Literature Search Habits of MIS Academics: Empirical Evidence on the Discovery of Impactful Research

Chiara Francalanci[1] and Paolo Giacomazzi[1]

[1] Politecnico di Milano, Milano, Italy
chiara.francalanci@polimi.it,paolo.giacomazzi@polimi.it

Abstract. The amount of academic literature published every year has increased at a steady 20% rate since the 1990s. With this impressive growth of available information, the discovery of *relevant* papers that are worth reading is recognized to be challenging. The search mechanisms of online archives are generally considered limited, as search keywords typically span multiple research areas and retrieve a large number of papers that are only partly pertinent to the user's interests.
The first research question of this paper is whether and to what extent academics perform their search online. The second research question is whether and to what extent academics use current advanced search mechanisms, as an indication of their commitment to online discovery. The third research question is on the role played by online search in different phases of the research process, that is choosing a research topic, finding readings on the topic, and selecting citations. To help answer these questions, the paper presents the results of an empirical survey conducted with academics in the MIS field. Findings from 326 respondents unveil interesting insights on the literature search habits of academics and, overall, indicate that despite the consensus on the low quality of current online search mechanisms, only a tiny minority of users seems to be willing to trade search simplicity for relevance.

Keywords: online search, search engine, impact factor, h-index.

1 Introduction

The number of academic papers published every year is continuously growing [23], [25] at an average 20% rate since the 1990s. With this impressive and steady growth rate, the size of the academic literature increases by a factor of 10 in 5 years. Although this demonstrates a growing interest towards research, previous literature has also pointed to some negative effects, which are summarized under the umbrella of a general *information overload* [4].
Efficient search mechanisms could obviate the risks of information overload by retrieving relevant papers irrespective of the size of the underlying archive. However, the search mechanisms of online archives are generally considered limited [4], as search keywords typically span multiple research areas and retrieve a large number of papers that are only partly pertinent to the user's interests. Users entering search keywords can

limit search results by selecting specific research areas, but it has been noted how the definition of research area provided by online archives is usually very broad and far from the more practical notion of research field [12].

This paper starts from the observation that working to improve online search is valuable only if the actual search habits of academics point to a tangible need for better online search mechanisms. To gather insights on this issue, this paper analyzes the search habits of academics by addressing the following research questions:

1. To what extend academics perform their search online?
2. To what extent academics use current advanced search mechanisms?
3. How to academics use online search in their research and, particularly, what is the role played by online search in different phases of the research process, that is choosing a research topic, finding readings on the topic, and selecting citations?

To help answer the three questions listed above, the authors of this paper have conducted a large-scale empirical survey with academics in the MIS field. Findings help understand whether and how online search can be improved to meet largely shared requirements.

The next section discusses the state of the art on search approaches and current online services. Section 3 presents our research hypotheses and testing results are reported in Section 4. Findings are discussed in Section 5 and conclusions are finally drawn in Section 6.

2 State of the Art

Online archives make a standard distinction among three types of basic parameters that users can specify to drive search: authors, publication outlets and content. Ideally, users entering a search would like to retrieve *all* the publications that are relevant to their search goals, possibly *ranked* by relevance. Users can set their search goals by specifying a value for the different types of search parameters, i.e. by formulating a query. This section reviews previous literature focusing on the main types of search parameters (authors, publication outlets, and keywords) and provides a comparative analysis of existing online archives in the last sub-section.

2.1 Authors as an online search variable

The use of author names as a search criterion is a feature provided by all online archives and search engines. Some archives, such as Scopus, provide it as a basic search functionality, while others, such as Google Scholar, provide it as an advanced functionality. In both cases, entering the full name of an author or the last name only, as well as providing the first and middle name initials or the full spelling can return significantly different search results. These inconsistencies can be due to homonymies among different authors or to the standards applied by different publication outlets in recording author names.

The issue of homonyms is central to the literature on the *h*-index (cf. [22]). It has been observed that the calculation of the *h*-index is error-prone due to the inclusion of publications from homonymous authors, the exclusion of relevant publication outlets and missing citations [22]. From an online search perspective, the *h*-index can provide an assessment of the impact of authors especially when users do not specify any author names in their query. However, it should be acknowledged that the *h*-index is a controversial measure of impact. The *h*-index has been initially welcomed as a significant improvement in the assessment of an author's impact compared to the mere number of publications [10]. It has been noted how the *h*-index takes into account the number of citations and encourages the publication of fewer, but more impactful papers [10]. On the other hand, the more recent literature tends to be more critical of the *h*-index, introduces alternative measures of impact [6], such as the 37 variants of the *h*-index reviewed in [11], and emphasizes the need for complementing the quantitative assessment of impact with a more qualitative approach [10].

Most search engines and online archives, including Google Scholar and Scopus, provide the total number of citations of the papers that are returned in response to a query. The number of citations can be used as an indication of the impact of a paper and help select a few readings from a long list of search results. To the best of our knowledge, the *h*-index of authors cannot be used as a search criterion. In some cases, users can click on an author's name to be provided an assessment of the author's publications which includes the author's *h*-index. For example, Google Scholar provides the author's list of most recent publications, the *h*-index and the i10-index, that is the number of publications with at least 10 citations. Scopus provides a variety of analytics describing an author's productivity, including the *h*-graph, visually showing the calculation of the *h*-index together with a list the most cited papers above the *h*-threshold.

Calculating the *h*-index is computing intensive. Most online archives seem to store author records with descriptive statistics including the *h*-index. In theory, this would enable them to include the *h*-index among their search criteria without incurring the risk of exceedingly complex queries. Whether users could benefit from using the *h*-index as a search variable remains an open issue.

