Discovery Is Never by Chance: Designing for (Un)Serendipity

Paul André, m.c. schraefel
Electronics and Computer Science
University of Southampton, UK
{pa2, mc}@ecs.soton.ac.uk

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

ABSTRACT
Serendipity has a long tradition in the history of science as having played a key role in many significant discoveries. Computer scientists, valuing the role of serendipity in discovery, have attempted to design systems that encourage serendipity. However, that research has focused primarily on only one aspect of serendipity: that of chance encounters. In reality, for serendipity to be valuable chance encounters must be synthesized into insight. In this paper we show, through a formal consideration of serendipity and analysis of how various systems have seized on attributes of interpreting serendipity, that there is a richer space for design to support serendipitous creativity, innovation and discovery than has been tapped to date. We discuss how ideas might be encoded to be shared or discovered by 'association-hunting' agents. We propose considering not only the inventor's role in perceiving serendipity, but also how that inventor's perception may be enhanced to increase the opportunity for serendipity. We explore the role of environment and how we can better enable serendipitous discoveries to find a home more readily and immediately.

Author Keywords
Serendipity, survey, creativity, insight, design suggestions.

ACM Classification Keywords
H.m. Information Systems: Miscellaneous.

General Terms
Design, Human Factors

INTRODUCTION
The tale of a lame, one-eyed, toothless camel [40] may not, at first glance, seem an auspicious start for ground-breaking discoveries of penicillin, x-rays, and chocolate chip cookies. However when Horace Walpole coined the word '*serendipity*' in 1754, based on the tale of The Three Princes of Serendip and the aforementioned camel, he was giving name to the *accidental sagacity* (i.e., accidental wisdom) involved in many scientific discoveries and inventions, where there is "*no* discovery of a thing you *are* looking for [40]."

C&C'09, October 26–30, 2009, Berkeley, California, USA.

Penicillin, for example, was discovered when Alexander Fleming failed to disinfect cultures of bacteria before leaving for his vacation. Upon his return he found them contaminated with Penicillium moulds which had killed the bacteria. The chocolate chip cookie was accidentally developed by Ruth Wilhelm when, depending who you believe, she either did not have the required chocolate to make chocolate drop cookies and so broke chunks of chocolate into the cookie mix instead, or was making sugar cookies when vibrations from a mixer caused bars of chocolate from a shelf above to fall into the mixing bowl.

In these examples, the discoverer was able to link together chance occurrences to arrive at a valuable insight. But people are not always capable of drawing the necessary connections. For example, when the first synthesis of copper phthalocyanine (later an important pigment and dye) was discovered, its relevance was not immediately apparent and the substance was not pursued for several years. Many other examples of potentially serendipitous discoveries missed for lack of sagacity can be found in previous work [19]. Indeed, there are two key aspects to serendipity, only the first of which is its accidental nature and the delight and surprise of something unexpected (e.g., the synthesis of copper phthalocyanine). The second is the breakthrough or discovery made by drawing an unexpected connection – the sagacity (e.g., using copper phthalocyanine as dye).

Computer scientists who have studied serendipity have seen it as a valuable part of creativity, discovery and innovation. For over 20 years, computer scientists have attempted to develop systems that deliberately induce serendipity [3, 20], and celebrated when it appeared as a side effect in systems built with other purposes in mind., for example the serendipitous discovery of something when browsing rather than searching hypertext documents [21]. However, most systems designed to induce or facilitate serendipity have focused on the first aspect, subtly encouraging chance encounters, while ignoring the second part, making use of those encounters in a productive way.

We propose that computing may be confusing the desired effect of serendipity (insight) with trying to recreate the cause (accidental finding). We hypothesize that a reconsideration of serendipity from numerous angles may help refine new opportunities for designing systems to support, if not serendipity exactly, then the desired effects of serendipitous revelation. We begin, therefore, with a deconstruction [11] of systems built to support different attributes of serendipity, particularly:

1) delightful chance encounters through subtle background systems; and

2) utilising aspects of serendipity to deliberately design explicit foreground systems to support insight, discovery and innovation.

In this deconstruction/reconstruction of serendipity, we want to achieve both an explicit review of how serendipity has been understood generally, but also particularly how serendipity has been understood within Computer Science. We are especially on the lookout for the two important aspects of serendipity discussed above: the chance encountering of information, and the sagacity to derive insight from the encounter.

Our goal is that through a more explicit understanding of serendipity and its value, we are able to identify new ways that computers might better encourage serendipity. Especially, however, we want to offer approaches to get at the desired effect of serendipity: insight. While developing systems to produce insights automatically is rather a contradiction in terms, we hope to show that we may be able to help someone optimize the opportunity for insight. These opportunities may be fostered by cross-domain comparisons [7] or, if "*fate favours the prepared mind* [26]," by facilitating light-weight efforts to enhance domain expertise which is often attributed to the foundation of seeing the serendipitous for new insights [7,15,26].

DEFINING SERENDIPITY

Let us begin by explicating our understanding of serendipity. Scattered over many different domains, there is much literature on serendipity and its importance in creativity. Medical discoveries [23,26], corporate settings [27], creativity and thinking [7,8], and historians [10] have all discussed accidental and serendipitous findings. Picasso's Blue Period has even been attributed [40] to one day finding he had blue and no other colour, inspiring him to use only blue and being intrigued with the effect.

