Library Hi Tech

Search engine optimization: A case study using the bibliographies of LG Science Land in Korea

Sungin Lee Wonhong Jang Eunsol Lee Sam G. Oh

Search engine optimization

A case study using the bibliographies of LG Science Land in Korea

Sungin Lee
School of Dentistry, Seoul National University, Seoul, Republic of Korea, and
Wonhong Jang, Eunsol Lee and Sam G. Oh
iSchool – Library and Information Science and Data Science, Sungkyunkwan University, Seoul, Republic of Korea

Received 5 February 2016
Revised 17 March 2016
Accepted 6 April 2016

Abstract

Purpose – The purpose of this paper is to examine the effect of, and identify core techniques of, search engine optimization (SEO) techniques applied to the web (http://lg-sl.net) and mobile (http//m.lg-sl.net) Science Land content and services at LG Sangnam Library in Korea.

Design/methodology/approach – In accordance with three major SEO guidelines, ten SEO techniques were identified and applied, and their implications were extracted on three areas: improved search engine accessibility, increased relevance between site content and search engine keywords, and improved site credibility. The effects were quantitatively analyzed in terms of registered search engine keywords and influx of visits via search engines.

Findings – This study shows that SEO techniques help increase the exposure of the library services and the number of visitors through search engines.

Practical implications – SEO techniques have been applied to a few non-Korean information service organizations, but it is not a well-accepted practice in Korean libraries. And the dominant search engines in Korea have published their own SEO guidelines. Prior to this study, no significant endeavors have been undertaken in the context of Korean library services that have adopted SEO techniques to boost exposure of library services and increase user traffics.

Originality/value – This is the first published study that has applied optimized SEO techniques to Korean web and mobile library services, in order to demonstrate the usefulness of the techniques for maximized exposure of library content.

Keywords Library services, Search engine, Content accessibility, Content relevance, Search engine optimization, Site credibility

Paper type Research paper

Introduction

It has become a cliché to say that there is overproduction of information on the web, almost to point of choking on it. Consequently, finding right information has, as opposed to general perception, become a rather sophisticated and specialized art, than a mechanical process with minimum human intervention required. Recent decades have seen clever technological advancements and so vast an expansion of the web data spaces in which high-quality information faces direct competition with the inaccurate and dubious; right information is buried and strewn all over the place on the web in the cacophony of competing voices for attention. A variety of new devices offer no repose in the scene, instead rather new challenges to libraries, which used to be the vital centers of information but have been gradually sidelined even by the core patrons, students and researchers, who now greatly resort to the giants – search engines (Gauder, 2011)

Studies report that on average 2.5 exabytes (1 followed by 18 zeros) of data are created daily, and this figure doubles in every 40 months. Now the amount of data infused into the internet per second surpasses the yearly addition of data 20 years ago

Library Hi Tech
Vol. 34 No. 2, 2016
pp. 197-206

0737-8831
DOI 10.1108/LHT-02-2016-0014

(McAfee *et al.*, 2012), and the abundance of data comes with a price – extra time taken for understanding, filtering, and extracting the information the user desires. And time is a precious commodity in the so-called "attention economy" of the web where organizations compete with others to grab scarce commodity of users' attention (Halavais, 2013). All these make search engines all the more attractive to users, a fact that has been substantiated, for example, by a study (Lewandowski and Gorman, 2015) that reports that in the USA only there are more than 18 billion search queries every month. In this competitive search environment, high-quality information does not necessarily translate into a higher traffic to your site, rather a new approach is required in content preparation and services in libraries.

Visibility is the motto in search engines; high visibility on search engines helps induce a higher traffic to your site (Killoran, 2013). In order to increase visibility, search engines provide search engine optimization (SEO), a set of technical processes and tasks that help organizations to secure relatively top positions in search engine result listings. It is high time, therefore, for libraries to delve into the details of SEOs and to prepare SEO-compliant contents and services for major search engines.

The search engine market in Korea is dominated by two major engines – Naver© and Daum© – parading a lion's share of the domestic search market, according to ReturnOnNow (72 and 77 percent for Naver, 18 and 20 percent for Daum in 2014 and 2015, respectively, http://returnonnow.com/internet-marketing-resources/2015-search-engine-market-share-by-country/). Naver© published its own SEO guidelines in 2013, which contain overlapping elements with the Google SEO. Daum© has adopted the SEO by Bing©.

