Spring 4-12-2016

SEARCH ENGINES - A SOURCE OF FRUITFUL RESEARCH IN INFORMATION SYSTEMS? (41)

Sophie Iredale
University of Salford, s.iredale@edu.salford.ac.uk

Aleksej Heinze
University of Salford, A.Heinze@salford.ac.uk

SEARCH ENGINES - A SOURCE OF FRUITFUL RESEARCH IN INFORMATION SYSTEMS?

S. Iredale and A. Heinze (University of Salford)

Abstract

This paper highlights the importance that search engines play in our daily lives as consumers of information. The short but eventful history of search engines has brought about a handful of leading search engines, which dominate their country specific markets. This rapid growth and market dominance highlight the long term sustainability of search engines as an information system (IS).

Google is chosen in this position paper as the case of the main search engine and offers a potential for three main areas of fruitful and important IS research: the sustainability of search engine systems, the long terms effect of cannibalisation of vertical search and the implications of search engine bias and competition.

These three research areas are argued as the main research problems in need of further study by IS researchers. Contributions to knowledge in these areas can help in sustainability of the industry.

Keywords: Search Engines, Search Engine Sustainability, Search Engine Bias, Search Engine Competition, Search Engine Optimisation, Search Engine Antitrust

1.0 Search Engines as Information Systems

Search Engines (SE) are a primary repository for a significant amount of information found on the World Wide Web. The information, commonly web pages, text and images (Whitmore, Agarwal and Xu, 2015), is aggregated, stored and managed in a SE's index for recall at a later date, making it a primary information retrieval tool (Croft, Metzler and Strohman, 2010).

People use SE's such as Google, Yandex and Baidu on a daily basis to satisfy their needs for information, thus becoming key online hubs of activity, offering users the only meaningful way to navigate the increasingly complex online world of information (Baye et al., 2016). Fundamentally, a SE's core purpose is to provide intelligent results to aid efficiency in decision making (Enge, Spencer and Stricchiola, 2015), which fits a definition of an information system, as defined by the UKAIS:

"*Information systems are the means by which people and organisations, utilising technologies, gather, process, store, use and disseminate information.*" (UKAIS, 2016).

As such, informaticians, Nguyen et al (2010) argue that SEs are an example of one of the most organised and intelligent IS applications available.

Whilst SE's have historically played a dominant role in the systematic management of information and information recall, research has shown current political, technological and economical implications reveal underlying ethical and sociological tensions that have the potential to impact sustainability of the SE market and operation. Therefore, this position paper aims to explore the history of SE's in order to provoke possible scenarios and implications for the future of the SE industry and to outline IS related research questions and topics.

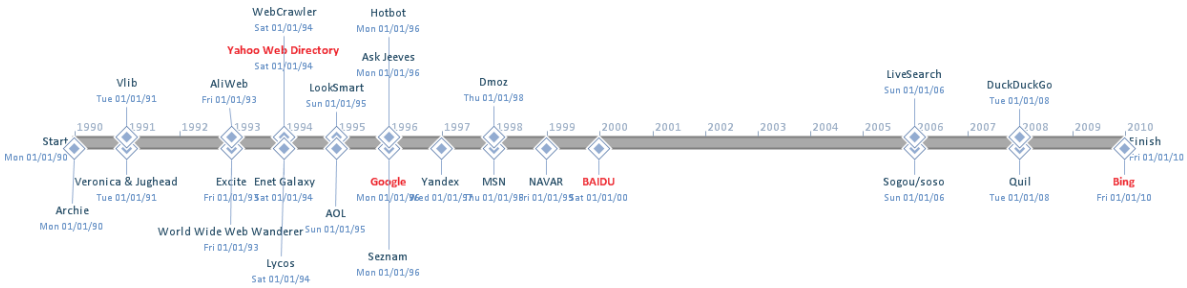
The paper is structured as follows: first, a history of SE industry development is outlined, highlighting the brief but fragile market of SE operators. Second, the tensions and sustainability implications surrounding SE operations are outlined and IS research directions are proposed for further enquiry.

2.0 A Brief History of SE's

The notion of "associative indexing" was first conceived in a pre-internet era by Vannevar Bush who outlined a conceptual method by which a vast amount of information could be stored, continually updated and consulted with speed and ease (Bush, 1945). Information scientists have since contemplated the ways in which the systematic organisation and management of an increasingly expanding bank of digital information can be made more intelligent, responsive and robust (Tavani, 2012).

