Expecting the Unexpected: Serendipity, Discovery, and the Scholarly Research Process

A SAGE White Paper

Alan Maloney
Senior Product Analyst, Taxonomy & Semantic Technology

and

Lettie Y. Conrad
Executive Manager, Product Analysis

Contents

Acknowledgments

The authors would like to thank all contributors to this paper. First, we are most grateful to all the students, researchers, and faculty members who completed surveys, interviews, and/or usability testing with us–your honest insights are most appreciated and always welcomed. Second, we extend thanks to everyone at SAGE who shared their thoughts, edits, and ideas that helped shape our thinking. Finally, we appreciate the time generously donated by our publishing and technology peers, in countless conversations about semantic solutions to scholarly discovery needs, namely the following:

- Ali Adair, Peer-J
- John Camarano, Scope eKnowledge
- Adrien Ginesty, Expert System
- Margie Hlava, Access Innovations
- Sam Molyneux, Meta
- Richard Padley, Semantico
- John Sack, HighWire
- Roger Schonfeld, ITHAKA S+R
- Marc Segers, GeoScienceWorld
- Richard Wallis, Data Liberate
- Alicia Warren, Wiley

Introduction

When considering researcher needs around scholarly information seeking and retrieval, the academic information industry often focuses on search—whether that's publishers indexing with a wide variety of search engines or libraries bringing Google-like search boxes to campus resources. Indeed, so much effort has been spent optimizing information systems for search that many providers consider searching for known information a solved problem. But even if that is the case, what happens when an information system is presented with a vague, fuzzy, or even unspoken information need, when users do not quite know what they're looking for? We set out to explore this question through interviews and surveys of students, researchers, and instructors across the globe about their habits and preferences with content recommendations and other chance encounters with scholarly materials relevant to their work. We then examined the information behaviour literature to better understand these responses in a wider context. Finally, we looked at what solutions were already out there and asked technologists and other publishers how information providers might support these unspoken, unplanned, and often unexamined methods of discovery, which, in this paper, we refer to as serendipity.

We found that there is a spectrum of discussion in the information studies literature: at one end, accidental discovery of unknown information is seen as a fundamental method of scholarly information seeking (Cooksey, 2004); at the other end, chance information encounters are rejected as having a useful role to play in academic practices at all (Gup, 1998). The purpose of this paper is not to take a position on that debate but to share some of what SAGE has learned about the dynamics of unplanned discovery and how information professionals can encourage this type of unplanned discovery to drive better research outcomes.

Background

Scholarly research workflows, like so many information practices, continue to evolve and change (Auclair, 2015), in part thanks to today's technological advancements and changing modes of communication. Therefore, it is critical for information professionals to keep pace with the researcher experience (Conrad, 2015) and scholarly information trends. In particular, a good, functional understanding of the diverse and ever-changing approaches to information seeking can have an impact on the quality of services provided by libraries, publishers, technologists, database providers, and other information professionals, as well as on the overall successes of those organizations.

With these goals in mind, SAGE Publishing launched this report to share the findings and outcomes of an initiative to attend to a greater variety of interdisciplinary information-seeking needs of our readers, which we believe ring true for many types of academic users. Building on what we know about search and browse behaviours, as well as cross-sector opportunities for enhancing scholarly discovery,[1] we set out to better understand other instances of discovery. In particular, to support in-context discovery of related materials, we wanted to understand our readers' needs and expectations for unplanned encounters with materials across the academic information ecosystem. We asked, "What role does serendipitous discovery play in the scholarly research process today?"

If "big data" is information that has overwhelmed a community's ability to deal with it (Choudhury, 2013), SAGE has a big data challenge even with its own content. This is not to mention its readers, who may be unaware of entire products, content types or corpora. For SAGE, an academic publisher with a 50-year history in social science books and reference, alongside journals across the disciplines, digital publishing technologies provide tremendous

opportunity to synthesize millions of pages of scholarly research and teaching texts. With more than 1.7 million journal articles, book chapters, and reference entries online, not to mention a growing amount of multimedia, numerical data, and other miscellaneous formats, accounting for all of the information that even this single publisher has becomes a challenge.

How, then, can we present and shape an information environment that is too large to be completely knowable? Reflecting our research into "chance" academic content discovery, this paper looks at one possible solution in the form of automated semantic recommendations, how SAGE has approached this challenge, and how this technology can facilitate one of the most important yet most elusive facets of discovery: serendipity.

Methodology and Demographics

This paper summarizes the results of SAGE research and development efforts in 2015, including user-experience (UX) research, semi-structured interviews, and surveys alongside reviewing the relevant published library and information science literature. Participation invitations for the surveys, interviews, and UX tests were distributed by e-mail to undergraduate and upper-level students and faculty members who signed onto the SAGE mailing list. Survey participants were entered into drawings for either $20 or $100 gift cards; interviews and testers were thanked for their time with $5 gift cards.

The UX research focused on modified usability testing of SAGE prototypes, with semi-structured interviews following talk-out-loud exercises. In total, 12 tests took place in February 2015, followed by ongoing usability tests during development in the spring and summer of 2015. All test participants were either faculty members or undergraduate and upper-level students in the social sciences, largely based in either North America or Europe. Additional interviews were conducted throughout the summer of 2015—with three faculty members in the social sciences, four publishing experts, and five technology suppliers—all in either North America or Europe.