2.2 Publication outlets as an online search variable

Publication outlets or sources can be used as a search variable by specifying their *name* or their *type*. Most search engines support the incomplete specification of source name and use simple string matching algorithms to map the incomplete specification to actual source names. This mapping often results into multiple candidate sources and, once again, search engines aim at avoiding false negatives by considering all candidate sources, resulting into information overload.

The type of a source represents a categorization of its mission and style, such as "journal," "conference," "magazine," "patent," and so on. Source types vary across search engines, with no standard categorization. In general, search results include a number of different sources, even if they are restricted to a specific source type. The number and variety of sources increases when search is based on keywords that are inherently interdisciplinary. It cannot be expected that users have a precise idea of the reputation of all sources, especially if the search engine is general purpose, such as

Google Scholar or Scopus. In theory, users may benefit from functionalities helping them assessing the quality of a source.

The issue of assessing journals is central to the literature on the *impact factor*. The impact factor represents a quantitative indicator of the impact that papers published on a given source have on average [9]. Similar to the *h*-index, the impact factor is based on citations and is highly controversial, both in principle and in practice. First of all, citations do not represent the only indicator of impact and, as a consequence, the impact factor does not provide a complete assessment of the quality of a source [14]. Furthermore, many researchers believe that they have published important research work in low-impact journals [14]. From a more practical standpoint, the impact factor is affected by several measurement problems. Self-citations represent the most widely discussed measurement problem. It has been noted how editors can increase the impact factor of their journal by inflating self-citations through editorials and readers' comments on published articles [13], [14]. A mismatch between citing and cited documents has also been observed, raising concerns on the precision with which the impact factor is measured [27]. An alternative metric to measure the impact of publication outlets is the *eigenfactor* [7].

Most search engines and online archives do not provide an assessment of the impact factor and none allows the impact factor to be specified among search parameters. Scopus is the only general-purpose engine providing three different measures of the impact factor of journals, the SCImago Journal Rank, the Impact per Publication, and the Source Normalized Impact per Paper. Including the impact factor among search parameters could be subject to strong criticism, as it would strengthen the role of a controversial indicator. Whether users would benefit from the practical use of the impact factor for search purposes remains an open issue.

2.3 Publication outlets as an online search variable

Search engines and online archives support the specification of a keyword-based Boolean expression as an input to search. Keywords are used syntactically by string matching algorithms that search for the specified keywords in the title, abstract or body of papers [9]. Papers are included among query results if they satisfy the Boolean expression entered by the user. This syntactic approach is recognized as a fundamental cause for the information overload experienced by users [26], who are returned papers that satisfy their Boolean expression, but are not relevant to their research domain.

The computer science literature provides several techniques to improve search, known as *semantic* or *content-based* search. The founding idea of semantic search is that keywords are ambiguous and need *disambiguation* to be used effectively. For example, the word "sustainability" has a different meaning in different domains, such as economics, agricultural sciences or computer science. Disambiguation can be achieved by understanding the meaning of a keyword, i.e. its semantics. Semantic search has been widely studied and experimented [26]. Semantic search engines have been proved to be effective, as long as a user provides the engine with enough knowledge for the engine to be able to disambiguate correctly. To the current state of the art, there is no general-purpose semantic engine that can be used effectively across different domains without prior instruction [5].

2.4 Comparative analysis of online archives

The literature makes a distinction between general-purpose and domain-specific online archives and search engines [17]. Google Scholar, Scopus and Web of Science represent the main general-purpose engines. There is a vast literature focusing on the comparison among these three main general-purpose engines. The majority of papers in this stream date back in the years 2005-2007, when the comprehensiveness of general-purpose engines became necessary for the calculation of bibliometric indices, such as the *h*-index and the impact factor.

In 2005, Jacsò [19] noted how Web of Science should not be used alone for locating citations to an author or title, while Scopus and Google Scholar can help identify a considerable number of citations not covered by ISI citation databases [19]. However, Web of Science has been proved to have the best coverage of specific research areas. For example, Web of Science has been found to have the best coverage of South African scholarly research [1]. More recently, Scopus has been claimed to offer about 20% more coverage of citations than Web Of Science, while Google Scholar offers results of inconsistent accuracy [17]. In that same research contribution, authors conclude that Google Scholar can help with the retrieval of "even the most obscure information", but citation information is updated less often. A general answer on which archive provides the most complete set of citing literature does not seem to exist and authors of [4] conclude that the answer depends on the subject and publication year of a given article. Science mapping software tools have emerged to integrate citation information provided by different archives and obtain a citation coverage better than that of any individual archive. Science mapping tools are extensively reviewed in [16], including Bibexcel, CiteSpace, Sci^2, and VantagePoint.

There is substantial agreement in the literature on the superior quality of Scopus search functionalities [12], [19], [24]. A recognized strength of Google Scholar is its simple interface consisting of a query box, but an equally obvious shortcoming is its lack of reliable advanced search functionalities [24]. A comprehensive review of the "odd search behaviours" of Google Scholar's advanced search can be found in [20]. As an example, Page numbers and ISSN four-digit numbers are found to be often interpreted as publication year and, in some cases, the *OR* logical operator reduces the hit count compared to the *AND* logical operator. Scopus is acknowledged to provide a more complete and dependable set of advanced search functionalities organized in a clear and usable interface [12]. In particular, Scopus offers interesting analytics under the "analyze search results" button, which are unique to this archive.

3 Research model and Hypotheses

We have adopted the three-step model of the research process shown in Figure 1. Our model identifies three different search goals that emerge at different times along a research timeline: choosing a research topic, finding readings on a topic, and selecting citations on a topic. Our three-step model is iterative, as each search goal can emerge multiple times as research unfolds. For example, researchers may need to look for readings at the beginning of a new research as well as several times along the research process as new issues arise. It can be observed that our three-step model takes a search perspective and does not make explicit reference to any research method. In this research, the three-step model is considered general and is used to explore the search habits of academics independent of the research method that they use.