In this paper, we focus on understanding serendipity in people's computer-based interactions. While researchers from some of the different domains mentioned above (notably corporate strategy) have suggested ways to promote serendipity and creativity, it is not until the computing and information systems literature begins to tackle the area that there starts to be a systematic exploration of, and attempts to facilitate and induce, serendipity.

Computer-based serendipity has historically most commonly been discussed in the information seeking and search literature. The definitions of serendipity used in this literature cover a broad range of chance encounters that, true to the interests of the information science research community, focus on the type of information encountered and the information activity engaged in during the encounter. Relevant research is summarized in Table 1 according to these two axes.

		Type of info. found	
		Relevant to goal	**Not relevant togoal**
Info. activity	**Directed browsing**	Serendipitous Info. Encountering [15]	Info. Encountering [13]
	Non-directed browsing		Serendipitous Info. Retrieval [38], Opportunistic Browsing [9]
	None		Involuntary Browsing [9]

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.

In the definition of digital information-based serendipity used in this paper, we choose not to focus on the task or information target involved in the chance encounter, but rather on the value the encounter provides to the person doing the encountering. For us, *serendipity* is:

1) the finding of unexpected information (relevant to the goal or not) while engaged in any information activity,

2) the making of an intellectual leap of understanding with that information to arrive at an insight.

Although some of the above definitions mention the need to link disparate entities [14], few highlight the sagacity and knowledge necessary to truly enable what we (and Walpole [40]) consider serendipitous insight. Instead, the value of information encountered by chance is described in many ways. It may reinforce an existing problem or solution or take it in a new direction [15], reject or confirm ideas [38], identify information relevant to a latent goal [9], or just plain be interesting [13]. For example, Spink et al. [34] found that unintentionally encountered partially-relevant search results, identified as "*containing multiple concepts, [or] on target but too narrow*," played an important role in a user's information seeking process and problem definition.

It's Not Just Semantics

Our discussion of the definition of serendipity, from Walpole's original to the continuum of today's usage, is not intended as a stern admonishment, recommending strict adherence to historical prescription. Rather, we fully embrace that serendipity is a word, like others in a living language, where meaning evolves and has multiple nuances. As such this term has been used to variously describe stumbling over an interesting titbit while reading a newspaper, to world-changing intellectual feats of medical discovery. In this paper, we suggest that by exploring its various uses we are able to 'deconstruct' it, in Dix's [11] sense of understanding the *attributes* of the construct in order to design for them explicitly, and how we may reproduce them deliberately to both delight and discover.

Indeed, we propose that the term serendipity itself may be ironic, in that as we will show in the work reviewed, more often than not a confluence of specific events, knowledge and attitude is needed to draw insight from chance encounters; in other words, no discovery is truly *by accident* (the *(un)serendipity* of our title). The circumstances may be termed luck, but as Gladwell [16] states, they are generally the particular advantage of experts.

In the following sections we study how serendipity has been examined, induced, and designed for. By looking separately at systems that support the two key attributes of serendipity (chance and sagacity), we are able to explore interesting new areas where computers might help produce great insight.

STUDYING SERENDIPITY

Studying serendipity appears to be even harder than defining it. Because serendipity is inherently rare, it is hard for researchers to capture or induce it for study and experimentation. In this section we examine how past studies have been designed to understand or create serendipity, and show that they have focused on the attributes (when, where, who) of chance encounters.

Studies Designed to Understand Serendipity

Erdelez [12] found that people were able and willing to discuss past experiences of serendipity, or "*bumping into information*," during discussion. She found the following elements useful in understanding such an experience:

- the *information user* who encounters the information – ranging from 'non-encounterers' to 'super-encounterers';
- the *environment* where the information encountering occurred – from libraries, bus stops, to the Internet;
- the characteristics of the *information encountered* – both problem-related and interest-related;
- the characteristics of the *information needs* that the encountered information addresses – either a current, past, or future need.

Similarly, Foster and Ford [15] found that serendipity was widely experienced amongst inter-disciplinary researchers, where it was categorised by reinforcing an *existing* problem, taking the researcher in a *new* direction, or by the location of the information: known valued information in an unexpected location, or unexpected finding of information that also proved to be of unexpected value. The study was based on naturalistic enquiry, a data exploratory approach of interviewing.

We conducted a small study to gain some insight into how frequent serendipitous encounters are in a common task: task-focused web search. We asked eight colleagues to review their search history. They were asked to examine clicked results on search result pages and report any clicked results that they deemed to be not directly task related. Each participant examined around 25 queries from 100 or more in their history. Six people reported that they did not click any results that were not directly related to their task. Three of these six mentioned that their searching at work was extremely focused and they do not allow themselves to 'wander off' as they may do at home. The remaining two colleagues did report encountering something unexpected in a search page and branching off, e.g., looking for candles for a Halloween-themed birthday cake, and broadening to look at Halloween party decorating ideas. One encounter even occurred for a very goal-directed query: getting distracted by a graduation photo album whilst looking for how to properly cite someone. It seems that people are happy to talk about serendipitous encounters, but it is rare they are able to point to, or find a specific instance. In our study, a number of participants remarked that they thought of themselves as 'serendipitous', and were surprised to find no instances of it in their search behaviour.

