Making your own luck: Tales of the unexpected in the search user interface

Paul Cleverley and Simon Burnett, Robert Gordon University, UK

Valuable discoveries

Serendipity (fortuitous discovery) has been responsible for some of the greatest scientific advances, from the discovery of Penicillin to the X-ray and Cosmic Microwave Background (CMB) radiation. The business world is also littered with these 'happy accidents', including the oil and gas industry. From the identification of new oil fields through to new completion engineering techniques. Serendipity may be responsible for leaps in value that cannot be predicted. Casting a wider net, we can all probably think of examples where we have serendipitous encounters as part of our daily lives. There are specific criteria for an event to be called serendipitous. It must be unexpected, insightful and valuable, which are subjective judgements creating the 'serendipity space'.

There is much debate about the extent to which serendipity is actually random. Certain people appear to have serendipitous encounters more than others, with some researchers suggesting that serendipity favours the *prepared mind* and *information rich environments*. Whilst it is unlikely that the phenomenon of serendipity can ever be controlled, it may be feasible to identify certain aspects that, if facilitated effectively, act as catalysts for increases in serendipitous opportunities.

Search Interface

The classic Internet search engine, digital library and their cousins deployed behind the firewall of companies (enterprise search) have traditionally focused on precision. Returning the 'ten blue links' concept or some derivation thereof. The rationale being as long as the specific web page or document you were seeking is on that first page, it does not matter how many results are returned. This approach has been incredibly successful, leading to some Internet search engines like Google attracting a crowd nearing one billion users a week, of which 94% never click past the first page of search results.

Filter bubble

A staggering 90% of the world's data stored on computers has been created in the past 2 years. Search result ranking algorithms continue to evolve to keep pace with these increasing volumes. Whilst some proclaim *"The Internet is the greatest serendipity engine in the history of human culture"*, others believe *"the Internet has become so good at satisfying our desires we spend less time seeking new ones"*. Increasingly smart algorithms recommend or suggest related information, trying to predict what we need or may find interesting. This contextual tailoring or personalisation has its benefits although concerns have been raised that algorithms which use historical usage patterns (collaborative filtering), facilitate information discovery via the "*rear view mirror*" placing the searcher in a "*filter bubble*" which constrains and limits accidental encounters in cyberspace. Suggestions based not just on what other people 'have done' may expand serendipitous possibilities. Social media and microblogging sites allow us to form networks that are capable of facilitating what purport to be serendipitous encounters, although we purposefully choose those connections. It has been said that serendipity "*favours the connected*". These connections can be to people, links created by people, or to artificially created networks generated from vast amounts of usage data and text.

Enterprise search

In an enterprise environment, significant frustration still exists where the success seen on the Internet seems harder to replicate inside an enterprise. Investments levels in search, vested interests, organizational culture, the nature of workplace tasks, information governance, small crowds, information structure, document permissions, lack of effective search monitoring and intervention, along with information behaviours of staff and management are among possible causal factors for unsatisfactory retrieval. An area of significant and ongoing interest is exploratory search. Unlike 'known item' (or lookup) search, the question is not *fully formed in the mind of the searcher*. It is possible the actual need may in part be stimulated by the search engine itself. The search engine acting like a creative member of the team making suggestions from initial inputs. The cognitive computing narrative is mainly based around complex reasoning, probabilities and decision making, and is now sufficiently advanced that some companies have recently appointed computers with voting rights onto their board. In our context only the searcher can determine if a filter suggestion is surprising to them but nonetheless the computer is acting as a type of provocative virtual assistant.