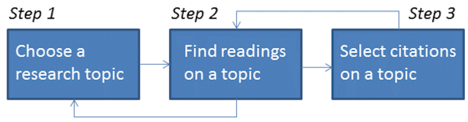

Fig. 1. Search-oriented model of the research process.

The search habits of academics have been rarely addressed in the literature. The research efforts on search habits are summarized in [28]. None of these studies focuses specifically on the literature search habits of academics. However, they are reported to concur that as a general rule "poor quality queries are the main reason for low precision in search engines." Less than 3% of all queries are found to use query operators, such as Boolean expressions. Computer literacy is indicated as an important driver of users' commitment to search and, ultimately, query quality. In this respect, MIS academics can be assumed to represent a highly computer literate community and, thus, a best case scenario for the exploration of search habits and commitment to search.
As an exploratory study, this research considers two general variables describing search habits, namely usage and satisfaction. Different steps in our research model have different search goals that, in turn, may be associated with a varying degree of usage of online search and level of satisfaction with online search functionalities. Consequently, search habits have been separately explored for the three steps in Figure 1.

Selecting a research topic seems to be a natural starting point of a research process. Academics need early insights on new and promising topics and, at the same time, they have to exchange views with colleagues to understand their opinions and interests. Participating in conferences, meeting with representatives from funding institution, visiting companies, and attending workshops and seminars seem more effective ways to choose new research topics. These considerations lead to our first research hypothesis.
H1 – The percentage of academics using online search to choose a research topic (in research step 1) is lower than the corresponding percentage of academics using online search to find readings on a topic (research step 2).

This need for knowing early about new and potentially hot topics as soon as they emerge is addressed by the literature on *weak signals* [21]. Weak signals are defined as hints of a new phenomenon that is growing quickly and, although currently small-scale, is likely to escalate in the near future [15]. In the scientific literature, weak signals represent emerging topics that have the potential to become mainstream soon. In research fields dealing with text analytics, such as information retrieval and social media analytics, weak signals can be discovered by measuring the occurrences of words or patterns of related words [5]. A word or pattern that has a number of occurrences lower than average (is not mainstream yet), but a growth trend higher than average (is growing fast) is a good candidate to represent a weak signal. Text analytic functionalities can support the discovery of weak signals [29]. These functionalities are not offered by online search engines and archives. In our second hypothesis, we posit that these functionalities represent a missing and potentially welcome functionality in the choice of a new topic step, especially among academics who use online search more often.

H2 – In research step 1, academics using online search more frequently rate the benefits from text analytic functionalities higher compared to academics using online search less frequently.

The second research step, finding readings on a topic, heavily involves literature search, in order for an academic to build a map of the state of the art and position her/his own contribution within previous literature. The search goals in step 2 can be considerably different from the search goals in step 1. Choosing a research topic can benefit from a wider breadth of readings, since the so called "weak signals" pointing to emerging and potentially interesting topics can be provided in multiple research fields and then develop into different research streams, each focusing on a specific set of research topics. Conversely, finding readings on a specific topic would benefit from more precise search results and, thus, in step 2 academics may come to the realization that online search, including advanced search, overloads them with information that is only partly related to their research issue. Accordingly, we put forward our third hypothesis:

H3 – In research step 2, academics using online search more frequently have a lower degree of satisfaction with online search and, particularly, with advanced search mechanisms to find readings on a topic, compared to academics who use online search less frequently.

As noted before, keyword-based search has evident limitations that cause information overload. To help cope with this overload, authors can rank search results according to different criteria. Publication time represents the default ranking criterion for several search services and is in fact useful to give higher priority to more recent publications. Other ranking criteria are the publication title or the so-called *relevance*, representing the percentage of matching criteria. None of these criteria guarantees that high-impact papers are ranked on top of search results. As a matter of fact, high-impact papers can be ranked low and be hidden below a large number of less impactful papers.

We have discussed in Section 2 how academics have strived to reach consensus on a quantitative definition of impact and how the *h*-index and the impact factor represent the most widely used quantitative indicators of impact of authors and journals, respectively. We have also noted that general-purpose search engines allow users to view the *h*-index of authors and the impact factor of journals as an aftermath of search by clicking on a specific paper. To the best of our knowledge, no search service allows to use *h*-index and impact factor as search criteria. Would academics appreciate the use of *h*-index and impact factor as search criteria to help filter results according to their impact and, thus, reduce information overload? In our fourth hypothesis, we put forward a negative answer to this question. Similar to H3, our fourth hypothesis is grounded on the observation that *h*-index and impact factors are widely criticized and, on the other hand, no quantitative indicator of impact has emerged as a widely accepted alternative.

H4 – In research step 2, finding readings is based on impact factor and h-index less than it is based on other impact assessment criteria.

We formulate a similar hypothesis also for research step 3 (selection of citations on a topic). Similar to H4, in H5 we hypothesize that selecting citations is based on impact factor and *h*-index less than it is based on other impact assessment criteria. The arguments leading us to this hypothesis start from the observation that there exists a common distinction between “strong” and “weak” citations [8], [18]. For example, a recent workshop where exploratory research results are presented is weaker to support a statement compared to a well-published survey paper providing more consolidated and conclusive evidence. It can be assumed that citations are chosen on the basis of their “strength”, among other criteria.