Facilitating or Inducing Serendipity

A small number of studies have tried to facilitate or induce serendipity in a lab setting. Toms [38] manipulated the purpose with which users approached a digital newspaper: goal directed (answering a set of questions), or no pre-defined goal (browse the paper), with two methods of access: search tool for prompting keywords, or a dynamic list of ten suggested articles. Those with a goal searched thoroughly and examined contents of articles with the intention of extracting information and moving on. Those without a goal, the serendipitous, were less concerned about meaningful content, but with coverage and exploration.

In a targeted study to induce serendipity [13], participants from a school were chosen with a common coursework task, and a new search task was reverse engineered (about buying a surfboard), so that one coursework relevant result would come up. Though 9 out of 10 participants noted the relevant result, none changed their task to look at it, highlighting the difficulty of measuring serendipity in a laboratory environment.

Though not looking to facilitate serendipity per se, in studies of relevance judgements [34] it has been shown that middle or partially relevant results may play an important role in informing and generating new directions in the information seeking process.

From the literature and our own studies, we see that if you want to understand serendipity there is value in coming at serendipity from many different directions. Serendipity is hard to induce and hard to automatically capture or identify in logs (e.g., via studies of entropy or random clicks). But by looking at it in many ways and confirming from many directions a picture of serendipity can begin to emerge.

CHANCE: IDENTIFYING SERENDIPITOUS CONTENT

Moving from trying to observe or artificially induce serendipity in a lab, in this section we examine real-world systems that have focused on supporting the 'chance' aspect of serendipity: the fortunate finding of an item. We look at previous systems that have tried to induce serendipity, elaborate on the measures that did and did not work in our own research in identifying serendipity, and provide a brief analysis. This, coupled with a discussion of how sagacity is used to connect chance encounters, allows us to explore in a later section, 'How (Else) Can a Computer Help?'

Browsing and Filtering

A number of systems to facilitate or induce serendipity take the form of an agent within a web browser. The first of these, Lieberman's Letizia [20], does not mention the term serendipity, but is described as trying to anticipate what items may be of interest to the user, using inferences from the user's browsing behaviour. No goals are predefined, and Letizia conducts a search of linked documents, providing recommendations to anticipate possible future needs.

Max [5] is an agent similar to Letizia. However, the system is set up explicitly with regard to programming for serendipity, quoting lateral thinking and stimulating the user with the precise information needed to provoke an insight as inspirations. Max is informed of domains and specific URLs of interest to the user. Max then submits queries to a search engine, with words chosen randomly across profiles (cross-domain integration has been suggested as key in producing insights [7]), and a best-first search is performed, with results e-mailed to the user. In an evaluation of 2 months, 100 messages were sent, 7 of which were considered interesting or valuable.

Mitsikeru [3] is an agent-based system to support internet browsing. It models the user's behaviour to look ahead at linked web pages and their word frequencies, using a Bayesian approach to determine relevance. It then colours links on the page depending on their relevance. In evaluation, the colouring was seen as successful, with people tending to follow the strongly advised links most of the time. There was no mention, however, of whether people found anything interesting.

The recommender systems literature has considered how going beyond pure accuracy metrics such as precision and recall may improve user experience. Herlocker [17] considers alternative measures of suitability of recommendations, including novelty and serendipity measures. Collaborative filtering systems have explored this concept. Sarwar et al. [30] alter their algorithm to recommend items that will be more preferred by a user than the population as a whole, helping users uncover less popular items they may like.

Based on a cognitive model of visual processing, de Bruijn and Spence [9] develop a behavioural model of 'opportunistic or involuntary browsing', suggesting that when a person's gaze falls upon an item of interest, a mere glance can trigger awareness of a possible solution to a problem. In evaluation of a coffee table with an embedded display, the table was found to support information acquisition through either opportunistic or involuntary browsing, therefore creating the possibility for serendipitous information retrieval.

Haiku, a system for interactive data mining [3], uses a 3D dynamic visualization with a genetic algorithm, aiming to support users in their search for relevant and interesting information. It is reported [43] that the visualization supports serendipitous discovery, for example the visualization of users internet browsing behaviour supporting the serendipitous discovery of related material.

Web Search

Serendipitous encounters can also potentially occur during directed Web search. Spink et al. [34] suggest that 'partially relevant' results can lead to the generation of new ideas and directions on the part of information seekers. We have previously reported research we have conducted into serendipity and directed search [2], and briefly summarize that work here.

To determine if web search results contained potentially serendipitous results, we conducted a study where we asked participants to rate results on two scales: Relevance (Relevant, Partially Relevant and Not Relevant); and Interestingness (Interesting, Partially Interesting, Not Interesting) (full methodology in [2]). We hypothesized that results judged as not highly relevant, but at least partially interesting are an area for serendipity.