We strived for as broad a reach as possible in the surveys and received responses from 87 undergraduate students and 152 faculty members/instructors. The undergraduate respondents represent an international sampling, with 53% from North America, 18% from Asia/Asia Pacific, 17% from Europe, 6% from South America, 5% from the Middle East, and 1% from Africa. Faculty responses were more heavily received from North America (70%), with 13% from Europe, 7% from Asia/Asia Pacific, 5% from Africa, 3% from South America, and 2% from the Middle East.

Many discipline perspectives are represented as well. The majority of survey responses (60% of faculty and 55% of students) were received by scholars in the social sciences. Responses were also received by faculty in the humanities (13%), the medical/health sciences (20%), and science, technical, or mathematical (STM) fields (7%); responses were also received from undergraduates declaring studies in the medical/health sciences (15%), STM (14%), humanities (9%), and the arts (7%).

Serendipity, Discovery, and Scholarly Research

Overview of Information-Seeking Behaviours

When considering exactly where serendipitous discovery fits into the modern research process, considering the wider context of information-seeking behaviours and techniques

may be helpful. In this context, serendipitous methods can be seen as one of several modes of discovery that might be used in tandem by a researcher. Delineating these modes of discovery is still an active conversation in the literature, with many competing theories and denominations of the constituent parts of information seeking within the research process (Russell-Rose & Tate, 2013). Some models use the granularity of the research need as the paradigm, placing individual queries at one end of the spectrum and overall research strategies at the other (Bates, 1990). Some models focus on the intent of each research action, whether it be comparing, assessing, or interpreting (O'Day & Jeffries, 1993).

Nevertheless, something each of these models has in common is the distinction between finding and exploring: between the need for known, well-defined things and fuzzier, less-known information. Donna Spencer (2006) further divides each grouping to define a four-part categorization of user behaviours (see Figure 1). *Finding* is characterized as either known-item searching (where users know exactly what they want and where to look) and refinding (where users know exactly what they want but do not necessarily know where to look); different types of *exploration*, on the other hand, are characterised by the stability of the user's information need, with Spencer identifying a special type of exploration in which users do not even know how to articulate their information need or where that information need changes in response to new findings. Here, we see the role for serendipity to lend assistance to the academic reader.

Whichever model we believe best fits the students and researchers we serve, each of these information strategies calls on different techniques and methods for discovery, all of which we need to support successfully. Progress has been faster in some of these areas than others. For example, when searching for a known item (*Finding* in Figure 1), if researchers have a citation and are aware that something is out there, they can find it with a high rate of success. *Exploring* is also traditionally well served by current information resources, which commonly allow researchers to browse subject hierarchies or other conceptual representations of a topic to refine a vague information need. These types of discovery entail an important relationship

Figure 1   Types of information seeking

	User knows where to look	User does not know where to look
User knows what they want	Finding	Refinding
User does not know what they want	Exploring	Serendipity

Source. Based on Spencer (2006).

between publishers and academic libraries. Of those we surveyed, 68% of faculty members and 70% of undergraduates rated searching library databases as "very important" when discovering new scholarly information—more important than any other resource. From here the researcher has several long-established information-seeking methods available to them, such as citation-based discovery or "pearl growing" as part of systematic literature reviews (Tucker, 2015).