There is no generally accepted definition of the “strength” of a citation. However, quantitative impact metrics may provide a measure of “strength” on the grounds that impactful papers constitute a stronger citation compared to less impactful papers and, similarly, impactful journals represent a stronger citation compared to less impactful journals. However, the criticisms raised against *h*-index and impact factor suggest that academics may choose their citations based on criteria different from impact metrics. For example, they may select citations because of their closeness to a research topic or to support specific statements. They may even select citations based on their personal knowledge (and judgement) of an author, rather than relying on quantitative indicators. These considerations lead us to our fifth hypothesis.

H5 – In research step 3, selecting citations is based on impact factor and h-index less than it is based on other impact assessment criteria.

4 Empirical evidence

This section describes the empirical testing of our hypotheses. The data sample is described in the next section, while Section 4.2 reports testing results.

Data Sample

Hypotheses are tested on a data sample collected with an extensive survey submitted to academics in the MIS field. Interviewees have been selected as authors of papers published in one of the basket of 8 MIS journals [2] in the period January 2013 – February 2015. The selected MIS journal are: European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of AIS, Journal of Information Technology, Journal of MIS, Journal of Strategic Information Systems, MIS Quarterly. A total of 3544 questionnaires have been submitted, collecting 326 complete responses with a 9.2% response rate. **Fig. 2.** Sample distribution by geographical region.

shows the distribution of respondents by continent.

The questionnaire has been piloted in the time frame March 1 – March 15, 2015 and then extensively submitted through Survey Face (www.surveyface.com) between March 20 and April 22, 2015. The questionnaire and a summary of responses are reported in Appendix (1). It can be noted that questions 6, 7, 10, 11, 15 and 16 address our research hypotheses directly, but the questionnaire includes additional questions that have been considered useful to gain an overall understanding of the search habits of academics. The qualitative results reported in Appendix (1) are discussed in Sect. 5.

Geographical Region		
	Count	*Percentage*
America	132	40.49%
Europe	138	42.33%
Asia	38	11.66%
Oceania	18	5.52%
Tot	326	100.00%

Fig. 2. Sample distribution by geographical region.

Results

All hypotheses have been verified by testing for differences between mean values of two separate groups of respondents. All t-Tests have been performed with a significance value of 0.05 and assuming unequal variances, as f-Tests have indicated that the probability of equal variances was lower than that of having unequal variances. The hypothesized mean difference reported in the tables is the highest value that allowed the rejection of the null hypothesis. Figures 3-7 report testing results for hypotheses 1-5, respectively. All hypotheses are verified with the exception of H3. H3 is not supported since the null hypothesis cannot be rejected (the *t Stat* is lower than minus *t Critical one-tail*). However, the mean values are different, consistent with H3.

5 Discussion

The testing results presented in the previous section provide insights on the literature search habits of academics. First of all, H1 is verified, indicating that academics use online search to find readings on a topic rather than to choose a topic for their future research. Academics do not seem to make decisions on the direction of their research on the basis of information that they gather online. A possible explanation is that the publication process of research results requires a significant amount of time and, as a consequence, the information available from online search engines and archives is not timely. Other information sources, such as workshops, symposiums, and other real-time dissemination initiatives can provide fresh insights on emerging research issues and trends.

The second hypothesis is also verified, suggesting that academics using online search more frequently rate the benefits from text analytic functionalities higher compared to academics using online search less frequently. Answers to question 7 in Appendix (1) show that, on average, 76% of respondents think that it would be useful to have a tool that shows the most frequent topics in scientific papers. As per H2, this percentage is higher for academics relying on online search more heavily. This type of text analytic functionalities supports the aggregate analysis of information and is suitable for the exploration of large data sets. Given the size of online archives, the large-scale aggregate exploration of information is likely to represent a useful tool and our results show that academics are largely aware of this opportunity. The current lack of this type of functionalities may contribute to explain the less intense use of online search for the choice of a new topic (H1), as this research step requires the examination of a broader set of information with an exploratory approach.

	Step1	*Step2*
Mean	3.067692	2.169231
Variance	0.742317	1.363248
Observations	325	325
Hypothesized Mean Difference	1	
df	596	
t Stat	-1.2615	
P(T<=t) one-tail	0.103811	
t Critical one-tail	1.647414	
P(T<=t) two-tail	0.207622	
t Critical two-tail	1.963952	

Fig. 3. Testing H1: Two-Sample t-Test Assuming Unequal Variances.

	More frequent usage	*Less frequent usage*
Mean	3.015748031	2.817258883
Variance	0.777527809	0.905210815
Observations	127	197
Hypothesized Mean Difference	0.35	
df	283	
t Stat	-1.463531791	
P(T<=t) one-tail	0.072215861	
t Critical one-tail	1.650255746	
P(T<=t) two-tail	0.144431722	
t Critical two-tail	1.968381923	

Fig. 4. Testing H2: Two-Sample t-Test Assuming Unequal Variances.

	More frequent usage	*Less frequent usage*
Mean	1.669081884	1.92156851
Variance	0.215272006	0.289537625
Observations	69	255
Hypothesized Mean Difference	0.2	
df	122	
t Stat	-6.936483196	
P(T<=t) one-tail	1.03581E-10	
t Critical one-tail	1.657439499	
P(T<=t) two-tail	2.07162E-10	
t Critical two-tail	1.979599878	

Fig. 5. Testing H3: Two-Sample t-Test Assuming Unequal Variances.

	Other criteria	*h-index and impact factor*
Mean	2.904615385	2.358461538
Variance	0.186475546	0.569411206
Observations	325	325
Hypothesized Mean Difference	0.6	
df	516	
t Stat	-1.116523842	
P(T<=t) one-tail	0.132358806	
t Critical one-tail	1.647812009	
P(T<=t) two-tail	0.264717613	
t Critical two-tail	1.964572029	

Fig. 6. Testing H4: Two-Sample t-Test Assuming Unequal Variances.