Besides collecting relevance judgements, we were also able to obtain additional information about each query and result by examining search logs. A toolbar allowed us to collect information about how personally relevant each result was to the participant, based on 1) how similar the text of each result was to the text of desktop content, including documents and e-mail, and 2) how similar each URL was to pages in the participants' browsing history and favourites.

Findings. Twenty-one percent of all results were judged to be interesting but not highly relevant to the query they were returned for – the area we hypothesized most likely to contain serendipitous results. We further explored whether it was possible to determine which queries and which results had the most potential for serendipity:

Types of queries that are serendipitous. Each query was characterised using several features. A number of query features were found to be promising for examination with a larger sample in future work. These included whether the query was informational vs. navigational, work related vs. not, or contained person's name vs. not. Click entropy, a direct measure of how varied the result clicks are for the query, was found to be significant. That is, a positive correlation between entropy and the number of 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.

A number of query features were found to not be significantly indicative of serendipitous results, including the length of the query, the number of words in the query, the number of returned results, the number of advertisements on query results page, and the popularity of the query.

Types of result that are serendipitous. If we are able to identify queries with the potential for serendipity, if we can also identify the results within those queries that are serendipitous, then we can build a system that return such results when people are ready to receive them. We looked at a number of results features to try to identify potentially serendipitous results, including: domain, top level domain, url depth, categorization of page, url length, popularity of url, and personalization scores. The majority of these features were not indicative of serendipitous results, aside from personalization scores.

The personalization score, as previously mentioned, measures how personally relevant a result is to the viewer, based on a content and behaviour score. We found that the behaviour score in particular significantly identified potentially serendipitous results.

Analysis

Of the systems discussed, only Haiku (interactive data mining) is designed for explicit interaction towards finding interesting information. Arguably however, almost all visualization systems are designed to support such a goal: identifying interesting, but unknown, trends or patterns in data that would not have been visible otherwise.

The other systems that explicitly try to induce serendipity largely work in the background or in the periphery, colouring or e-mailing links for instance, with no significant interaction with the user needed. This corresponds to the definition of serendipity being accidental and fortuitous – if a user is forced to interact, it is hard to pretend it is accidental anymore. But are there other reasons for this focus on background serendipity? We explore two possibilities.

Role of delight. The surprise of finding something unexpected but interesting is delightful. By only working in the background, systems are able to provide many recommendations, and even if only one or two are interesting, the delight of a completely unexpected finding would be worth it, compared to a system that required interaction. Which leads us to our next point.

It's just a bad recommender system. If instead we view the induction of serendipity as a recommender system, Agent Max's [5] seven out of 100 items thought interesting could be considered a disappointing return. By not asking the user to invest any effort, or even expectation, in the system, it is free to generate recommendations that lead to nothing interesting, but be free of blame. We come back to this comparison to a recommender system in a later section.

SAGACITY: CONNECTING SERENDIPITOUS CONTENT

In this section we examine how people, using the chance information they have come across, connect their findings into serendipitous insights.

Creativity insight and serendipity literature has highlighted the role of the state of the mind of the person. Similar to Louis Pasteur's "*chance favours the prepared mind* [26]," Van Andel [40] states he agrees with Pattle [6] that discoveries are never by chance, and insists on the key role of an opening and questioning mind. These people are perhaps Erdelez's [12] so-called 'super-encounterers', encountering unexpected information on a regular basis, even counting on it as an important element in information acquisition. Rice and McCreadie [25] count knowledge about the encountered resource, and knowledge about the task the person is engaged in, as among four dimensions within serendipity and browsing. Toms [38] asserts the serendipitous encounter is influenced by the person's prior knowledge and recognition of affordances of the item. Simonton [33] and Seifert [31] both suggest that creativity originates in a preparation of mind that allows subsequent recognition of the serendipitous when it is encountered. In Csikszentmihalyi and Sawyer's model of creative insight [7], the preparation stage involves hard work and research to accumulate raw information, before periods of incubation and insight, with domain expertise vital in creative insight [28].

In considering a collection of more than one thousand examples of serendipity, Van Andel [40] suggests seventeen 'serendipity patterns' – ways in which unsought findings have been made. To illustrate the connections involved in making these serendipitous discoveries, we present a small number of examples:

Analogy. Laennec invented the stethoscope after seeing children playing. They scratched with pins on one end of a piece of wood and listened with their ears on the other end.

Successful error. The 'bad and discarded' glue, the 'temporarily permanent' adhesive on removable post-it notes, was unintentionally invented at 3M.

Inversion. McLean, looking for blood clotting factors, discovered heparin as an anticoagulant (a factor preventing blood clotting).

In the following section, we consider how knowledge from the literature into how people draw serendipitous connections can aid in thinking about future design opportunities for serendipity and discovery in ways that go beyond supporting chance encounters.

HOW (ELSE) CAN A COMPUTER HELP?

If the power of the computer is to automate a process, and that automation's effect is to accelerate practice, what can the computer offer to the seeming chance that is serendipity? So far we have mainly seen systems subtly suggesting or aiding users towards content that may be perceived to be serendipitous. That they are in the background may be because the rate of success is rare.