This experimental study applies SEO techniques to the bibliographies of Science Land hosted by LG Sangnam Library. The new SEO-compliant contents were applied to Naver© and Daum© search engines, in order to investigate whether the SEO techniques would induce more traffic and increase the number of visits, and to identify core SEO techniques that showed the most promising results in terms of, for example, the number of visits to the site.

Background and literature review

SEO

Information-publishing organizations feel the strategic need for maximized exposure to end users of their contents and services, and SEO has become the main means by which to achieve the goal. SEO in general refers to scientific methods and processes to induce web traffic to websites (Davis, 2006). It defines structured and goal-driven techniques that would streamline website structures, languages used, and site interaction mechanisms, and others, all of which are designed to promote the site to land itself at the top of search engine listings in light of the keyword(s) entered by the user. Development of SEO has gone in tandem with search engines, and as such, websites tend to orient themselves to SEOs developed by major search engines such as Google©, Microsoft©, and Yahoo©. Unlike most other countries where Google and Microsoft dominate the markets, the search engine market in Korea is rather unique in that two local giants, Naver© (www.naver.com) and Daum© (www.daum.net), enjoy far surpassing dominance than Google, which only recently shows a higher than before market share, thanks in particular to the popularity of Android-based devices. This idiosyncrasy has led the authors to investigate the applicability of the well-accepted SEO by Google in combination with the two SEOs published by the two search engines, and to identify the SEO elements that are appropriate for Science Land hosted by LG Sangnam Library.

SEO and libraries

Library websites were not the main information hubs (Cantrell *et al.*, 2005) that college students and researchers had used for research, and the trend seems in no wise to have reverted (Gauder, 2011). Inaccuracy and dubious credibility of information sources that search engines present to end users have not been working as deterrents to their widespread adoption, which on the contrary seems only to have been fortified by the perceived convenience and accessibility of search engines. In short, the centrality of search engines in information discovery and retrieval on the internet is now undeniable (Coopee, 2000; Enge *et al.*, 2012). Libraries will greatly benefit from search engines in increasing the exposure of library contents (Cahill and Chalut, 2009), especially when they are armored with professional awareness as regards the biases of search engines and SEO – for example, exclusion of certain types of information non-conformant to the policy set by a search engine – and the corresponding distortion of search results.

A number of studies have emerged that looked into SEO in library contexts. Onaifo and Rasmussen (2013) provides an extensive overview of SEO in libraries as well as characteristics of library websites that contributed to exposure by search engines. Arguably the most relevant work related to SEO in libraries in general, and to the present study in particular, may be Rushton *et al.* (2008). Their pilot study optimized a handful of web pages in Binghamton University Libraries, and expounded optimization processes involved. Though the study scope was rather narrow, their work has provided much insight and guidance for the present study, which targets all the web pages in Science Land. Science Land is a digital library service, which suffered from the same dilemma, portrayed by DeRidder (2008), of "the Deep Web, inaccessible to general web crawlers and hence to generalized search engines such as Google," and as such, SEO was viewed as a crucial component in our efforts to enhance content visibility for Science Land. A more recent study by Gasparotto (2014) argues for SEO in heightened visibility for ethnic sexuality studies contents that seemed to be in direct competition with other contents not related to keywords, proving the necessity of quality metadata for high visibility. A thorough introduction to SEO adoption as institution-level endeavors is introduced by Arlitsch *et al.* (2013). In Korea, KoreaScience (www.koreascience.or.kr) of KISTI (Korea Institute of Science and Technology Information, http://en.kisti.re.kr) provides open-access journals indexed by Google. Park *et al.* (2013) demonstrates more traffic to, and higher placement of, contents housed at KoreaScience, by inserting HTML metatags representing machine-processable bibliography metadata, and by creating XML site maps.

Methods

Site usage

Since its launch in 2003, LG Science Land (http://lg-sl.net) at LG Sangnam Library has been providing digitalized scientific contents mainly targeted toward the young. The patron distributions shown in Figure 1 shows that elementary, middle, and high school students comprise the majority of the user base, and similar distributions were noticed for visitors as well at Figure 2.

The popularity in terms of visitors had peaked in 2007, after which the numbers of yearly site visitors have been gradually in decline as shown in Figure 3, and starting from 2008, the numbers of page views have started to decrease as well shown at Figure 4. This decline may be due in part to a wave of similar services that opened in Korea since 2007. Hence, it was considered imperative to maximize the exposure of the LG Science Land service to a wider audience, to boost the number of visitors and page views.