	Other criteria	*h-index and impact factor*
Mean	2.552307692	1.792307692
Variance	0.204559926	0.486823362
Observations	325	325
Hypothesized Mean Difference	0.8	
df	555	
t Stat	-0.867245856	
P(T<=t) one-tail	0.193091057	
t Critical one-tail	1.647603773	
P(T<=t) two-tail	0.386182113	
t Critical two-tail	1.964247525	

Fig. 7. Testing H5: Two-Sample t-Test Assuming Unequal Variances.

Empirical data do not support our third hypothesis. H3 posits that academics using online search more frequently have a lower degree of satisfaction with the advanced functionalities of online search. The qualitative results reported in Appendix (1) show that respondents have a generally high degree of satisfaction with advanced search functionalities, as they rate them mostly *good*, although not *excellent* (question 11). This result contrasts with previous literature indicating a low degree of satisfaction with online search functionalities [3]. This contrast is partly mitigated by the fact that the mean values of satisfaction are different in the direction hypothesized in H3 (see **Error! Reference source not found.**), suggesting that users relying more heavily on online search might have a lower degree of satisfaction.

Results indicate that *h*-index and impact factor are used less than other criteria both to find readings (H4) and select citations (H5). In fact, most authors (73%) disagree on the statement that a publication is worth reading when at least one of the authors has a high *h*-index (see Appendix (1) question 15). The majority of respondents (60%) disagree on the statement that they select citations based on the impact factor of the journal where they are published. Almost all respondent (93%) disagree on the statement that they select citations based on the *h*-index of their authors.

Although H4 is verified, most respondents weakly (47%) or strongly (23%) agree on the statement that a publication is worth reading when it is published on a journal with a high impact factor (Appendix (1) question 15). This indicates that academics tend to trust information (the publication) if the information source (the *journal*) is impactful.

A more direct online interaction through academic social media, such as Academia.edu and ResearchGate, does not seem to be among the priorities of our respondents. More than 90% of our interviewees have answered that they do not use Academia.edu or ResearchGate to choose their research topic (Appendix (1) question 6). Similarly, less than 10% uses them regularly (at least once a week) to find readings (Appendix (1) question 10). Very few respondents (less than 20%) think that a publication is worth reading when at least one of the authors belongs to their research circle in either ResearchGate or Academia.edu (Appendix (1) question 15). Less than 10% uses ResearchGate or Academia.edu to disseminate their work, while going to conferences and working on common projects represent the most common dissemination mechanisms (see Appendix (1) question 17). Despite the success of social media in other domains, in academia the peer-reviewed publication system is still seen as the main mechanism for knowledge sharing and consensus building.

In summary, our results confirm that online search plays an important role in satisfying the literature search requirements of academics. They also indicate that academics mainly use Google Scholar and Google's general purpose engine and that they are generally satisfied with them. Advanced search mechanisms, including alerting services and the analytic functionalities of Scopus, are rarely used. This is consistent with previous literature indicating that users are not willing to put an effort into search [28].

6 Concluding Remarks

Our empirical survey conducted with academics in the MIS field has provided the following main results: MIS academics use online search to find readings on a topic, rather than to choose a topic for their future research. Academics using more frequently online search rate the benefits from text analytic functionalities higher than academics using online search less frequently. Both *h*-index and impact factor are used less than other criteria both to find readings on a research topic and select citations.

Query results provided by search engines can be made more precise, and information overload can be reduced, by applying semantic search techniques. However, improving search to reduce information overload does not seem to be a priority for MIS academics, especially if improvements must be achieved at the expenses of simplicity (at the current state of the art, semantic search places an additional burden on users).

The majority of respondents would welcome text analytic functionalities to identify new and potentially trending topics, by performing aggregate analyses on a large number of publications. Whether their need for simplicity applies to this type of functionalities remains an open question for future research efforts.

Acknowledgements

We would like to thank Marco Zamperetti, former Master student at Politecnico di Milano, for his assistance in the early phases of this research. This research has been supported by The Association of Information Technology Trust (AITT, London UK).