Is that as good as it can get, though? Some researchers believe so. Van Andel [40], for example, opines that, "*Pure serendipity is not amenable to generation by a computer.*

The very moment I can plan or programme 'serendipity' it cannot be called serendipity anymore." It may be argued that by foregrounding serendipity, designing specifically for serendipity, we remove all elements of chance and accidental finding, ending with something barely recognisable as serendipity.

We, however, are not entirely persuaded that *a priori* it is impossible to design a system that would facilitate serendipitous discovery. It may perhaps be possible for a computer searching for patterns of association or of related interest to be able to surface something that to its user would be perceived as a serendipitous discovery. We propose an automation, acceleration and aid for the first half of serendipity – the discovery of a new piece of information. The second half of serendipity – the sagacity and wisdom needed to make the connection between pieces of information – remains dependent on the human. The connections, though they may be guided, must remain unlooked for specifically to be considered serendipitous. Computer systems, however, may be able to help potential discoverers be as primed as possible to make unexpected connections in such a way that they are able to take advantage of them. Instead of treating serendipity as arcane, mysterious and accidental, we embrace the ability of computers to help us perceive connections and opportunities in various pieces of information.

Regardless of whether or not it is possible to design to generate serendipity, we see several possibilities to design for at least some aspects of serendipity. In the following sections, we propose three areas where we see critical design opportunities for creativity-focused designers and computer science researchers to collaborate with other fields to develop new tools for a change of behaviour to enhance serendipity as a foreground activity. Computers can foster serendipity by surfacing interesting connections, by providing mechanisms to enhance the expertise of the would-be discoverer to be better attuned to recognizing such connections, and by supporting means for enabling either the growth of the idea or the sharing of it so it can be developed by those more keenly interested in the connection. With these approaches, we believe it is possible to leverage the computer's function to automate processes, accelerate discovery and improve accuracy.

Better Support Chance Encounters

The majority of systems discussed earlier in this paper try to automate a chance encounter, relying upon a user's current knowledge to make a connection and make use of the content. Though we argue for a more explicit focus on insight and discovery, we believe there is real value in such background systems.

Present Serendipitous Content at the Appropriate Time

When potentially serendipitous content is presented can affect whether it leads to interest, delight or a form of discovery. Utilising attributes discussed in the previous 'Identifying Serendipity' section to extract web and search content, we suggest two ways potentially serendipitous content could be presented to the user.

1) Display at time of search. Although users do not generally look past the top few search results [37], we observed the potential for interesting and serendipitous results within the top 50 [2]. Some web search engines (e.g. cuil.com) have experimented with a grid view of results, drastically changing the concept of ranking within a list. Others (e.g., spezify.com) present media from a large number of websites in different visual ways. Within such an environment, extracting the potentially serendipitous results and displaying as part of a less-obviously ranked view could aid serendipitous encounters.

2) Personalized 'BoingBoing'. With links, web pages, papers, and search results extracted from one's interactions during the day, a personalized site could present them for navigation to in idle moments; Csikszentmihalyi and Sawyer [7] even suggest that insights occur during idle times. By explicitly requesting the content in this way, the user is also freed from potential unwanted information either in their inbox or presented as they browse.

Support Creativity and Play

Additionally, there is a long history examining the relationship between creativity and play that could be exploited to enable people to encounter information in unexpected ways, as seen in work by Russ [29] and Vandenburg [41], and more recently in The National Institute for Play (nifplay.org) and the Serious Play Conference. We note this as a key area for future work, but focus on other attributes of serendipity we have extracted from the literature in the following sections.

Varying degrees of intelligent extraction tools could be experimented with. Even pure randomness may be beneficial, as with the concept of aleatoricism, the creation of art by chance used by surrealists, writers and musicians. Similarly, the French literary group Oulipo [22] use constrained writing techniques to trigger ideas and inspiration.

Mitigate the Cost of Extra Information

However potentially serendipitous information is identified – through search results, browsing behaviour, play, or even randomly – presenting such information to users has the potential to increase the overall information the user must interact with. This can lead to two problems: distraction or overload, and the negative consequences of incorrect or problematic recommendations or assumptions

We hypothesize that in the case of a tool that merely creates recommendations, users will tolerate a level of distraction or incorrectness if there is even a minor benefit (as seen with the agent Max [5]), or they will just discontinue usage quickly. The threshold of how soon a tool produces interesting information that the user is able to synthesize, and users' frustration with such a tool remains to be seen.

Relationship to Recommender Systems

One may ask whether a system that supports chance encounters is not simply another kind of recommender system. The answer is both yes and no.

Recommender systems aim to help individuals more effectively identify content of interest from a potentially overwhelming set of choices [24], guiding the user in a personalized way to interesting or useful objects [4]. There are three approaches in these systems: content-based, collaborative, or hybrid. A content-based approach uses commonalities among things rated highly by the user in the past to suggest new items for the user, where a collaborative filtering approach recommends new items based on items previously rated by other users [1].

Clearly, a serendipity-inducing system has similar goals: recommend something interesting and unknown. However, it is the type of *unknown* and *unexpected* that sets such a system apart from recommender systems, at least as they generally exist today. Not only should it be unknown to the user, it may be unknown to almost everyone with similar interests to the user. In some sense, perhaps we want a serendipity hunter to be an *anti- or un-recommender* system: give me things that other people who have looked at this have not seen (but are related, in some way).