From the inception of the current day commercial SE's in the early 1990s, there has been a considerable and significant development in their complex search technologies and algorithms, which have aimed to organise the world's data and make it easily accessible to the masses (Google, 2016). Figure 1. provides a high level overview of the history of prominent SE's from around the world, from early downloadable directories to the contemporary and semantically driven web crawling algorithms of today.

Figure 1. History of SE's (adapted from Wordstream, 2016)



Initial analysis of growth and patterns of competitive structures between SE's in the 1990s showed oligopolic and balanced competition, as many attempted to gain stability in an embryonic market. As of the 'Web 2.0' era, from the early 2000's to date, an allocative inefficiency occurred as emergent behaviour of new SE's decreased. The five years post 2000, mark a clear impact on SE entrants post dot.com bubble, which made investment in speculative tools such as search engines less attractive (Adamuthe et al., 2016). Key SE's, such as Google, Bing and Baidu began to quickly monopolise their power and aggressively develop the quality of their service through alterations of their algorithms, with reported claims of over 500 changes per year (MOZ, 2016).

As of 2016, the global market share of SE's stabilised with Google taking a dominant share: Google 65.4%, Bing 15.8%, Yahoo 8.3%, Baidu 8.3%, Ask 0.2%, AOL 0.2%, other 1.8% (Net Market Share, 2016). However, literature suggests that Google's dominance could be

challenged in the coming years, which presents a range of fruitful IS research topics, highlighted in the next section of the paper.

3.0 SE related tensions and potential future research areas

Available literature largely considers Google as a primary case for further enquiry due to its near monopoly position in many countries, therefore, this section of the paper will continue to enumerate key themes of SE sustainability, welfare loss, bias, trust and competition, in relation to Google (Gillespie, Boczkowski and Foot, 2014).

3.1 Google's Dominance and Search Engine Optimisation (SEO) Sustainability

Google is often the first point of reference for the majority of online consumer journey's and decision making processes (Smith, 2013). This raises the question of equality and entitlement for those seeking the privilege of ranking their content in Google's search engine results pages (SERP), in order to benefit from such visibility.

In a bid to rank above competition in SERPs, the system has given rise to Search Engine Optimisation (SEO) practitioners, who purport to help organisations rank their digital content in SERPs (Gunjan et al., 2012). SEO practitioners utilise a variety of optimisation strategies, techniques and tactics, to improve the quality, relevancy and authority of digital content, which Google's algorithm interprets as being worthy of a high ranking position or not (Baye et al., 2016).

Google's ranking algorithm started as an open academic publication (Page et al., 1999), which is now compounded in a secretive and complex computational code, undergoing a significant number of modifications each year (Schwartz, 2015). Google does, however, release a compendium of 'best practice guidelines' to those looking to exercise SEO techniques, and can

be defined as a series of high level recommendations that provide some indication of how to enhance the quality metrics of a website or digital content, to improve its chances of ranking (Google Webmaster Guidelines, 2016).

Whilst working within the paradigms of webmaster guidelines is endorsed by Google it is not enforceable and with no guarantee of the efficacy of results (Elliott, 2011). It has also been hypothesised that alternative SEO methods, outside of Google's guidelines are also effective (Boutet and Quoniam. 2012), which makes the sustainable operation of the SEO industry that more complicated.

Due to the opacity of Google's algorithm it can be argued that complete transparency within the practice of SEO cannot be achieved as this would require SEO practitioners to be "*able to explain why any particular outcome was produced*" (Vieth and Bronowicka, 2015). This has the very real potential to affect the welfare of SEO clients as well as risking the sustainability of the SEO industry (Livingstone and Wang, 2013).

As the system currently stands, there is no regulatory body and no ethical or professional code of conduct. This triggers a series of responses for further research such as a potential need for regulation, standardisation and enforcement (Raval, 2013), as well as practitioner responsibility, accountability and liability (Lewis, 2013; Saam, 2013).

3.2 Cannibalisation of Search

As search technologies have advanced over time, the fundamentals of SE's as IS's that collect, analyse and disseminate information will likely not change (Calero, Moraga and Piattini, 2008). SE's act as intermediary information mechanisms between companies, webmasters and SE users, by generating search results to any given search query. Its core model is founded upon two streams of search results listings: organic results and paid results. Paid results are

bought by advertisers and organic results are achieved naturally through creation of content and its optimisation using SEO techniques (Corniere and Taylor, 2014).

In recent years, SE's, particularly Google, have come under scrutiny for the way in which listings are displayed, which typically sees paid results listed above organic results, pushing organic listings further down the page where click through rate (CTR) can be negatively affected (Slegg, 2016). Such activity has been criticised for manipulating unaware SE users by prominently displaying paid ads above organic listings. The ulterior motive is to influence user behaviour to click on the paid ad, which poses the question of whether SE's display search results that are in the best interest of SE users or for the benefit of those buying paid advert space (Anderson et al., 2016).