Citation-based discovery occurs when a reader follows cited works from relevant papers to discover additional related items that might be important to his or her current information need or research question. Abstracting and indexing databases have made this method more effective, for example, with easy creation of citation networks. The principle that the citation is accepted as a mark of quality and relevance remains unchanged: If the author has seen fit to cite a paper, it must somehow be a source of relevant information. This practice can have a recursive effect on discovery, however. If a paper is discovered, and then cited again just because it has already been cited, then its popularity and success is self-perpetuating, often out of proportion to its relevance to the information need at hand.
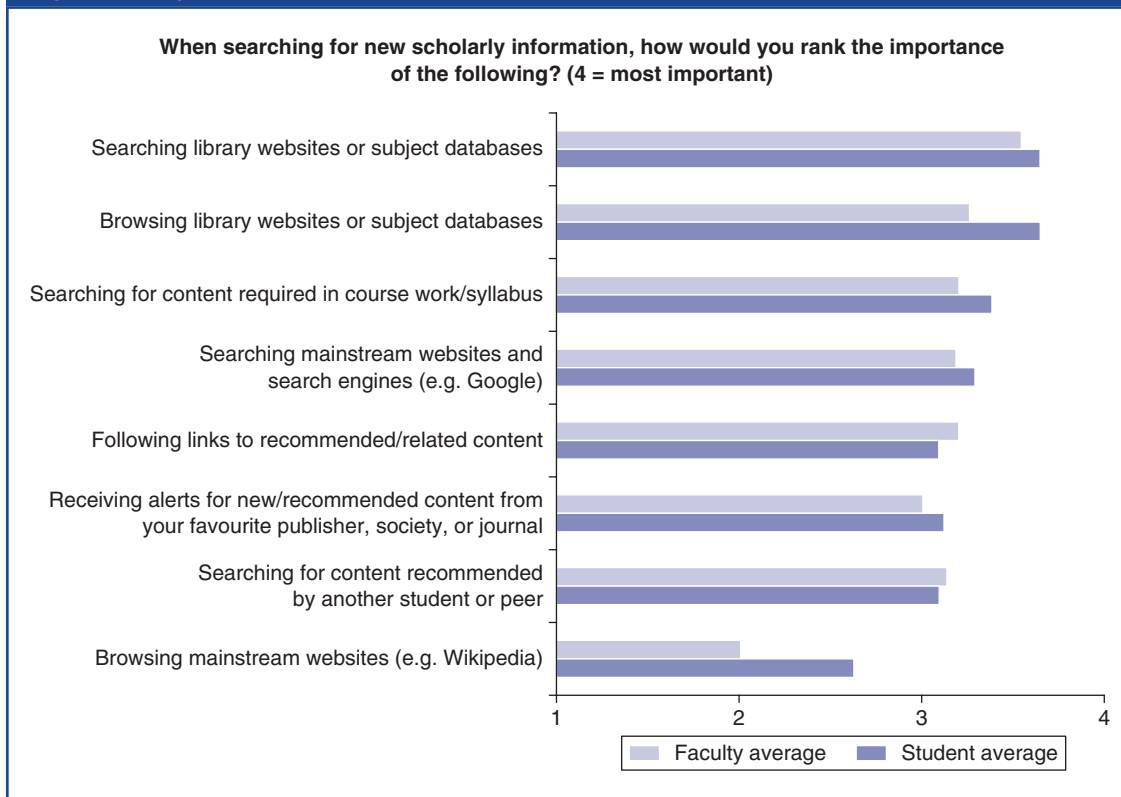
Bibliometric studies have established that the citation count of an average paper rises exponentially until around the third year of publication, though the introduction of alternative discovery methods, such as social media, can disrupt this cycle (Muglia, Lea, & McDonald, 2015). As the researcher only has a finite amount of time for discovery, the gravity of highly cited papers can reduce exposure to less popular (and perhaps more interesting) information. This effect can be seen elsewhere, for example, in the "most-viewed" sections in online news websites (Yang, 2015). This, together with the trade-off between item popularity and recommendation accuracy (Steck, 2011), means popularity-based strategies are an often criticised method of discovery, yet, for completely understandable reasons, they remain protocol in scholarly research.

Systematic search and browse is another common method of exploration: As well as walking through citations, researchers also draw commonly on review articles, bibliographies, and journal tables of content to establish the current state of the literature and then set up saved searches and alerts to remain abreast of new developments. However, this systematic method of exploration requires the researcher to have a priori knowledge of the relevant journals, keywords, and other resources, which can be particularly challenging for undergraduates as well as for those in emerging fields of study.

Paying attention to the role that serendipity plays in the research process might enable information professionals to help researchers overcome some of these barriers. Introducing some informed randomness into the discovery process, for example, may help researchers break the cycle of self-perpetuating, citation-based popularity or discover new and emerging concepts and keywords, promoting the discovery of information that may otherwise have been overlooked or even uncited. This is placed in even sharper relief by the exponential increase of scholarly information, as focusing on a small number of familiar sources becomes proportionally even more limiting (Cooksey, 2004), especially in interdisciplinary areas where relevant findings may be dispersed across different facets of the humanities and the sciences.

More fundamentally, information providers should think about how to handle unexpressed, unarticulated information needs, largely because this is the reality of how research is carried out. Our research shows that a majority of undergraduates (78%) and faculty members (91%) are inclined to click on links to recommended or related content during the course of their online research. Web analytics for any information resource will establish that these academic users rarely perform a single search and have their need satisfied by a single, ideal set of documents. Instead, researchers reframe their information needs as they go, based on

Figure 2 Importance of various information sources

When searching for new scholarly information, how would you rank the importance of the following? (4 = most important)

Searching library websites or subject databases

Browsing library websites or subject databases

Searching for content required in course work/syllabus

Searching mainstream websites and search engines (e.g. Google)

Following links to recommended/related content

Receiving alerts for new/recommended content from your favourite publisher, society, or journal

Searching for content recommended by another student or peer

Browsing mainstream websites (e.g. Wikipedia)

1 2 3 4

■ Faculty average ■ Student average

information they encounter. Not only are information needs fluid and changeable; they are also not always satisfied intentionally.

Furthermore, serendipitous discoveries are enjoyable: creating user satisfaction and increasing engagement with a publisher's resources. Although the business case for recommendation engines in retail is obvious—despite Amazon's stated objective with product recommendations is simply "to delight our customers by allowing them to serendipitously discover great products" (Mangalindan, 2012) —the case for recommendation engines in academic publishing has become a natural extension of our responsibility to disseminate scholarly research and knowledge.

A Brief History of Serendipity

Before considering how information providers can promote and encourage serendipitous discovery, it may be useful to consider the concept in wider focus and break it out into its constituent parts.