On the other hand, our serendipity hunting applications may be a super-personalised recommender system that not only may take into account what others have looked at who looked at this problem, but also take into account task knowledge and domain expertise, to understand what would be most likely to succeed, and at what time would that information be most beneficial. This focus on previous knowledge, (and our desire to enhance that knowledge to enhance frequency of serendipity-finding success) is related to the sagacity part of serendipity – in what context is a person best able to make the connection between the recommended item and their current knowledge.

As previously mentioned in the 'Chance' section, researchers in this space have already considered some of this work (e.g., serendipity measures [17], less popular items [30], and diversification of recommendation lists [45]). Our work here is to push on the differences and uncover potential design opportunities.

In the following section we consider how we can use the attributes of serendipity for *enhancing* our existing sagacity, towards facilitating discoveries.

Enhance Sagacity

Support Domain Expertise

It is widely acknowledged that serendipitous discoveries are preceded by a period of preparation and incubation [7]. They are, in that respect, not as 'serendipitous' as we might expect, being the product of mental preparation as well as of an open and questioning mind [15, 26]. Domain expertise is therefore considered vital, both in serendipity in particular, and in creative responses to situations in general [28]. In the example of penicillin, Fleming realised the significance of the mould killing the bacteria, but he had already carried out extensive research into antibacterial substances. While he had the favourable, indeed necessary trait of having a mind set willing to see new ideas in accidental happenings, he had the background knowledge necessary to identify what was happening in the dish as an antibacterial process rather than just a spoiled sample. We see as a complementary challenge to serendipity hunting, therefore, the enhancement of the inventor's or discoverer's own domain knowledge to enhance the likelihood of being able to make a serendipitous connection when one surfaces.

In order to enhance domain knowledge, one aspect is to track *existing* domain knowledge. In the most automatic of scenarios drawing from life logging literature, one might imagine a system that could develop a fairly comprehensive view of a person's domain knowledge. Such a system integrating heterogeneous sources such as: a (set of) courses in a particular domain, the topics covered, reading list, exam results, confidence ratings, as well as other related resources from one's own writings, publications (and perhaps rejections), would be able to calculate what the current domain knowledge may include. From this, it may be possible to derive gaps around more current literature or programs that may be if not of interest, then relevant.

In the interim of such a complete domain knowledge appliance, an assessment of one's own work in a domain via various similarity measures may help automate selection of papers from recent conferences to read.

The challenge from a design perspective may not necessarily be discovering domain literature opportunities, but defining mechanisms for presenting these suggestions in ways that are effective for the investigator. Further to creating a reading list is defining the space to deliver them opportunistically. Recently, Wilson has explored porting a conference schedule of manually selected 'interesting papers' to the iTunes music player where the abstracts of the papers are read to the listener at their leisure to prepare for the conference [42]. This experiment in repurposing usually bland information into something more interesting is one kind of mechanism that may not only enhance domain knowledge building, but if connected to automated discovery mechanisms, surface serendipitous discovery.

Google's retroactive answering of search queries [44] uses a person's search history to understand interests, recommending a URL if it addresses a specific, unfulfilled need from the user's past, where unfulfilled need may be a new URL the user has not seen before. These are just a few ideas of design opportunities motivated by a desire to design to support serendipity.

Build a Common Language Model

Another part of serendipitous discovery can be the ability to compare models across domains. Computer Science for the past decade has deliberately been working with biologists to develop new computing models informed by organic processes [39]. Here, computer scientists have very deliberately been studying biology. There are examples of such cross-domain model inspiration without one domain having to become an expert in the others. In a more accidental pairing, recently the behaviour of ants as a superorganism [18] has been seen as a potentially valuable new model to understand our brains. This comparison and contrasting of models has sparked new collaborations and much creative thought recently across science domains, via serendipitous discovery by one domain of another's model.

How might we reduce the barrier of one field discovering another field's similar and useful model, especially given that each field may have its own very different language for describing what may be very similar concepts.

Physicists, engineers and mathematicians address the cross domain specificity by using a shared metalanguage – mathematics. But even in this space, there are instances where different terminology for similar concepts means that, for example, robotics researchers miss relevant references in biology, with little chance of ever uncovering the related work.

Swanson's work on complementary but disjoint literature addresses a similar problem: two arguments may exist separately that when considered together lead to new insights, but neither argument is aware of the other [36]. His ArrowSmith system [35] is one effort towards modelling concepts to enable such connections to be drawn.

Extending the idea, could we accelerate the automation of such discovery by developing a shared semantics, a new way to abstract ideas? In the linked data domain, a key rationale for the Semantic Web efforts is to enable ontology mapping between domains, where different terms for shared contexts could be recognised. But encouraging non-ontology experts to create mappings, let alone ontologies, is a significant problem. There is a clear role for interaction design to play in developing useful and usable designs to enable concept mapping for creative, cross-discipline concept discovery.