Whilst some commercial eye tracking studies show that SE users recognise ads and prefer to view and click on the organic results (Maynes, 2016), independent academic studies are needed to explore this issue. Are consumers happy with their service where companies are worst affected, or, does profound change needs to occur in future to promote "fair-play" (Knowledge@Wharton, 2015)?

3.3 Exposing hidden bias and promoting fair competition

The internal view of Google also offers some initial research considerations. In an end of year fiscal report by Google, it stated that their operation is, and will continue to be, at threat of competitors, both direct and indirect (United States Securities and Exchange Commission, 2014). Direct competitors include other similar SE's, such as Bing, Yahoo and Apple's speculative search platform for iOS and OS X Users (Oliver, 2015), whilst indirect competitors encompass vertical SE's, such as Yelp, Amazon, Indeed, and Rightmove (Drexl and Porto, 2015). This makes the conventional definition of a SE far broader than initially conceived.

Some researchers have also suggested that SE's harm the promulgation of fair content and information exposure, presenting 'popular' SE results from commercial partners above more appropriate listings (Salinger and Levinson, 2013). SE's have the greatest responsibility to reflect true market share to avoid replication of the "*power structure of the conglomerates that dominate the media landscape*" (Granka, 2010: 365), but evidence suggests that this is not being upheld.

Using Google as a continued example of SE bias and unfair competition, it has been claimed that this SE favours its own content above other companies and competitors within search results (Friedman, 2015), in a bid to direct consumers to stay within the confines of Google's own search content. Content includes Universal Search elements and blended results such as product listings ads, job search functionality and hotel and travel booking systems (Southern, 2015). This indicates that Google is not being objective in its own practice and investing in territorial monopolies of information as a form of self preservation, as seen historically with other technological monopolists (Innis, 2008; Patterson, 2012).

Some researchers have theorised that Google's choice to list its own content over competitors could adversely affect the relevancy of results and degrade user search experience (Ratliff and Rubinfeld, 2014). This being true contradicts a fundamental aim of IS's, which seek to improve the quality of information provision and perhaps explains Google's drop of the "don't be evil" motto in 2015 during the move to a new holding company Alphabet (Basu, 2015).

Whilst some researchers dismiss Google's dominance as natural competition (Jamison, 2012), this is not an isolated case of SE bias and competition. As of April 2015 the European Commission registered a formal antitrust complaint against Google for its apparent misuse and monopolisation of search results as well as violation of the EU's competition law (European

Commission, 2015). The long term implications of which are yet to be determined; however, a central hypothesis exists for competition law, which suggests that "*market concentration has harmful effects for consumers*" (Loiselle, 2012: 1).

Further to this, studies suggest that Google reduces the incentive for entrepreneurs and innovators to invest in the SE for fear of having its products and service cloned, resulting in actionable exclusion of new investment and a lack of innovation (Luca et al., 2015). This raises questions of welfare loss on behalf of competitors.

There is, however, an opposing school of thought that considers SE's as key agents for innovation, exercising a 'Schumpeterian hypothesis' (Schumpeter, 1942). It could be argued that monopolistic SE's can stimulate change through their autopoietic nature. In the context of IS's, autopoiesis is characterised as the ability of a system to continually learn from and adapt to its users needs (Schatten and Baca, 2008), which is fundamentally one of the central purposes of its continual algorithm experimentation and updates.

Whilst competition is a prevalent force that has the capacity to drastically alter Google's future and that of its stakeholders, as of yet, there have been no longitudinal studies that determine the effect of SE bias against its own search affiliates and partners. The topic of competition and bias in IS's is identified as important by Turpin (2004), as it has the potential to impact trustworthiness in SE informed decision making, adversely affecting business and consumer choice, utility and welfare.

4.0 Conclusion

Originally primitive in nature, SE's have technologically evolved into an integral part of the way in which the public discovers and consumes information, playing a mediating role in public knowledge and communicative discourse. A historical analysis revealed how ubiquitous

certain SE's have become and the potential ramifications such power has upon decision making and stakeholder welfare. This paper proposes three main themes for future IS research directions, which hinge upon themes of sustainability, bias and competition.