Throughout the literature, the concept of serendipity is defined by two key components: *accident* and *sagacity*. The latter aspect is sometimes lost, to the objection of researchers on creativity and innovation, who consider any definition of serendipity as a mere happy accident to only tell half the story (Ferguson, 1999). For example, when Alexander Fleming encountered a contaminated petri dish, an unprepared mind might have thrown the dish away rather than recognizing that the substance that grew on the dish had killed the bacteria on it and conceiving antibiotics. By contrast, when copper phthalocyanine was first synthesized by accident in 1927, its useful properties were not considered, and several years passed before

its use as a dye was realized. Few would consider the latter event to be serendipitous, even though it was a fortunate accident, because the sagacity was lacking. This bilateral conception of serendipity has survived into modern information studies literature.

It is important for information professionals to be mindful of both of these aspects, as "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" (Andre, Teeven, & Dumais, 2009). One could argue that this is only natural and that it is the information professional's job to point researchers to authoritative and potentially relevant information, but ultimately, the researcher's job is to make sense of that information and to turn it into insight. However, whether publishers and libraries can do more to prepare the minds of their readers and patrons for the chance discovery is an important question.

This question begins to touch on the domain of information literacy, which, among other things, often focuses on the economics of the scholarly publishing industry, digital literacy, the evaluation of source material, and how students and researchers can make the most of their relationship with the library (Association of College and Research Libraries, Working Groupon Intersections of Scholarly Communication and Information Literacy, 2013). The ability to spot new connections and insights when encountering new material, on the other hand, is more often seen as a role for academic departments in terms of equipping students with the expertise and knowledge required to turn research into insight. As elusive as this quality is, publishers have a role to play in making sure not just that the right information is presented at the right time but also that the material is presented and contextualised in such a way— or that the overall researcher experience is shaped in such a way—that making these connections is easier.

Unplanned Discovery in Context

Systems designed to facilitate serendipity are increasingly common in many industries—quite prominently in online retail, where inducing a customer's unarticulated or previously unknown need is potentially lucrative. Amazon, for example, draws on data from customer behaviour (both individual and collective) as well as manually curated content similarity to "integrate recommendations into nearly every part of the purchasing process" (Mangalindan, 2012). With more than 200 million products on sale in the United States alone, and 153 million page views per day, the pool of data that Amazon can draw on is remarkable and illustrates the need for recommendation systems that, by helping the customer negotiate such overwhelming choice, reduce big data to small data. Amazon and other retailers are also able to draw on decades' worth of insights from consumer psychology to maximise engagement, accuracy, and, by turns, revenue.

But do these mechanisms apply to scholarly publishing? Most important, in comparison to mainstream e-commerce, academic information professionals have a primary obligation to the scholarly research enterprise, and our underlying goals of providing discovery tools are to support an effective and efficient research experience—these goals are often prioritized over commercial interests. Moreover, practically speaking, no academic resource (other than, arguably, Wikipedia) receives as many page views as the major online retailers, which also gather personal details as a by-product of every transaction. The pool of data on which information providers can draw is, by comparison, small. Additionally, insights from consumer psychology tend not to consider the unique environment of scholarly publishing and research—and so knowledge about the psychology of the consumer dwarfs knowledge about the psychology of researchers. Finally, the business case for these recommendation

systems is also challenging to prove with hard metrics. Often, drawing a direct link between increased discoverability and the bottom line is difficult because this connection is filtered and complicated through the lens of usage, where the commercial interactions are indirect. Publishers and libraries alike face an uphill battle in setting out the case for investment in discovery.

In this respect, the situation for academic publishers and libraries is perhaps less akin to Amazon and Apple and more akin to Spotify, Pandora, or Netflix, where the business case is driven by discovery and usage of existing content rather than by increasing shopping cart revenue. These services have drawn on both behaviour-based analytics and content-based methods to encourage serendipitous discovery on their platforms. Spotify, for example, draws on user behaviour to connect listeners with music liked by people with similar taste. This powerful method inherently creates a feedback loop. A successful recommendation will be listened to again and again, further refining the algorithm's sense of the user's tastes and improving the accuracy of recommendations and giving the listener an implicit control over his or her own recommendations. Pandora, on the other hand, focuses much more on content rather than behaviour, classifying each song in its database against its ontology of musical attributes, such as "level of distortion" or "type of background vocal." Netflix adopts a similar approach for movies and TV, allowing the system to show viewers, among other things, "films with a strong female lead."

One thing all these services have in common is a certain consistency in the types of content they provide, whether it is music or movies. This approach has its advantages: For publishers of any kind, having a homogenous inventory of products enables extremely detailed "fingerprinting" of each individual object, allowing for a finely grained comparison and classification of content and enabling useful filters and features for readers. This situation is often as true for publishers of academic content as it is for publishers of other media, so not only are academic publishers in a similar situation to these mainstream media services when it comes to proving an indirect business case; there are also opportunities for sharing knowledge and learning about approaches, technology, and the user experience.

But even when the mechanisms tested in the wider world do not apply to scholarly communication, information providers being aware of this wider context for our readers is important. Although e-commerce sites are not usually direct competitors to libraries and academic databases, we share a user base whose expectations of online discovery are being set by the progress made in these other industries. Or, as Schonfeld (2015) puts it, "researchers' expectations are being set not by improvements relative to the past but rather by reference to consumer internet services."