Faceted search

In these cases interesting information, the *hidden gems*, may be buried deeply within the search results. The traditional 'ten blue links' and ranking model may not be enough. Following well known commerce websites such as Amazon and Ebay, digital library search engines along with some enterprise search engines, enable the 'what's related' and 'faceted search' concept. Faceted search shows a breakdown of what exists in the search results by author, date and various topic categories with counts, normally shown on the left hand side of the screen inviting further human interaction to filter results. These may be potentially useful options when you consider most enterprise searchers enter two words or less searching increasingly larger haystacks of information. It is therefore not uncommon for most search results to deliver hundreds if not thousands of results. Whilst these prompts aid information discovery, they rarely display surprising or intriguing associated concepts mainly because the metadata used to generate the topics *represents the information items as a whole*, not the matched search context. For example, it is difficult to represent the richness of a 50 page report with 6 metadata tags. Furthermore, the same information item will always be represented by those same 6 categories, regardless of what search terms are used and where relevant matches are found inside the document. Automated techniques can enrich manually added metadata but still represent the information item as a whole not the matched search context within. Text co-occurrence techniques using words that appear in proximity to the search terms found in documents produce vast amounts of data. The most statistically popular or commonly associated terms tend to be the ones displayed, often used in tag cloud derivations and as filters in some search and digital library systems.

A need for the surprising

Recent research by Robert Gordon University published in the *Journal of Information Science*, identified certain information needs with respect to faceted search refiners. Research was conducted using stimuli generated from data provided by the Society of Petroleum Engineers (SPE), Geological Society of London (GSL) and the American Geological institute (AGI). The stimuli was used to gather survey data from 54 petroleum engineers from over thirty oil and gas industry organizations. A need was identified for the 'surprising' as a search filter. The research found the most statistically frequent associations (to search terms) were often *"too vague and no promise of telling me anything I didn't already know"*, *"relevant but not interesting"* and *"contained few surprises"*. However, algorithms such as mutual information measure appeared to generate more intriguing associations "*useful for deep dives*", "*might learn something*" and *"high on interestingness quotient, you can't say where these results may lead you".*

Algorithms

Further research presented at the International Conference on Knowledge Management (ICKM) used *discriminatory* text analytics techniques based on set theory to create colour coded data driven networks surfacing potentially 'surprising' associations to search terms. Initial results were promising. In an observational study of 53 geoscientists in two oil and gas organizations, 41% felt current search interfaces used by their organization facilitated serendipity to a moderate/large extent, increasing to 73% with the introduction of certain algorithmically

generated filters. As put by one participant *"It's like open up the box for me and I'll pick what does not fit with my brain, like one of those games"*. Surprising and serendipitous encounters occurred giving rise to learning experiences, *"It is clear I underestimated the importance of... this is immediately important for the research I am undertaking now"*. Surprising associations can be unusual words, *"some of them attract my attention because they are very unique"*, or quite common terms but appearing in an unusual or discriminatory context *"What is interesting is that Halite is there for the Permian, but technically it could occur for Tertiary, Jurassic, (others), what is surprising is that it has not"*. This may be detached from any initial specific intent, the surprising nature of the association enticing the searcher to drill down further which may lead to a serendipitous encounter.
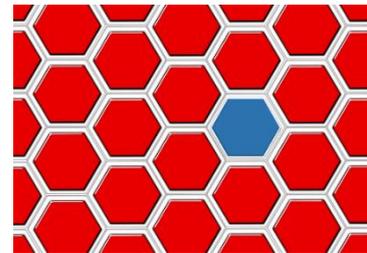
Enhancing creativity and innovation

What is deemed 'surprising' or 'intriguing' by one person, may not be by another as suggested filter

terms are compared with their own cognitive map, like a game of *spot the difference*. However, it appears that certain algorithms are more likely to produce more surprising filter suggestions than others. The challenge with text co-occurrence is to decide what to present to the user, minimizing distraction but offering potential surprises, combining with traditional knowledge organization controlled vocabulary (taxonomy) approaches. If the capability to present the 'surprising' could be embedded in software system design and deployment principles for faceted search, this may enhance learning, creativity and innovation within the enterprise, leveraging the search user interface as a *creative influence*, not just a time saver. Companies that adopt such practices, may experience more "happy accidents" in the user interface than those which do not.