In order to locate the access points whose traffic mattered the most, we analyzed the access patterns to LG Science Land. The pie chart in Figure 5 shows different access methods by which our patrons visited the site during the month of March, 2014. The Organic Search denotes traffic from search engines; the Direct search refers to traffic originated from direct input of URLs (including favorites); the Referral shows traffic

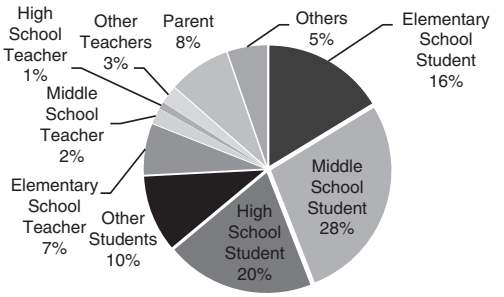

Figure 1.
Patron distribution (as of May 2015)

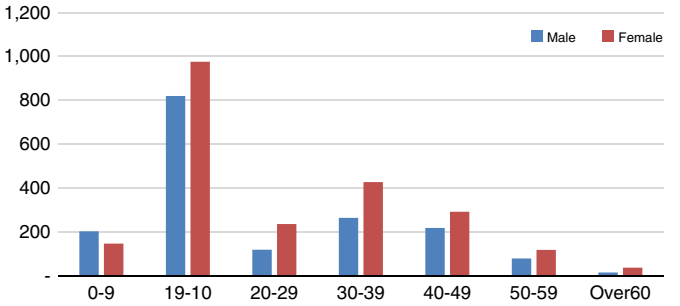

Figure 2.
Visitors by age (May 1, 2014-May 31, 2015)

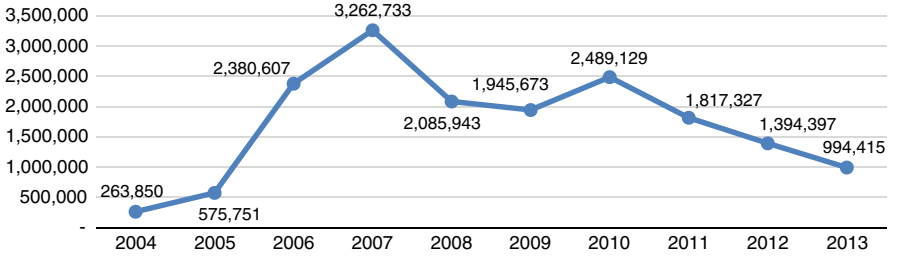

Figure 3.
Yearly visitors from 2004 to 2013

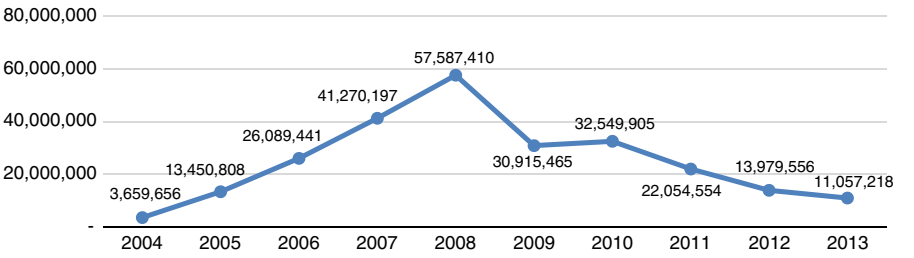

Figure 4.
Yearly page views from 2004 to 2013

from external links; and the Social denotes traffic from social network services. Both Organic and Direct were the main access methods to the site. The traffic generated by the Direct method came from those users already aware of Science Land, and in order to generate more traffic, we opted to focus our work on search engines.

Science land and search engines

Prior to applying SEO techniques, we analyzed the status-quo of our services in search engines. We first identified 195 relevant keywords, classified into three groups: first, 29 keywords related to LG Science Land brand; second, 137 keywords relevant to science songs created by the library; and third, 29 keywords suitable for general sciences, as shown in Table I. The top 10 keywords for each category are listed in Table AI. We then analyzed how these keywords fared in terms of search engine result listings in three major search engines, Daum, Naver, and Google, as shown in Table II.