Mechanisms of Serendipity in Academic Research

One way for information providers to address the relative paucity of user data collected from academic researchers is to recognise this inherent limitation and focus efforts on approaches to serendipity that play to their own strengths. This is perhaps why we are more accustomed to seeing content-based or popularity-based recommendations on academic information resources, where we might see behaviour-based or social-based recommendations on e-commerce or social networking sites. If there is one thing an information provider has, it is insight into its content, and so it is far more common to see resources that recommend materials based on their conceptual similarity rather than materials that similar users read. But if information providers are drawn to content-based methods partly out of necessity, it is perhaps serendipitous for them that this appears to be an approach that users respond to.

Solutions in Serendipity

Case Study A: Meta

Meta (originally ScienceScape) started from founder Sam Molyneux's experience as a scientist who needed to be able to identify areas of research in high-impact fields, including areas not yet known. Mapping the state of an ever-changing field through traditional methods such as systematic review is challenging, so Meta brings the same network principles to scientific research that services such as Spotify and Facebook apply to other content. Meta starts with machine intelligence, organizing content against entities such as concepts, people, places, and institutions. The resulting knowledge graph is used as the basis for providing the user with content recommendations, but even that can present the user with an overwhelming amount of information. Meta's solution is to predict what will be of interest to the user based on their network, all the while preserving what Sam considers the special quality of scholarly papers that set them apart from other content like music and movies: "Academic papers are linked; they're part of a global conversation."[2]

Case Study B: Semantico

Semantico is a supplier of digital publishing solutions to the scholarly and professional market. Through its platform Scolaris, and its access and identity management system SAMS, Semantico is focusing efforts on analysing user visits and profiling what content individual users are accessing. The semantic profile of each item a user reads can be rolled up to build a detailed profile of their interests and suggest content that they already have access to, but may have missed, or even to potentially offer free samples of restricted content. For chairman and CEO Richard Padley, academic content is not only a special case because it is not mass market, but because of its distinctiveness: "All pieces of academic content are unique in their own right, whereas consumer products are innately more comparable."[3]

Case Study C: Access Innovations

For Marjorie Hlava, president and founder of the information management company Access Innovations, serendipitous search is an important mode of discovery but one that might ultimately be an illusion: Search systems can often "feel serendipitous while delivering high levels of contextually relevant results."[4] This form of discovery still must be precise because users do not tolerate high levels of irrelevance in any

(Continued)

When we surveyed our users, asking what motivates them to click on recommended links, the apparent relevance of the link was by far the most important factor, with the vast majority of respondents considering the relevance or interestingness of the recommendation itself important (see Figure 3). A distinction is made between relevance and interestingness, which is why we asked about "interestingness" separately. Interesting links might not necessarily be relevant to the current information need, but they may be relevant to a different one, which meets precisely the definition of serendipity from the *International Encyclopedia of Information and Library Science*: "information encountered whilst not actively being sought at the time" (Feather & Sturges, 2003). Here it seems that interestingness is a good bellwether of serendipity because none of the other information-seeking strategies—known-item searching, saved searches, and so on—necessarily allow the user to facet their search by interestingness and therefore purely interesting content may be missed.

In addition to an inclination to content-based discovery, our users also showed disinclination to some of the other alternative methods. If, indeed, user expectations are being set by consumer Internet services, there is apparently room for negotiating exactly what method to use to meet them. Behaviour-based and popularity-based methods, for example, were the least trusted

source of recommendations by our users, with fewer than 20% of users motivated by links listed as "also read" by other users (see Figure 4). Also, some participants had additional suspicions that behaviour-based recommendations have ethical implications: One faculty member summed up her concerns with behaviour-based recommendations that draw on browsing history as "often both creepy and unhelpful."

Distinct from behaviour-based recommendations are those that draw on personal networks or social principles. We have learned from student and faculty survey respondents that recommendations from peers or instructors were seen as more important and trustworthy than were recommendations based simply on either their own or others' online behaviour. Some

set of search results. Semantic enrichment resources, such as taxonomies or thesauri, must be integrated from start to finish, from content acquisition through to the search system itself, to achieve reliably relevant results.

Case Study D: SAGE Recommends

In late 2015, leveraging the findings of the research outlined in this white paper, SAGE launched a new discovery feature that provides users with links to related SAGE content for extending and supporting their academic research journey. SAGE Recommends suggests further reading within the context of what a user is currently viewing, referring readers to related SAGE journals, data sets, and other reference materials. The feature brings together the knowledge mined from SAGE's 1.5 million journal articles, more than 250,000 book chapters and reference works, and much more. SAGE Recommends enables a serendipitous research experience across all of SAGE's award-winning content platforms, leading students and faculty to uncover relationships between topics across the social science disciplines. For more, see "The Story of SAGE Recommends" at https://us.sagepub.com/sites/default/files/SAGERecStory.pdf.

providers strive for a blended approach, adding a layer of network intelligence onto content-based relevance (see the "Solutions for Serendipity" sidebar). Meta (originally ScienceScape), for example, allows users to easily share recommendations and has plans to filter and prioritise related content suggestions by looking at users' contacts and whom they follow—in other words, training content-based recommendations with assumed trust within a reader's network.