References

Adamuthe, A. C., Surve, A. V., & Thampi, G. T. (2016). Technology Forecasting: A Case Study of Software Technology Product Families. *International Journal of Information Engineering and Electronic Business*, *8*(1), 11-20. doi: 10.5815/ijieeb.2016.01.02

Anderson, S., Waldfogel, J., & Stromberg, D. (2016). *Handbook of Media Economics: Volume 1A*. Oxford: Elsevier

Basu. (2015). New Google Parent Company Drops 'Don't Be Evil' Motto. Retrieved from http://time.com/4060575/alphabet-google-dont-be-evil/

Baye, M. R., De los Santos, B., & Wildenbeest, M. R. (2016). Search Engine Optimization: What Drives Organic Traffic to Retail Sites?. *Journal of Economics & Management Strategy*, *25*(1), 6-31. Retrieved from http://kelley.iu.edu/babur/research/SEO_What_Drives_Organic_Traffic_to_Retail_Sites.pdf

Boutet, C., & Quoniam, L. (2012). Towards Active SEO (Search Engine Optimization) 2.0. *Journal of Information Systems and Technology Management, 9*(3)*,* 1807-1775. doi: 10.4301/S1807-17752012000300001

Bush, V. (1945). As We May Think. Retrieved from http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

Calero, C., Moraga, M.A. & Piattini, M. (2008). *Handbook of Research on Web Information Systems Quality*. New York: Information Science Reference, Hershey.

Corniere, A., & Taylor, G. (2014). Integration and Search Engine Bias. *RAND Journal of Economics*, *45*(3), 576-597. doi: 10.1111/1756-2171.12063

Croft, W., Metzler, D., & Strohman, T. (2010). *Search Engine Information Retrieval in Practice*. USA: Pearson Education

Drexl, J. & Porto, F. (2015). *Competition Law as Regulation*. Cheltenham: Edward Elgar Publishing

Elliott, C. (2011). *How to Save Your Money and Find Better Service in a World of Schemes, Swindles, and Shady Deals*. New Jersey: John Wiley & Sons

Enge, E., Spencer, S. & Stricchiola, J. (2015). *The Art of SEO*. USA: O'Reilly Media

European Commission. (2015). Commission opens Google antitrust proceedings. Retrieved 6 March 2016, from http://ec.europa.eu/news/2015/04/20150415_en.htm

Friedman, A. (2015). Has Google gone too far with the bias towards its own content? Retrieved from https://moz.com/blog/has-google-gone-too-far-with-the-bias-toward-its-own-content

Gillespie, T., Boczkowski, P., & Foot, K. (2014). *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge: The MIT Press

Google Webmaster Guidelines. (2016). Webmaster Guidelines. Retrieved 6 March 2016, from https://support.google.com/webmasters/answer/35769?hl=en

Google. (2016). Google's mission is to organize the world's information and make it universally accessible and useful. Retrieved 6 March 2016, from https://www.google.com/about/company/

Granka, L. (2010). The Politics of Search: A Decade Retrospective. *The information Society*, *26*, 364-374. doi: 10.1080/01972243.2010.511560

Gunjan, V., Pooja., Kumari, M., Kumar, A., & Appa Rao, A. (2012). Search Engine Optimization with Google. *IJCSI International Journal of Computer Science Issues*, *9*(1). 216-214. Retrieved from