The site has used multiple domains, each of which had its own robots.txt file. This effectively limited user exposure to the pages housed in the site. The rankings of 195 keywords relevant to the site were low in the search results from the search engines, which translated into the relatively low percentage of visitors originated from the search engines, 39.19 percent, as shown in Table II. These results convinced us of the need for SEO, which would increase the exposure of the keywords to the search engines, and raise the relative percentage of visitors originated from the engines.

Selected SEO techniques

The SEO techniques chosen were organized around the three goals: improved accessibility for search engines, increased relevance between site content and search engine keywords, and increased site credibility. The techniques were stepwise applied to the site from April 2014 to September 2014.

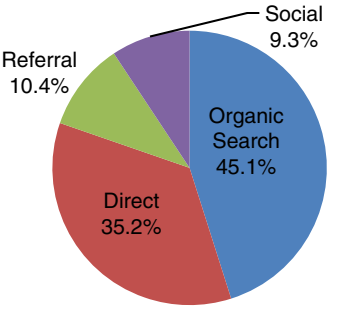

Figure 5.
Methods of access (March 1, 2014-March 31, 2014)

Category	Description	Count
Keywords related to LG Science Land Brand	Keywords that describe LG Science Land, and its content (ex: LG Science Land, Science Land, science song, etc.)	29
Keywords related to science songs	Keywords that describe titles of science songs and animations created by the site (ex: digestion song, geological layer song, herbivorous dinosaur song, etc.)	137
Keywords related to general sciences	Keywords that describe general terms in science (ex: scientific experiment, learning material for science, etc.)	29
Total		195

Table I.
Main keywords relevant to the site

Category	Description
Domains in use	11 domains, including www.lg-sl.net, www.lg-sl.co.kr, www.lgsl.kr
robots.txt	1. Access blocked except for major search engines 2. Varying robots.txt files for major search engines, depending on content
Registered pages by search engine	Google: ~16,700 pages Naver: 490 pages Daum: 94 pages
Percentage in which search results show Science Land among the top 10 entries when the 195 main keywords were queried	An average of 60.01% in Google, Naver, and Daum
Percentage in which search results show Science Land among the top 30 entries when the 195 main keywords were queried	An average of 74.91% in Google, Naver, and Daum
Visits originated from search engines	39.19% (37,753) of all the visits (measured up to March 2014)

Table II.
Search engine exposure

Improved accessibility for search engines

The following five techniques were taken to facilitate more streamlined access to the site by the search engines.

URL structure optimization. Since web crawlers collect data while traversing through URLs, it is desirable that simplified URL structures would enable facilitated traversal by the crawlers. Science Land, therefore, has purged superfluous parameter values from its URLs.

Internal redirect. In case of changes in web page URLs, or page deletions, internal redirect forwards URL requests to alternative internal URLs.

XML sitemap. A sitemap is a page that shows the structure of the website. LG Science Land already provided a sitemap for users, but none for search engines. A new XML sitemap for a search engine was created to help search engines to discover the pages on Science Land.

Robots.txt. A robots.txt file tells search engines whether or not certain pages in your site are accessible. Previously, the robots.txt file at Science Land permitted crawling by selective search engines (Google, Naver, Daum, and Bing) within pre-set time frames. To maximize its exposure, the robots.txt was modified to allow unlimited access to the site by any search engines any time of the day.

Naver content syndication. Content syndication enables pushing contents of a website onto other websites, by means of snippet, link, and others. Naver owns its proprietary content syndication system, and Science Land transfers its contents to Naver in formats compliant to the guidelines set by Naver for content syndication.

Increased relevance between site content and SE keywords

Two tags were used to increase the correspondence between the content of our site and the keywords of the search engines.

Meta tag. The < meta > tag provides metadata such as short descriptions and keywords about the web page, to be used by search engines. Science Lab previously used a metatag that applied to all the pages at the site, lacking

distinctness and detailed information. The site now use distinct metatags that fit the content of each page.

Title tag. The title tag of a page, as the name implies, tells search engines the topic of the page. Our site used some title tags before, which were not as descriptive as one would expect them to be. We now use descriptive content for each page in the title tag, so that users can in greater ease recognize the relevance of the pages from the site.

Improved site credibility

A search engine uses its own algorithm to decide site credibility – the more credible the site, the higher the chance of the site being listed among the top entries. The following three techniques were employed to increase site credibility.

Redirect for multiple domains. A website may face a low credibility, rated by search engines, of its content when duplicate pages exist for a page. Since LG Science Land used multiple domains, each domain was found with duplicate pages. In order to increase the credibility of the site, one domain is set to be representative domain to which all other domains are redirected.