References

1. Adriaanse L., Rensleigh C., 2013. “Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison,” The Electronic Library, 31(6):727-744.
2. AISNET, 2016. https://aisnet.org/?SeniorScholarBasket.
3. Baez, M., Birukou, A., Casati, F., Marchese, M. (2010). "Addressing Information Overload in the Scientific Community," IEEE Internet Computing, Vol. 14(6):31-38.
4. Bakkalbasi N., Bauer K., Glover J., Wang L., 2006. “Three options for citation tracking: Google Scholar, Scopus and Web of Science,” Biomedical digital libraries, 3(7).
5. Barbagallo, D. , Bruni, L., Francalanci, C., Giacomazzi, P., Merlo, F., Poli, A., 2011. “Semi-automated Methods for the Annotation and Design of a Semantic Network Designed for Sentiment Analysis of Social Web Content,” Proc. of 10th International Workshop on Web Semantics (WebS11), Toulouse, France, August.
6. Belkadhi, K., & Trabelsi, A. (2020). Toward a stochastically robust normalized impact factor against fraud and scams. *Scientometrics, 124*(3), 1871-1884. doi:10.1007/s11192-020-03577-4
7. Bergstrom, C. T., West, J. D., Wiseman, M. A., 2008. “The Eigenfactor™ Metrics,” *The Journal of Neuroscience*, 28(45):11433-11434, November.
8. Boell, S. K., Cecez-Kecmanovic, D., 2014. “On being ‘systematic’ in literature reviews in IS,” *Journal of Information Technology*, 30:161–173.
9. Boonyoung, T., Mingkhwan, A., 2014. “Semantic search: Document ranking and clustering using computer science ontology and N-grams ,” *Journal of Digital Information Management*, 12(6):369-378, December.
10. Bornmann, L., 2014. “h-index Research in Scientometrics: A Summary,” *Journal of Informetrics*, 8(3), 749-750.
11. Bornmann, L., Mutz, R., Hug, S. E., Daniel, H. D., 2011. “A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants.” *Journal of Informetrics*, 5(3):346–359.
12. Bosman J., van Mourik I., Rasch M., Sieverts E., Verhoeff H., 2006. “Scopus reviewed and compared,” Utrecht University Library, June.
13. CAIS, Special Issue on Self Citation, 2009.
14. Campbell P., 2008. Escape from the impact factor. Inter-Research, 8(1):5-7.
15. Charitonidis, C., Rashid, A., Taylor, P. J., 2015. “Weak Signals as Predictors of Real-World Phenomena in Social Media,” Proc. of IEEE/ACM Int. Conf. on Weak Signals as Predictors of Real-World Phenomena in Social Media, August.
16. Cobo M. J., Lopez-Herrera A. G., Herrera-Viedma E., Herrera F., 2011. “Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools,” Journal of the American Society for Information Science and Technology, 62(7):1382-1402.
17. Falagas M., Pitsouni E., Malietzis G., Pappas G., 2008. “Comparison of PubMed, Scopus, Web of Science, and Google Scholar: stengths and weaknesses,” The FASEB Journal, Life Science Forum, 22:338-342.
18. Hao Wu; Yu Hua; Bo Li; Yijian Pei, 2012. “Enhancing citation recommendation with various evidences,” 2012. IEEE Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD 12), 1160 – 1165, May.
19. Jacsò P., 2005. “Savvy searching, Google Scholar: the pros and the cons,” Online Information Review, 29(2):208-214.
20. Jacsò P., 2008. “Savvy searching: Google Scholar revisited,” Online Information Review, 32(1):102-114.
21. Rossiter, S., Noble, J., Bell, K. R. W., 2010. "Social simulations: Improving Interdisciplinary Understanding of Scientific Positioning and Validity," Journal of Artificial Societies and Social Simulation, 13(1):1-34.

22. Schreiber, M., 2015. "A variant of the h-index to measure recent performance," *Journal of the Association for Information Science and Technology*, 66(11):2373–2380, November.
23. Scopus, Content Coverage Guide, 2014. http://www.elsevier.com/__data/assets/pdf_file/0007/69451/sc_content-coverage-guide_july-2014.pdf
24. Shultz M., 2007. "Comparing test searches in PubMed and Google Scholar," J Med Libr Assoc, 95(4), Oct.
25. Thomson Reuters Community, Citation Impact Center, 2016. http://community.thomsonreuters.com/t5/Citation-Impact-Center/Web-of-Science-Coverage-Expansion/ba-p/10663.
26. Uyar, A. , Aliyu, F.M., 2015. "Evaluating search features of Google Knowledge Graph and Bing Satori: Entity types, list searches and query interfaces," *Online Information Review*, 39(2), 197-213, April.
27. Vanclay JK, 2012. Impact factor: outdated artefact or stepping-stone to journal certification? Scientometrics 92(2):211-238.
28. Wells, J., Truran, M., Goulding, J., 2007. "Search Habits of the Computer Literate," ACM Hypertext Conferece 2007 (HT 2007), Sept.
29. Yoon, J., 2012. "*Detecting weak signals for long-term business opportunities using text mining of Web news*," Elsevier, November.

Appendix (1) – Questionnaire and Qualitative Analysis of Responses

How would you define a "hot research topic"?

		Strongly agree	Weakly agree	Weakly disagre	Strongly disagree
1	A research topic is hot if many researchers are working on it	38.65% (126)	48.47% (158)	9.51% (31)	3.37% (11)
2	A research topic is hot is many companies are investing in research on it	31.9% (104)	47.55% (155)	14.72% (48)	5.83% (19)
3	A research topic is hot if it is included in many calls for funded research	38.04% (124)	46.93% (153)	11.96% (39)	3.07% (10)
4	A research topic is hot if it is the focus of many research publications	32.21% (105)	46.63% (152)	18.1% (59)	3.07% (10)
5	A research topic is hot if papers focusing on it are more likely to be cited	26.99% (88)	43.25% (141)	24.54% (80)	5.21% (17)
6	A research topic is hot if it provides more opportunities to cooperate with other researchers	17.79% (58)	40.18% (131)	34.05% (111)	7.98% (26)
	answered question :	326			
	skipped question :	0			

How do you choose your research topics?

		Always	Frequently	Seldom	Never
1	I do research on what I think is a "hot topic" in my field	5.52% (18)	40.18% (131)	48.47% (158)	5.83% (19)
2	I do research on topics that are most likely to be cited	2.15% (7)	27.61% (90)	55.21% (180)	15.03% (49)
3	My research topics are a consequence of research contracts with companies	2.76% (9)	21.17% (69)	41.41% (135)	34.66% (113)
4	My research topics are a consequence of peer reviewed funded projects	6.44% (21)	28.83% (94)	41.72% (136)	23.01% (75)
5	I choose my research topics as a consequence of suggestions from peer/senior members of my research group	2.76% (9)	26.69% (87)	48.77% (159)	21.78% (71)
6	I choose my research topics according to my personal research interests	70.25% (229)	28.22% (92)	1.53% (5)	0% (0)
7	I choose my research topics according to cooperation opportunities with researchers outside of my research field	2.45% (8)	33.44% (109)	51.23% (167)	12.88% (42)
8	I choose my research topics according to cooperation opportunities with other researchers in my research field	7.98% (26)	57.98% (189)	27.61% (90)	6.44% (21)
9	I choose my research topics based on the directions provided by the most impactful authors in my field	2.45% (8)	23.01% (75)	48.16% (157)	26.38% (86)