Networks to Help Serendipity Flourish

While we posit that serendipitous tools should help develop expertise and behaviours to better identify interesting connections, what we cannot imagine doing is enabling someone to see a connection who is unwilling to step outside the blinkers of their perspective. The history of science is littered with examples of lost opportunities [19]. For serendipitous discoveries to happen, it is necessary that the person making the connection have the ability to see a connection and the infrastructure available to see that connection flourish.

For example, Ernest Duchesne documented penicillin in 1897, 30 years before Fleming forgetfully went on holiday. But his paper was rejected by the Institut Pasteur because of his youth. As a consequence, humanity would have to wait another thirty years for the person with the insight to recognize the discovery and the infrastructure to publish the finding to make it available to be mechanized for delivery as a drug.

And where Columbus may have lived his life in denial that he had accidentally bumped into anything, there were sufficient other people in his party who recognized the value of their serendipitous 'discovery' of America to take advantage of it with a vengeance.

Likewise the copper phthalocyanine of our Introduction was discovered but deemed not interesting enough to pursue at the time. The initial discoverers did not investigate the structure's potential, writing that they were, "*Busy with other investigations,... and would be glad if colleagues who are specialists... would be willing to investigate* [46]." There was no great mechanism, it seems, to share an interesting idea with a community to find a person interested in taking the idea up. Again, humanity would have to wait for 20 years for technical production.

These are all examples of failed networks. If there had been a form of network available to share interesting but uninterested-in results for others in the field to attempt a 'Eureka!' moment, the discovery of penicillin or uses for copper phthalocyanine may have happened much earlier. If only Duchesne had had a blog. But a blog is too limited as well: it assumes that someone else will actively read it. More important is the ability to publish a discovery such that those serendipity hunting agents can find it and connect it with the domain expert who may be able to make something of it, too. Here of course the ideal model would enable the idea to be set free for others to use with appropriate acknowledgement or be part of a collaboration – perhaps what Duchesne would have appreciated.

This idea again supposes a form of common language model, a way to express interest or expertise in particular areas, and a way to search for results. In some cases, it may not even be expertise that is required. For Ernest Duchesne, merely asking if someone out there has the right resources, the right connections, or the right marketing department would no doubt have been useful.

We recognize that some organizations are taking the initiative to develop such discovery networks. Eli Lilly for instance has collaborative agreements with many universities world wide to enable them to share their IP with universities, and have universities work with them. But let us suppose that these networks do work flawlessly to enable discovery of resources across it, it is a closed network. How would we design open, automated systems to guide the publication of the *shape* of an idea for the automatic detection and uptake of an idea by an idea hunting agent on another inventor's behalf?

Summary

By leveraging similar existing technology (e.g. newsreaders to aggregate content, article summarization to abstract and deliver targeted content), but coupling them to the presented creativity and serendipity ideas, we can expand them to support these processes more deliberately. In doing so, there is a higher incentive for creativity researchers and designers to collaborate more with researchers in agent based computing, Linked Data/Semantic Web, and computational linguistics to name a few. Serendipity is a new scenario around which to challenge our development and design efforts. Better applications to support serendipity, especially across disciplines, will have benefits beyond any one community.

CONCLUSION

Our contributions in this paper have been to propose that a formal consideration of serendipity, both in terms of how it is understood in the literature and how it has been adopted in computer science, enables us to think about new opportunities for designs to deliberately enhance creativity, discovery and innovation. In this review we have shown that rather than one understanding of serendipity, we have seen a kind of continuum of understandings from inadvertently finding something of personal interest, to the critical breakthrough of a domain expert making a key 'sagacious' insight between a perceived phenomena and an opportunity for a new invention.

In this spectrum, we have also demonstrated that computer science has spent most of it's design effort perhaps overly focused on trying to create insight (effect of serendipity), by recreating the cause (chance), rather than on, for instance, increasing the rate and accuracy of proposed candidates for serendipitous insight, or developing domain expertise.

Based on these observations we have proposed several design areas where we might more deliberately develop applications to enhance opportunities for making connections leading to new discoveries. Considering the history of serendipitous discovery, we see that success of serendipitous discovery is not just the find itself, but being able or willing to do something with it. Our approach has been to consider ways where we can enhance the likelihood and potential for serendipity and insight: for example, through surfacing connections, play, enhancing domain expertise, and mechanisms to share discoveries.

Each of these mechanisms, grounded in our formal investigation of serendipity, is challenging but plausible. By taking a broader view of serendipity than simply having a serendipitous insight, but by looking at also what it means to have a prepared mind and an infrastructure to support discovery, we have presented, we think, a more holistic picture of serendipity, and thus perhaps ideas that may improve the creativity, innovation and discovery process.