About the authors: Paul Cleverley and Simon Burnett are researchers in the department of Information Management at the Aberdeen Business School at Robert Gordon University in Aberdeen, UK. For more information contact: p.h.cleverley@rgu.ac.uk and sburnett@rgu.ac.uk

Further Reading

Church K and Hanks P. (1991). Word Association Norms, Mutual Information and Lexicography. Computational Linguistics 1991; 16(1): pp. 22-29.

Sergey Brin and Larry Page (1998). The Anatomy of a Large Scale Hypertextual Web Search Engine. Proceedings of the 7th International Conference on the World Wide Web, 107-117

Marchionini G. (2006). Exploratory Search: From Finding to Understanding. Communications of the Association for Computing Machinery (ACM); 49 (4): pp. 41-46.

Bill Thompson (2006). Serendipity casts a very wide net. http://news.bbc.co.uk/1/hi/technology/5018998.stm

Mark de Rond and Iain Morley (2010). Serendipity. Fortune and the prepared mind. Cambridge University Press

HartEnergy E&P (2010). Serendipity is alive and well at Eagleford. http://www.epmag.com/serendipity-alive-and-well-eagle-ford-655606

Barry Friedman (2010). Serendipity is Explorationsists Friend. http://archives.aapg.org/explorer/2010/04apr/mobilebay0410.cfm

Siegler (2010). Google now serving 1 Billion users each week. http://techcrunch.com/2010/09/08/google-one-billion/

Peter Morville and Jeffrey Callendar (2010). Search Patterns. O'Reilly Media.

Luke Allnutt (2011). If the Internet Killed Serendipity, it probably never existed anyway. http://www.rferl.org/content/if_the_internet_killed_serendipity_it_probably_never_existed_anyway/24361774.html

Lori McCay-Peet and Elaine Toms (2011). Measuring the dimensions of serendipity in digital environments. Information Research 16(3).

Mirian Meckel (2011). Save our Serendipity. http://www.miriammeckel.de/2011/10/11/sos-save-our-serendipity/

Eli Pariser (2011). The Filter Bubble. What the Internet is hiding from you. Penguin Books.

Todd Jensen (2011). 2nd page rankings: You are the #1 loser. http://www.gravitateonline.com/google-search/2nd-place-1st-place-loser-seriously

Tom Hulme (2012). Serendipity favours the connected. http://www.wired.co.uk/magazine/archive/2012/09/ideas-bank/serendipity-favours-the-connected

Martin White (2012). Enterprise Search. O'Reilly Media.

Ian Leslie (2012). In search of serendipity. http://moreintelligentlife.com/content/ideas/ian-leslie/search-serendipity?page=full

Steven Makri and Ann Blandford (2012). Coming across information serendipitously – Part 1. Journal of Documentation 68(5), 68-705

John Kelly and Steve Hamm (2013). Smart Machines. IBM's Watson and the era of cognitive computing. Columbia Business School Publishing.

Daniel Rasmus (2013). How IT Professionals can embrace the serendipity economy. Harvard Business Review (HBR) https://hbr.org/2013/08/how-it-professionals-can-embrace-the-serendipity/

SINTEF (2013). Big Data for better for worse: 90% of the Worlds data generated over the past 2 years. http://www.sciencedaily.com/releases/2013/05/130522085217.htm

Sophie Brown (2014). Investment firm appoints robot to its board. http://edition.cnn.com/2014/09/30/business/computers-ceo-boardroom-robot-boss/

Paul Cleverley and Simon Burnett (2014). Retrieving Haystacks: a data driven information needs model for faceted search. Journal of Information Science. Article ahead of print published 5th November 2014: http://jis.sagepub.com/content/early/2014/11/04/0165551514554522.abstract

Paul Cleverley and Simon Burnett (2014). Creating sparks: comparing search results using discriminatory search term word co-occurrence to facilitate serendipity in the enterprise. Submitted to the Journal of Information and Knowledge Management (JIKM) October 2014.