, Bing could examine this issue using their integration with Facebook. Further, Feegli used only a relatively simple technique to model and match user interests, and so much smarter methods, including those used in prior work, could be used to examine more design opportunities.

6. CONCLUSIONS

To investigate naturally occurring serendipity in web search a technology-probe search engine, Feegli, was deployed that highlighted Google API search results that related to their Facebook ‘Like’ data. Feegli was deployed to 14 people to use as their primary search engine for 7 days. 506 search queries were made by the user during the diary study and 445 results were recommended by the search engine as serendipitous. 57 distinct events were found where users clicked on the serendipitous search results. Whilst generally confirming empirically grounded models of serendipity [16], the logged behaviours, and insights from the diary entries, support Bogers & Bjorneborn’s model of microserendipity [5].

7. ACKNOWLEDGMENTS

Thanks to our participants, and the support from EPSRC Grants EP/L019981/1 and EP/M000877/1.

8. REFERENCES

- [1] P. André, J. Teevan, and S. T. Dumais. From x-rays to silly putty via uranus: serendipity and its role in web search. In *Proc. CHI*, pages 2033–2036. ACM, 2009.
- [2] P. André, J. Teevan, S. T. Dumais, et al. Discovery is never by chance: designing for (un) serendipity. In *Proc. C&C*, pages 305–314. ACM, 2009.
- [3] S. S. Bateman, C. A. Gutwin, and G. I. McCalla. Social navigation for loosely-coupled information seeking in tightly-knit groups using webwear. In *Proc. ACM CSCW 2013*, pages 955–966. ACM, 2013.
- [4] R. Beale. Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *IJHCS*, 65(5):421–433, 2007.
- [5] T. Bogers and L. Björneborn. Micro-serendipity: Meaningful coincidences in everyday life shared on twitter. In *Proc. iConference 2013*, pages 196–208, 2013.
- [6] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.
- [7] J. Campos and A. d. Figueiredo. Searching the unsearchable: Inducing serendipitous insights. In *Proc. ICCBR-2001*, 2001.
- [8] S. R. Clegg, S. Mendonça, et al. On serendipity and organizing. *European Management Journal*, 28(5):319–330, 2010.
- [9] E. B. Cooksey. Too important to be left to chance-serendipity and the digital library. *Science & Technology Libraries*, 25(1-2):23–32, 2004.
- [10] D. Elsweiler and I. Ruthven. Towards task-based personal information management evaluations. In *Proc. SIGIR ’07*, pages 23–30. ACM, 2007.
- [11] S. Erdelez. Information encountering: It’s more than just bumping into information. *ASIST Bulletin*, 25(3):26–29, 1999.
- [12] G. A. Fine and J. G. Deegan. Three principles of serendip: insight, chance, and discovery in qualitative research. *IJQSE*, 9(4):434–447, 1996.
- [13] A. Foster and N. Ford. Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59(3):321–340, 2003.
- [14] H. Lieberman et al. Letizia: An agent that assists web browsing. *IJCAI*, 1995:924–929, 1995.
- [15] S. Makri. Serendipity is not bullshit. In *Proc. EuroHCIR2014*, 2014.
- [16] S. Makri and A. Blandford. Coming across information serendipitously—part 1: a process model. *Journal of Documentation*, 68(5):684–705, 2012.
- [17] L. McCay-Peet and E. G. Toms. The process of serendipity in knowledge work. In *Proc. IiX 2010*, pages 377–382. ACM, 2010.
- [18] M. A. Meyers. Science, creativity, and serendipity. *AJR*, 165(4):755–764, 1995.
- [19] A. Muralidharan, Z. Gyongyi, and E. Chi. Social annotations in web search. In *Proc. ACM CHI 2012*, pages 1085–1094. ACM, 2012.
- [20] V. L. Rubin, J. Burkell, and A. Quan-Haase. Facets of serendipity in everyday chance encounters: a grounded theory approach to blog analysis. *Information Research*, 16(3), 2011.
- [21] A. Singh. Serendipity is bullshit. *The Stanford Daily*, 2014.
- [22] H. Von Restorff. Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, 18(1):299–342, 1933.