Despite the opportunities for hybrid approaches, based on our research, it appears that approaches to encourage serendipity that do not place the content front and centre might encounter problems. This phenomenon is interesting because providers in other fields, such as music, have successfully developed serendipitous discovery tools that largely disregard aspects of the content itself. Often it is said that music is a personal experience, but perhaps academic research is even more so—and therefore less susceptible to fashion, or to network, social, or behavioural effects. If research can be a personal, intimate experience, then finding that some users might consider the behaviour of some information systems to be unacceptably "creepy" if it is seen to be too intelligent or intrusive is not surprising.

The over-personalization of the research experience can have other unintended consequences that are antithetical to serendipitous discovery. If discovery is too exacting or too precise, it can end up reinforcing habits rather than exposing students and researchers to new information, sharply limiting the researcher's view of the world of information. Eli Pariser (2011) observed this phenomenon as the "filter bubble," arguing that information providers should take care not to separate users from contrasting viewpoints. Besides the implications of the filter bubble on civic discourse, there is also simply the effect on the aforementioned interestingness of the information that a user is exposed to. Predictable discoveries are not usually very useful, or very interesting: suggesting to someone who has just watched *Home*
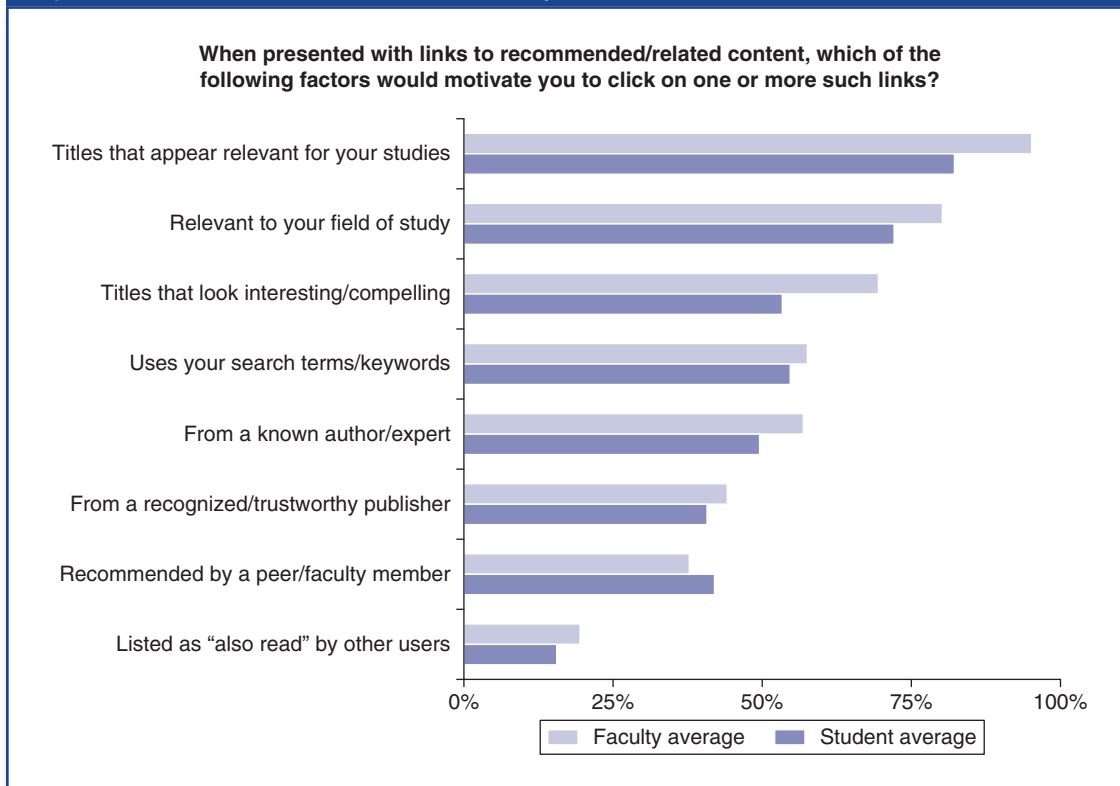
When presented with links to recommended/related content, which of the following factors would motivate you to click on one or more such links?