https://www.researchgate.net/profile/Vinit_Gunjan/publication/259761919_Search_engine_optimization_with_Google/links/00b7d52dbcae5f1a62000000.pdf
Innis, H. (2008). *The Bias of Communication*. Toronto: University of Toronto Press
Jamison, M. (2012). Should Google be regulated as a public utility? Retrieved from http://warrington.ufl.edu/centers/purc/purcdocs/papers/1205_Jamison_Should_Google_Search.pdf
Knowledge@Wharton. (2015). Why the Google Antitrust Case Is a 'Step in the Negotiation Process'. Retrieved 29 February 2016, from http://knowledge.wharton.upenn.edu/article/why-the-google-antitrust-case-is-a-step-in-the-negotiation-process/
Lewis, J. (2013). Why the SEO industry needs standards (and possible regulation). Retrieved from http://www.huffingtonpost.co.uk/judith-lewis/why-the-seo-industry-needs-standards_b_2408140.html
Livingstone, S. & Wang, Y. (2013). Media literacy and the Communications Act: what has been achieved and what should be done? A 2013 update. Media Policy Brief 2 (Update). Retrieved from http://www.lse.ac.uk/media@lse/documents/MPP/LSE-Media-Policy-Brief-2-Updated.pdf
Loiselle, M. (2012). Flawed Competition Laws: The Case of Google. Retrieved from http://www.iedm.org/files/note1112_en.pdf
Luca, M., Couvidat, S., Seltzer, W., Wu, T. & Frank, D. (2015). Does Google Content Degrade Google Search? Experimental Evidence. Retrieved from http://people.hbs.edu/mluca/SearchDegradation.pdf
Maynes, R. (2016). Eye tracking in 2016: How searchers interact with mobile vs. desktop. Retrieved from https://moz.com/blog/eye-tracking-2016-how-searchers-interact-mobile-serps-desktop
MOZ. (2016). Moz Algorithm Change History. Retrieved , 7 March 2016 from https://moz.com/google-algorithm-change
Net Market Share. (2016). Search Engine Market Share. Retrieved 7 March 2016, from netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=
Nguyen, N., Zgrzywa, A., & Czyżewski, A. (2010). *Advances in Multimedia and Network Information System Technologies.* Advances in Intelligent and Soft Computing. doi: 10.1007/978-3-642-14989-4
Oliver, S. (2015). Apple challenges Google with growing web search program, fueled by Topsy acquisition. Retrieved 7 March 2016, from http://appleinsider.com/articles/15/05/06/apple-challenges-google-with-growing-web-search-program-fueled-by-topsy-acquisition
Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the web. Retrieved from http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf
Patterson, M. (2013). Google and Search-Engine Market Power. *Harvard Journal of Law & Technology Occasional Paper Series*. 1-24. Retrieved from http://jolt.law.harvard.edu/antitrust/articles/Patterson.pdf
Ratliff, J. & Rubinfeld, D. (2014). Is there a market for organic search engine results and can their manipulation give rise to antitrust liability? *Journal of Competition Law & Economics, 0*(0), 1-25. Retrieved from https://www.ebscohost.com
Raval, V. (2013). Trust in Information Systems. *ISACA Journal, 2.* 1-2. Retrieved from http://www.isaca.org/Journal/archives/2013/Volume-2/Documents/13v2-Trust-in-Information.pdf
Saam, C. (2013). Suing your SEO: Can an Agency be held liable for poor results? Retrieved from http://searchengineland.com/suing-your-seo-agency-168370

Salinger, M. & Levinson, R. (2013) The Role for Economic Analysis in the FTC's Google Investigation. Retrieved from http://www.law.northwestern.edu/research-faculty/searlecenter/events/internet/documents/Salinger_Economics_of_Google_and_Antitrust_Case_Searle_conference_version.pdf

Schatten, M. & Baca, M. (2008). A critical review of autopoietic theory and its applications to living, social organizational and Information Systems. Retrieved from https://www.researchgate.net/profile/Markus_Schatten1/publication/209023775_A_critical_review_of_autopoietic_theory_and_its_applications_to_living_social_organizational_and_information_systems/links/07af210357b3c9e8881fed3a.pdf

Schumpeter, R. (1942). *Capitalism, Socialism & Democracy.* USA: George Allen and Unwin

Schwartz, B. (2015). We make thousands of updates to search algorithms each year. Retrieved from https://www.seroundtable.com/google-updates-thousands-20403.html

Slegg, J. (2016). Google Adwords switching to 4 ads on top, none on sidebar. Retrieved from http://www.thesempost.com/google-adwords-switching-to-4-ads-on-top-none-on-sidebar/

Smith, P. (2013). Search, the start of the customer journey. Retrieved from http://www.thedrum.com/knowledge-bank/2013/12/18/search-start-customer-journey

Southern, M. (2015). Biased Google search results are hurting users, Harvard study claims. Retrieved from https://www.searchenginejournal.com/biased-google-search-results-are-hurting-users-harvard-study-claims/135824/

Tavani, H. (2012). Search Engines and Ethics. Retrieved from http://plato.stanford.edu/entries/ethics-search/#IntOve

Turpin, M. (2004). Decision-making Biases and Information Systems. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.2627&rep=rep1&type=pdf

UKAIS. (2016). Definition of IS. Retrieved 4 March 2016, from http://www.ukais.org.uk/about/DefinitionIS.aspx

United States Securities and Exchange Commission. (2014). End of year Fiscal Report, Google 2014. Retrieved from https://www.sec.gov/Archives/edgar/data/1288776/000128877615000008/goog20141231 10-k.htm

Vieth, K., & Bronowicka, J. (2015). Should algorithms decide your life? Retrieved from https://cihr.eu/wp-content/uploads/2015/07/EoA_web.pdf

Whitmore, A., Agarwal, A., & Xu, L. D. (2015). The Internet of Things - A survey of topics and trends. *Information Systems Frontiers, 17*(2), 261-274. doi: 10.1007/s10796-014-9489-2

Wordstream. (2016). History of Search Engines. Retrieved 7 March 2016, from http://www.wordstream.com/articles/internet-search-engines-history

Lifelong Learning Programme

This project has been funded with support from the European Commission. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.