Canonical Tag (or URL). A canonical URL indicates the preferred URL for a page that can be accessed through multiple URLs. When multiple URLs exist to access the same page, it may cause some challenges to search engines to reach the page. Prior to SEO, Science Land intentionally used duplicate pages to maximize usability and access to its content. Now Science Land uses canonical URL (rel = "canonical" link element).

Dead links. A dead link refers to the one which does not exist on the web any more. Science Land used lots of external links, many of which were dead links and subsequently removed from the site.

Results

The analysis was performed to evaluate the effects of the SEO techniques, and its results are presented in three headings below: percent change of visitors via search engines, keyword ratings, increased number of pages indexed by search engines.

Percent changes of visitors via search engines

After SEO, there has been a gradual and steady increase in the numbers of visitors coming from the search engines, as shown in Table III.

Keyword ratings

Table IV shows the benefits of SEO in terms of promoting the keywords relevant to Science Land content. Prior to SEO in April 2014, out of the 195 keywords, on average 60.01 percent of them were ranked among the top ten results, and 74.91 percent among the top 30. In September 2014 after SEO, the rankings moved up to 77.31 and 70.87 percent, respectively.

Increased number of pages indexed by search engines

Prior to SEO application in April, 2014, Science Land had 61,300 pages indexed in Google, 2,872 pages in Naver, and 45,500 pages in Daum. After SEO in September, 2014, the numbers increased to 221,000 pages in Google, 23,000 pages in Naver, and 53,600 pages in Daum.

Date	No. of visitors	Visitors from SE	%
March 2014	96,345	37,753	39.19
April 2014	174,069	67,926	39.02
May 2014	132,136	57,655	43.63
June 2014	131,113	58,476	44.60
July 2014	108,051	47,423	43.89
August 2014	84,036	42,597	50.69
September 2014	98,589	50,441	51.16
October 2014	109,845	60,244	54.84
November 2014	119,351	64,293	53.87
December 2014	86,635	43,786	50.54
January 2015	58,621	31,715	54.10
February 2015	53,268	30,144	56.59
March 2015	104,214	60,991	58.52
April 2015	138,207	78,858	57.06
May 2015	96,656	59,507	61.57

Table III.
Percent Change of visitors (2014.03 – 2015.05)

		April 2014 (%)			September 2014 (%)		
Keywords	Of search results	Google	Naver	Daum	Google	Naver	Daum
LG Science Land Brand	Among the top 10	93.10	75.90	65.50	96.60	72.40	93.10
	Among the top 30	100.00	93.10	72.40	100.00	72.40	93.10
Science songs	Among the top 10	94.20	48.90	83.20	99.30	97.10	89.10
	Among the top 30	94.20	89.10	97.80	99.30	98.50	93.40
General sciences	Among the top 10	41.40	20.70	17.20	72.40	31.00	44.80
	Among the top 30	48.30	44.80	34.50	75.90	34.50	51.70
Average	Among the top 10		60.01			77.31	
	Among the top 30		74.91			79.87	

Table IV.
Keyword ratings in search results

One limitation of this study comes from its inability to understand the inner workings of the ranking algorithms of the search engines, since they are proprietary. Otherwise, it would have shed much more light into our results, and helped us analyze them in finer detail.

Conclusions

This study presents an application case in which well-established SEO techniques were selected and applied to the whole library content. It shows the benefits of SEO in terms of increased traffic from the search engines, and a greater number of web pages indexed by the search engines, and higher keyword ratings. It is the first systematic attempt to apply SEO techniques to the whole library site in Korea, in order to demonstrate the usefulness of SEO in local libraries.

This study has not addressed the differentiation between the applied SEOs, nor identified the varying impact of each keyword on SEO visibility. In-depth investigation of open-source-based search engines and their ranking algorithms will expose more data on relations between ranking algorithms and SEO techniques. These research venues will be explored in our subsequent studies. We also envision to carry out a survey-based study on users perception on how easy it is to locate the content and service from the search engine(s) of their choice. And keywords related to new services and content will be identified and applied to SEO.

References

Arlitsch, K., OBrien, P. and Rossmann, B. (2013), "Managing search engine optimization: an introduction for library administrators", *Journal of Library Administration*, Vol. 53 Nos 2-3, pp. 177-188.