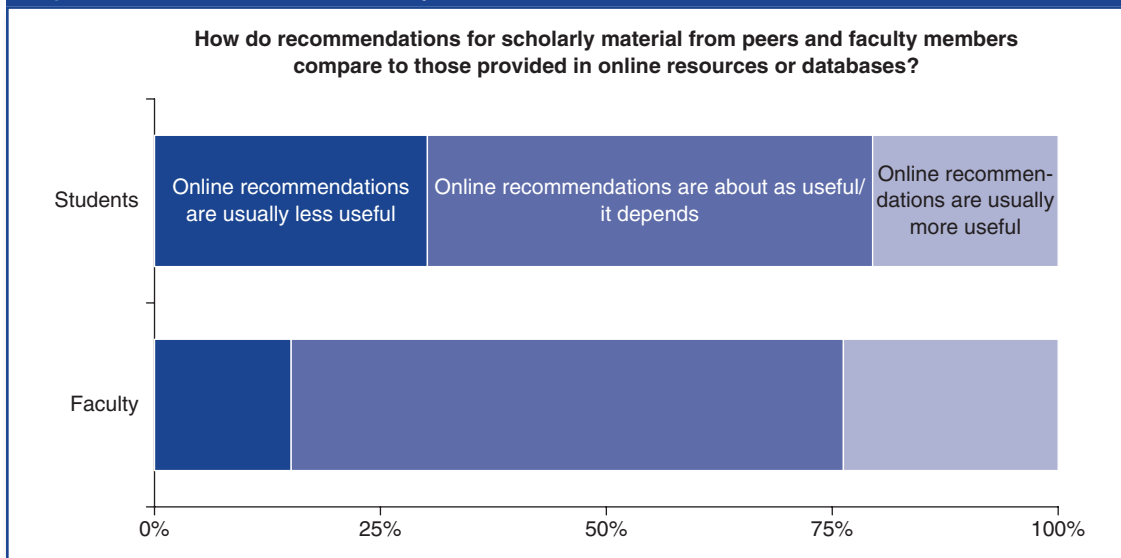|	
Faculty average	Student average

Alone that they should watch *Home Alone 2* is a solid recommendation but one that promotes convenience rather than discovery. Or, as one of our students noted on the survey, he or she wouldn't click on a recommendation "when the title is almost exactly similar to my search terms."

On the other hand, a content recommendation that is based on a false positive can still direct users to interesting content. We might even suggest that there is room for errors and luck in recommendation systems; a serendipitous system that does not include some element of chance is hardly serendipitous at all. Or, if not pure chance, a serendipitous recommendation system should at least anticipate what the user would have found using other search strategies, as well as offering something the user might not have necessarily found by other means. In one interview, a researcher noted that he generally sees online recommendations, such as favourite-journal e-mail alerts, as offering a "reactive stance to research," in contrast to those moments when he is "more proactive in my approach to looking for relevant literature."

At this point we might reflect that we have been doing an awful lot of talking about the first aspect of serendipity—the chance encounter itself—and not much about the second, equally important aspect: the transformation of that encounter into some kind of insight. It is fair to say that the web, in general, has been slower to catch on to this second aspect of discovery, but the move from encounter to insight is now visible in how the architecture of the web is changing. We may argue about whether Google is moving away from, or returning to, the presentation of the "10 blue links" of Google search results that revolutionized web-based information seeking at the end of the last millennium. But clearly more context and insight are being built directly into search responses, such as an info box presenting structured context

Figure 4   How students and faculty members evaluate content recommendations

How do recommendations for scholarly material from peers and faculty members compare to those provided in online resources or databases?

Students	Online recommendations are usually less useful	Online recommendations are about as useful/ it depends	Online recommendations are usually more useful
Faculty			

0%   25%   50%   75%   100%

and answers directly within search results or a semantic search that queries a knowledge graph to quickly present the user with an answer and links to related concepts. Whether this encourages insight and connections for the end user has yet to be proven, but it is encouraging that attempts are being made to allow users to act more quickly and more meaningfully on the information that they encounter.

Conclusion

The importance that serendipity plays in the scholarly discovery process is often overlooked, and there is room for much more to be said about the topic than one paper can address. However, this facet of discovery should not be neglected in favour of more predictable, easily scrutinised activities—such as quick search and browsing—given today's researcher workflows

Recommendations for Serendipitous Discovery

- Serendipity is just one means of discovery that might be used in tandem with others by a researcher, but it is often overlooked, despite its being part of the reality of how research is performed. Information providers should consider this aspect of discovery in their user personas and user stories to stimulate internal discussion on how to handle unexpressed, unarticulated information needs.

- The role for publishers to play is making sure not just that the right information is presented at the right time but also that the material is presented and contextualised in such a way—or that the overall researcher experience is shaped in such a way—that making the leap from discovery to insight is easier. In the wider context of the web, attempts are being made to allow users to act more quickly and more meaningfully on the information that they encounter.

- Information providers should play to their own strengths and focus on gaining insight into their own content to deliver a better researcher experience.

- Serendipitous discovery should be of particular interest to information providers precisely because there is so little precedent; there is still tremendous scope for individual organizations to bring their own priorities and values to bear on how they recommend or otherwise help researchers discover their content.

and the potential benefits to discovery and usage when directly serving a diverse range of researcher information needs. Perhaps serendipitous discovery should be of particular interest to information providers precisely because there is so little precedent; there is still tremendous scope for individual organizations to bring their own priorities and values to bear on how they recommend or otherwise help researchers discover their content.

For SAGE, these values are threefold. First, organizations whose core competence is scholarly content, such as academic publishers, should focus on making the content king in all aspects of their operations, including recommendations, serendipitous discovery, and the content enrichment processes that underlie them. It is a serendipitous coincidence that our readers seem to prefer this approach to methods that rely on behavioural or social network analysis. Second, SAGE has always believed in the power of interdisciplinary research; therefore, our approach to serendipity tries to draw on relationships from other fields by drawing on ontological relationships between concepts and promoting some fuzziness in the algorithm (Maloney, 2016). Third, we believe that consideration of the user is as important as what goes on behind the scenes, which is why the same amount of research goes into the design and presentation of our interfaces as to the content enrichment processes underneath them. There is a time and a place for discovery, and potential insights will be overlooked if publishers and libraries alike do not deeply understand academic users' workflow and the full range of their information needs.