REFERENCES

1. Adomavicius, G.,Tuzhilin, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Knowledge and Data Engineering*. 17(6) (2005). 734-749.
2. André, P., Teevan, J., & Dumais, S. T. From x-rays to silly putty via Uranus: serendipity and its role in web search. *Proc CHI '09*. 2033-2036.
3. Beale, R. Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *Int. J. Human-Computer Studies* 65 (2007).
4. Burke, R. Hybrid recommender systems: survey and experiments. *User Modeling and User-Adapted Interaction*. 12(4) (2002) 331-370.
5. Campos, J., & Figueiredo, A. D. Searching the unsearchable: Inducing serendipitous insights. In R. Weber, & C. G. von Wangenheim (Eds.), Case-based reasoning: Workshop program at *ICCBR-2001*. http://tinyurl.com/c47c6a
6. Comroe, J.H. *Retrospectroscope. Insights into Medical Discovery.* Menlo Park, CA: Von Gehr Press p. 177. 1977.
7. Csikszentmihalyi, M., & Sawyer, K. *Creative insight: The social dimension of a solitary moment*. In R. J. Sternberg & J. E. Davidson (Eds.), The nature of insight 329–363. Cambridge, MA: MIT Press. 1995.
8. De Bono. E. *Lateral Thinking*. Penguin Books. 1990.
9. 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.
10. Delgadillo, R & Lynch, B.P. Future historians; their quest for information, *College & Research Libraries*, Vol 60,(1999) 245-259.
11. Dix, A. Deconstructing experience: pulling crackers apart. In *Funology: From Usability To Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, Eds. Kluwer Academic Publishers, Norwell, MA, 165-178. 2004.
12. 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.
13. Erdelez, S. Investigation of information encountering in the controlled research environment, *Information Processing and Management* 40,6 (2004), 1013--1025.
14. Ford, N. Information retrieval and creativity: towards support for the original thinker. *Journal of Documentation*, 55(5), 1999.
15. Foster, A. & Ford, N. Serendipity and information seeking: An empirical study, *J. Documentation* 59,3 (2003) 321--340.
16. Gladwell, M. *Outliers: The story of success*. Little, Brown, and Company. 2008.

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

18. Hölldobler, M. & Wilson, E. *The superorganism: the beauty, elegance, and strangeness of insect societies*. W.W. Norton. 2008.

19. Kohn, A. *Fortune or failure: missed opportunities and chance discoveries*. Cambridge, MA: Blackwell. 1989.

20. Lieberman, H. Letizia: An Agent that Assists Web Browsing, *Proc. of IJCAI 95*.

21. Marchionini, G., Shneiderman, B. Finding facts vs. browsing knowledge in hypertext systems. *IEEE Computer*, 21, 1, (1988) 70-80.

22. Mathews, H. & Brotchie, A. *Oulipo compendium*. Atlas Press, 1998.

23. Meyers, M.A. Glen W. Hartman Lecture. Science, creativity, and serendipity. *American Journal of Roentgenology*, Vol 165, (1995) 755-764.

24. Resnick, P. & Varian, H. R. Recommender systems. *Commun. ACM* 40, 3 (Mar. 1997), 56-58.

25. Rice, R.E., McCreadie, M.M., & Change, S.L. *Accessing and browsing information and communication*. Cambridge, MA: MIT Press. 2001.

26. Roberts, R.M. *Serendipity: Accidental Discoveries in Science*. New York: Wiley VCH. 1989.

27. Robinson, A.G., Stern, S. *Corporate creativity: how innovation and improvement actually happen*. San Francisco: Berrett-Koehler. 1998.

28. Runco, M.A., Pritzker, S.R. *Encyclopedia of creativity*. Elsevier 1999.

29. Russ, S.W. *Affect and Creativity: The role of affect and play in the creative process*. Lawrence Erlbaum 1983.

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

31. Seifert, C,. Meyer, D., Davidson, N., Patalano, A., & Yaniv, I. *Demystification of cognitive insight*, In R.J. Sternberh and J.E. Davidson (eds), The nature of insight. MIT Press, Cambridge, MA. 1995.

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

33. Simonton, D.K. *Foresight in Insight? A Darwininan Answer*. In: Sternberg, R.J., Davidson, J.E. (eds.): The Nature of Insight. Cambridge: MIT Press. 1995.

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

35. Swanson, D. R. & Smalheiser, N. R. Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. *Library Trends*, 48, (1999) 48-59.

36. Swanson, D. R. Complementary structures in disjoint science literatures. *Proc SIGIR '91*. 280-289.

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

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

39. Tsuda, S., Zauner, K. P. & Gunji, Y. P. Robot control with biological cells. *BioSystems*, 87(2007) 215-223.

40. Van Andel, P. Anatomy of the Unsought Finding. Serendipity: Origin, history, domains, traditions, appearances, patterns and programmability, *The British Journal for the Philosophy of Science*45, 2 (1994), 631--648.

41. Vandenburg, B. Play, problem-solving and creativity. *New Directions for Child and Adolescent Development*, 9 (1980), 49-68

42. Wilson, M.L. & schraefel, m.c. EPrintCast – A Document Repository Podcast. In: *The 3rd International Conference on Open Repositories*, 2008.

43. Wood, A., Drew, N., Beale, R., Hendley, B. HyperSpace: Web Browing with Visualization, *Proc. WWW'95*, 21–25.

44. Yang, B. & Jeh, G. Retroactive answering of search queries. In *WWW '06*. 457-466

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

46. Zollinger, H. *Color chemistry: syntheses, properties, and applications of organic dyes and pigments*. New York: Wiley VCH, 2003.