Cahill, K. and Chalut, R. (2009), "Optimal results: what libraries need to know about Google and search engine optimization", *The Reference Librarian*, Vol. 50 No. 3, pp. 234-247.

Cantrell, J., Cellentani, D., De Rosa, C., Hawk, J., Jenkins, L. and Wilson, A. (2005), *Perceptions of Libraries and Information Resources*, OCLC, Dublin, OH.

Coopee, T. (2000), "How to climb search engine rankings", *Info World*, Vol. 22 No. 24, p. 61.

Davis, H. (2006), *Search Engine Optimization*, O'Reilly Media, Inc., Sebastopol, CA.

DeRidder, J.L. (2008), "Googlizing a digital library", *The Code4Lib Journal*, No. 2, available at: http://journal.code4lib.org/articles/43 (accessed January 5, 2016).

Enge, E., Spencer, S., Stricchiola, J. and Fishkin, R. (2012), *The Art of SEO*, O'Reilly Media, Inc., Sebastopol, CA.

Gasparotto, M. (2014), "Search engine optimization for the research librarian: a case study using the bibliography of us latina lesbian history and culture", *Practical Academic Librarianship: The International Journal of the SLA Academic Division*, Vol. 4 No. 1, pp. 15-34.

Gauder, B. (2011), "Perceptions of libraries, 2010: context and community", a report to the OCLC Membership, OCLC Online Computer Library Center, Inc., Dublin, OH.

Halavais, A. (2013), *Search Engine Society*, John Wiley & Sons, Hoboken, NJ.

Killoran, J.B. (2013), "How to use search engine optimization techniques to increase website visibility", *Professional Communication, IEEE Transactions on*, Vol. 56 No. 1, pp. 50-66.

Lewandowski, D. and Gorman, G. (2015), "Living in a world of biased search engines", *Online Information Review*, Vol. 39 No. 3, pp. 278-280.

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D. and Barton, D. (2012), "Big data. The management revolution", *Harvard Business Review*, Vol. 90 No. 10, pp. 61-67.

Onaifo, D. and Rasmussen, D. (2013), "Increasing libraries' content findability on the web with search engine optimization", *Library Hi Tech*, Vol. 31 No. 1, pp. 87-108.

Park, M., Choi, S.-h. and Seo, T. (2013), "Search engine optimization for a scholarly website: a case of KoreaScience", *The Korea Contents Association, International Conference on Convergence Content*, Vol. 11 No. 2, pp. 171-172.

Rushton, E.E., Kelehan, M.D. and Strong, M.A. (2008), "Searching for a new way to reach patrons: a search engine optimization pilot project at Binghamton University Libraries", *Journal of Web Librarianship*, Vol. 2 No. 4, pp. 525-547.

(The Appendix follows overleaf.)

Appendix

Table AI.
Top 10 keywords for each category

Group	No.	SEO keywords
Brand	1	LG Science Land
	2	LG Science
	3	Science Land
	4	Elji Science Land
	5	Dr. ChukChuk Laboratory
	6	Science Laboratory
	7	Science Story
	8	Creative Expedition
	9	Creative TongTong Experiential Activities
	10	Science Song
Science Song	1	Science Song Plus
	2	LG Science Land Perceptible Material Song
	3	LG Science Land Digestion Song
	4	LG Science Land Nutrition Song
	5	LG Science Land Magnet Song
	6	LG Science Land Hormone Song
	7	Earthquake Song
	8	Gene Song
	9	Acid and Base Song
	10	Rock Song
General sciences	1	Science App
	2	Virtual Science Experiment
	3	Science Playground
	4	Science in Magic
	5	Hot Science
	6	Science in Cooking
	7	Science with explanation
	8	Pictures of the Universe
	9	ByulByul Meteorological Station
	10	Science with Reasoning

About the authors

Sungin Lee is a Doctoral Candidate at the Seoul National University, Medical Informatics.

Wonhong Jang is a Doctoral Student at the Sungkyunkwan University (SKKU) iSchool – Library and Information Science.

Eunsol Lee is a Master's Student at the Sungkyunkwan University (SKKU) iSchool – Library and Information Science.

Sam G. Oh is a Professor at the Sungkyunkwan University (SKKU) iSchool – Library and Information Science and Data Science and currently serving as an Executive Committee Member of the iCaucus. Sam G. Oh is the corresponding author and can be contacted at: samoh21@gmail.com