4. How frequently do you change your set of research topics?

		Response Percent
1	Every year	4.29%
2	Every 2 years	10.43%
3	Every 3 years	22.39%
4	Every 4 years	18.10%
5	Every 5 years	21.47%
6	Between 5 and 10 years	21.47%
7	Never	1.84%

5. Why do you move away from a research topic?

		Strongly agree	Weakly agree	Weakly disagre	Strongly disagree
1	Because it is no longer a "hot topic" in my field	5.83% (19)	27.61% (90)	39.57% (129)	26.99% (88)
2	Because the related project has ended	27.61% (90)	46.32% (151)	17.18% (56)	8.9% (29)
3	Because of lack of funds	7.06% (23)	28.53% (93)	31.9% (104)	32.52% (106)
4	Because peer/senior members in my research group advised me to do so	3.37% (11)	10.43% (34)	27.61% (90)	58.59% (191)
5	Because I plan to move to a different research institution	3.68% (12)	15.64% (51)	21.17% (69)	59.51% (194)
6	Because I plan to move to a different research group	4.6% (15)	16.87% (55)	23.31% (76)	55.21% (180)
7	Because of a personal loss of intellectual interest	55.52% (181)	33.13% (108)	6.13% (20)	5.21% (17)

6. Do you use online archives and search engines to choose a research topic?

		Strongly agree	Weakly agree	Weakly disagre	Strongly disagree
1	I do not use them to choose a research topic	41.41% (135)	19.33% (63)	19.63% (64)	19.63% (64)
2	I use Google Scholar	30.67% (100)	26.69% (87)	14.11% (46)	28.53% (93)
3	I use Google's general purpose search engine	18.4% (60)	27.3% (89)	19.33% (63)	34.97% (114)
4	I use Scopus	3.68% (12)	10.12% (33)	17.48% (57)	68.71% (224)
5	I use the analytic functionalities of Scopus ("Analyze search results" button)	1.53% (5)	4.29% (14)	17.79% (58)	76.38% (249)
6	I use Web of Science	7.98% (26)	16.26% (53)	17.79% (58)	57.98% (189)
7	I use ResearchGate.net	2.15% (7)	15.34% (50)	26.38% (86)	56.13% (183)
8	I use Academia.edu	1.23% (4)	6.75% (22)	20.55% (67)	71.47% (233)
9	I use my field's vertical search engines (e.g. IEEE for engineering, Pubmed for medicine, etc.)	14.72% (48)	19.02% (62)	14.11% (46)	52.15% (170)

7. Do you think that it would be useful to have a tool that allows you to automatically gather the most frequent topics in scientific papers?

		Response Perc	Response Count
1	Yes	26.07%	85
2	May be yes	50%	163
3	May be no	11.66%	38
4	No	12.27%	40

8. How do you find readings related to your research topics? - Conferences

		More than 4 tir	2-4 times a yea	Once a year	Less that once a	Never
1	I attend conferences in my field	10.43% (34)	49.39% (161)	29.14% (95)	11.04% (36)	0% (0)
2	I serve as a reviewer for conferences in my field	29.14% (95)	45.4% (148)	17.79% (58)	4.91% (16)	2.76% (9)

9. How do you find readings related to your research topics? - I serve as an editor for journals in my field

		Response Perc	Response Count
1	More than 3	9.20%	30
2	Three	8.59%	28
3	Two	16.26%	53
4	One	24.85%	81
5	None	41.10%	134

10. How do you find readings related to your research topics? - Other sources

		Every day	Every week	Every month	A few times in a	Never
1	I use Google Scholar	21.17% (69)	43.25% (141)	19.33% (63)	10.74% (35)	5.52% (18)
2	I use Google's general purpose search engine	22.7% (74)	30.98% (101)	21.47% (70)	14.11% (46)	10.74% (35)
3	I use Scopus	1.53% (5)	5.21% (17)	11.66% (38)	20.86% (68)	60.74% (198)
4	I use Web of Science	2.76% (9)	9.2% (30)	15.03% (49)	27.91% (91)	45.09% (147)
5	I use ResearchGate.net	0.31% (1)	10.12% (33)	17.79% (58)	26.38% (86)	45.4% (148)
6	I use Academia.edu	0% (0)	1.23% (4)	7.98% (26)	18.1% (59)	72.7% (237)
7	I use my field's vertical search engines (e.g. IEEE for engineering, Pubmed for medicine, etc.)	4.6% (15)	19.02% (62)	18.1% (59)	19.33% (63)	38.96% (127)
8	I select interesting readings amond the references of papers that I have read	17.48% (57)	43.87% (143)	25.46% (83)	10.43% (34)	2.76% (9)
9	I receive paper versions of the relevant journals in my field	3.99% (13)	10.12% (33)	17.79% (58)	23.31% (76)	44.79% (146)

11. How do you rate the effectiveness of the following tools to search for readings related to your research topics?

		Excellent	Good	Fair	Bad	I never use it
1	Google Scholar's advanced search	31.9% (104)	47.85% (156)	13.19% (43)	1.84% (6)	5.21% (17)
2	Scopus advanced search	3.68% (12)	15.95% (52)	12.88% (42)	1.84% (6)	65.64% (214)
3	Web of Science's advanced search	6.44% (21)	19.33% (63)	21.47% (70)	4.29% (14)	48.47% (158)
4	The advanced search of your field's vertical search engines (e.g. IEEE for engineering, Pubmed for medicine, etc.)	10.43% (34)	25.77% (84)	20.86% (68)	2.45% (8)	40.49% (132)
5	ResearchGate.net	0.92% (3)	10.43% (34)	26.69% (87)	5.21% (17)	56.75% (185)
6	Academia.edu	0.61% (2)	5.52% (18)	12.88% (42)	2.15% (7)	78.83% (257)