However, in spite of the potential benefits to scholarly discovery of individual content enrichment efforts, the need to consider the bigger picture is clear, especially when considering what happens *after* discovery. The less exciting, but equally as important, corollary to discovery is *delivery*, or access: providing the patron with the material once they have found it. Given that "the researcher's discovery-to-access workflow is [already] much more difficult than it should be" (Schonfeld, 2015), improving discovery before solving the challenges of infrastructure and access is perhaps kicking the can down the road. This is not to say that there is no value to tools and solutions that promote discovery within an isolated silo, but their potential is limited until publishers, libraries, and discovery vendors make interoperability a priority.

Notes

1. See the SAGE Discoverability white paper series at https://us.sagepub.com/en-us/nam/sage-white-papers.

2. Interview with Sam Molyneux, September 2015.

3. Interview with Richard Padley, August 2015.

4. Interview with Majorie Hlava, August 2015.

References

Andre, P., Teevan, J., & Dumais, S. T. (2009, October). *Discovery is never by chance: Designing for (un) serendipity*. Paper presented at the C&C '09, Berkeley, CA.

Association of College and Research Libraries, Working Group on Intersections of Scholarly Communication and Information Literacy. (2013). *Intersections of scholarly communication and information literacy: Creating strategic collaborations for a changing academic environment*. Chicago, IL: Association of College and Research Libraries.

Auclair, D. (2015). *The changing state of researcher workflow* (White paper, pp. 1–30). Burlingame, CA: Outsell, Inc.

Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing and Management, 26*(5), 575–591.

Carr, P. L. (2015). Serendipity in the stacks: Libraries, information architecture, and the problems of accidental discovery. *College & Research Libraries*, *76*(6), 831–1842.

Choudhury, S. (2013, December 4). *The research data revolution*. Paper presented at STM Innovations Seminar 2014, London, UK.

Conrad, L. Y. (2015, April 21). Guest post: Lettie Conrad on understanding the researcher experience. *The Scholarly Kitchen.* Retrieved from http://scholarlykitchen.sspnet.org/

Cooksey, E. B. (2004). Too important to be left to chance—serendipity and the digital library. *Science & Technology Libraries*, *25*(1–2), 23–32. Retrieved from http://doi.org/10.1300/J122v25n01_03

Feather, J., & Sturges, P. (Eds.). (2003). *The international encyclopedia of information and library science* (2nd ed.). London: Routledge.

Ferguson, A. (1999). The lost land of Serendip. *Forbes*, *164*(8), 193–194.

Foster, A., & Ford, N. (2003). Serendipity and information seeking: An empirical study. *Journal of Documentation, 59*(3), 321–340. Retrieved from http://doi.org/10.1108/00220410310472518

Gup, T. (1998). Technology and the end of serendipity. *The Education Digest*, *63*(7), 48–50.

Maloney, A. (2016). *The story of SAGE Recommends* (White paper). Thousand Oaks, CA: SAGE Publishing. Retrieved from https://us.sagepub.com/sites/default/files/SAGERecStory.pdf

Mangalindan, J. P. (2012, July 30). Amazon's recommendation secret. *Fortune*. Retrieved from http://fortune.com/2012/07/30/amazons-recommendation-secret/

Muglia, C., Lea, B., & McDonald, J. (2015, November). *To tweet or not to tweet? That is only the first question.* Paper presented at the Charleston Library Conference, Charleston, SC. Retrieved from http://www.slideshare.net/CarolineMuglia/to-tweet-or-not-to-tweet-that-is-only-the-first-question

O'Day, V. L., & Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. In *Proceedings of the INTERCHI conference on human factors in computing systems (CHI'93)*. Amsterdam, the Netherlands: IOS Press.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York, NY: Penguin Press.

Russell-Rose, T., & Tate, T. (2013). *Designing the search experience: The information architecture of discovery.* Waltham, MA: Morgan Kaufmann.

Schonfeld, R. C. (2015, March 26). Meeting researchers where they start: Streamlining access to scholarly resources. Retrieved from http://sr.ithaka.org?p=241038

Spencer, D. (2006). Four modes of seeking information and how to design for them. Retrieved from Boxes and Arrows website: http://www.boxesandarrows.com/view/four_modes_of_seeking_information_and_how_to_design_for_them

Steck, H. (2011). Item popularity and recommendation accuracy. In RecSys 11 Conference Committee, *Proceedings of the fifth ACM conference on recommender systems (RecSys'11)*. New York, NY: ACM.

Tucker, V. M. (2015). Sharpening the search saw: Lessons from expert searchers. *iSchool Student Research Journal, 5*(1), 6.

Yang, J. (2015). Effects of popularity-based news recommendations ("most-viewed") on users' exposure to online news. *Media Psychology*. Advanced online publication. doi:10.1080/15213269.2015.1006333