12. What type of readings do you look for?

		Always	Frequently	Seldom	Never
1	Academic papers	85.58% (279)	14.11% (46)	0.31% (1)	0% (0)
2	Company white papers	2.76% (9)	21.17% (69)	52.45% (171)	23.62% (77)
3	Patents	0% (0)	1.84% (6)	22.7% (74)	75.46% (246)
4	Slides	1.53% (5)	16.26% (53)	50.92% (166)	31.29% (102)
5	Videos	1.53% (5)	11.66% (38)	48.47% (158)	38.34% (125)

13. What are the characteristics of your readings?

		Always	Frequently	Seldom	Never
1	They discuss the results of theoretical research	33.44% (109)	54.29% (177)	11.96% (39)	0.31% (1)
2	They discuss the results of empirical research	33.13% (108)	63.19% (206)	3.68% (12)	0% (0)
3	They describe case studies	11.35% (37)	50.92% (166)	32.82% (107)	4.91% (16)

14. Do you use push services that alert you when new research is available that might be of interest for you?

		Always	Frequently	Seldom	Never
1	I use Scopus alerting services that let me know when there is a new publication that matches one of my searches	2.45% (8)	3.68% (12)	9.51% (31)	84.36% (275)
2	I use ResearchGate alerting services that let me know when a researcher in my network has uploaded a new publication	6.75% (22)	11.96% (39)	16.87% (55)	64.42% (210)
3	I use Academia.edu alerting services that let me know when a researcher in my network has uploaded a new publication	1.23% (4)	1.84% (6)	11.66% (38)	85.28% (278)

15. What are the drivers that make you think that a publication is worth reading?

		Strongly agree	Weakly agree	Weakly disagre	Strongly disagree
1	A publication is worth reading when it has a high number of citations	22.7% (74)	49.08% (160)	19.94% (65)	8.28% (27)
2	A publication is worth reading when it tackles a practical problem	34.36% (112)	49.39% (161)	14.11% (46)	2.15% (7)
3	A publication is worth reading when it tackles a theoretical problem	38.96% (127)	50% (163)	8.59% (28)	2.45% (8)
4	A publication is worth reading when it is closely related to your reasearch topics	80.98% (264)	16.56% (54)	2.15% (7)	0.31% (1)
5	A publication is worth reading when it presents research results that are practically applicable and useful	46.32% (151)	39.88% (130)	11.66% (38)	2.15% (7)
6	A publication is worth reading when at least one of the authors has a high h-index	4.91% (16)	21.47% (70)	32.21% (105)	41.41% (135)
7	A publication is worth reading when at least one of the authors is a well known established researcher in your field	15.34% (50)	48.47% (158)	26.99% (88)	9.2% (30)
8	A publication is worth reading when I personally know at least one of the authors	13.19% (43)	40.8% (133)	28.22% (92)	17.79% (58)
9	A publication is worth reading when at least one of the authors belongs to my research circle in either ResearchGate or Academia.edu	2.45% (8)	18.71% (61)	23.62% (77)	55.21% (180)
10	A publication is worth reading when it is published in a journal that has a high impact factor	22.7% (74)	46.63% (152)	20.25% (66)	10.43% (34)

16. How do you choose the citations to be included in your own publications?

		Always	Frequently	Seldom	Never
1	I choose citations based on their closeness to my research topics	62.27% (203)	33.13% (108)	3.37% (11)	1.23% (4)
2	I choose citations that provide important evidence to support specific statements in my publication	73.31% (239)	24.54% (80)	1.23% (4)	0.92% (3)
3	I choose citations based on my personal knowledge of at least one of the authors	3.99% (13)	19.33% (63)	45.4% (148)	31.29% (102)
4	I choose citations based on the impact factor of their publication outlet	7.36% (24)	32.52% (106)	31.29% (102)	28.83% (94)
5	I choose citations where at least one of the authors has a high h-index	0.61% (2)	6.44% (21)	25.46% (83)	67.48% (220)
6	I choose citations based on my target publication outlet	13.19% (43)	43.87% (143)	30.98% (101)	11.96% (39)
7	I make an effort to include highly cited papers among my citations	7.98% (26)	28.53% (93)	35.28% (115)	28.22% (92)
8	I cite authors who are most likely to cite me back	1.23% (4)	5.52% (18)	18.71% (61)	74.54% (243)

17. How do you work to disseminate your own research work?

		Always	Frequently	Seldom	Never
1	I personally go to the conferences where my papers are published	30.37% (99)	51.84% (169)	14.72% (48)	3.07% (10)
2	I reach out to other researchers in my field on ResearchGate	3.68% (12)	11.35% (37)	30.37% (99)	54.6% (178)
3	I reach out to other researchers in my field on Academia.edu	1.23% (4)	2.45% (8)	15.03% (49)	81.29% (265)
4	I create a network of contacts through cooperation in common research projects	11.66% (38)	44.79% (146)	25.15% (82)	18.4% (60)
5	I circulate my own papers before publication to get early feedback	8.59% (28)	32.21% (105)	39.88% (130)	19.33% (63)
6	I circulate my own papers before publication to engage other researchers	4.91% (16)	20.55% (67)	44.79% (146)	29.75% (97)