



**The impact of Google on discovering scholarly information:
Managing STM Publishers' visibility in Google**

Journal:	<i>Collection and Curation</i>
Manuscript ID	CC-01-2020-0002.R1
Manuscript Type:	Research Paper
Keywords:	discoverability, promotion strategies, search engine results, search visibility, STM publishers, search engine optimization

SCHOLARONE™
Manuscripts

The impact of Google on discovering scholarly information: Managing STM Publishers’ visibility in Google

Abstract

Purpose: The purpose of this paper is to discuss Google visibility of five large STM publishers (Elsevier, Emerald Publishing, Springer, Taylor & Francis and John Wiley & Sons) with the aim to focus on and investigate various upcoming current issues and challenges of the publishing industry regarding discoverability, promotion strategies, competition, information seeking behavior, and the impact of new information technologies on scholarly information.

Design/methodology/approach: The study is based on data retrieved through two commercial online tools specialized in retrieving and saving the data of the domain's visibility in search engines: SEMrush (“SEMrush - Online Visibility Management Platform”) and Ahrefs (“Ahrefs - SEO Tools & Resources To Grow Your Search Traffic”). All data gathering took place between the 15th of April and the 29th of May 2019.

Findings: The study exhibits the significance of Google visibility in the STM publishing industry taking into consideration current issues and challenges of the publishing activity.

Originality/value: This is a “new” trend, certainly of great significance in the publishing industry. The research is conducted in this paper and the theoretical background will be offered to the study of this issue.

Keywords: discoverability, promotion strategies, search engine results, search visibility, STM publishers, search engine optimization.

Article Classification: Research paper

Introduction

During the last decades, a number of significant changes has taken place in the publishing industry. Main features of the publishing industry have been discussed and highlighted in a number of works (Thompson, 2010; Phillips, 2014; Banou, 2017; Clark & Phillips, 2014; Miller, 2007; Striphas, 2009; Greco *et al.*, 2007). The Scientific, Technical, and Medical (STM) publishing industry is actually characterized by mergers and acquisitions leading to the dominance of large publishing companies and conglomerates, the emerging role of new information and communication technologies, novel publishing business models, open access, new promotion and marketing strategies, reader's engagement, online scholar communities etc. Discoverability as a key issue means that visibility in search engines is considered to be a valuable aspect in publishers' promotion that also influences and develops the information seeking behavior. In that framework, old roles in the publishing chain are redefined while new ones mostly towards the digital publishing space have been emerged (Banou, 2017), engaging authors, readers and other stakeholders in the content creation chain to a circular, networked, in many cases web-based environment (Lloyd, 2008).

It is estimated that there are about 10.000 journal publishers existing worldwide. According to Johnson *et al.*, (2018), about 50% of the total journal output by title (about 11.550 journals) represents 650 publishers of the main English-language trade and professional associations of journal publishers. 73% of these (about 480 publishers) and 20% of the journals (about 2.334 journal titles) are not-for-profit. Also, an earlier directory between 2000 and 2013 suggested that about half of all journals came from not-for-profits. According to Ware (2006), an Elsevier analysis of the ISI Journal Citation database indicated that the article proportions' output were: 64% of commercial publishers (including publishing for societies), 30% of society publishers, 4% of University presses, and 2% of other publishers. The distribution of journals by publisher is highly asymmetric; it is noteworthy that eleven (11) publishers (2% of the publishers) publish more than the 70% of the journals in this group which is about the 35% of all journals. Indicatively, two publishers (Elsevier and Springer) produce 2.000 journals each. Thus, a concentration on scientific publishing can be observed, as in other publishing fields, with undeniable "leaders" a few publishing companies which exploit the opportunities are provided by new information and communication technologies. Online scholar communities of readers are being developed as well. In that context, discoverability and visibility are high on the publishers' priorities, as they aspire to reach new readers and penetrate into new audiences.

Processes and strategies of the STM publishing companies for delivering and making scientific information visible have notably changed over the last few years (Rzepa & Murray-Rust, 2001; Thomspson 2010). It is **also** noteworthy that discoverability and accessibility of online scholarly information are among the core values and aims of STM publishing companies (Zhu & Kelley, 2015; Phillips, 2014), **that** aspire to build a relationship of trust and communication with their audiences. Over the last decade, web-based services and strategies that aim to the discovery, dissemination and sharing of scholarly information have emerged as new products and tools, and undeniably as a novel commodity (Frederiksen, 2015) that reshapes information and consumption cultures. On the other hand, we notice that provided scholarly visibility is assured **and** academic or government research departments that may not even regard themselves as “publishers” can actually lounge successfully STM journals (Ware and Mabe, 2015). Publishers’ visibility in search engines is rather crucial. Currently, search engines, such as Google and Yahoo, **could** be considered as the primary platforms **of** online scholarly information discovery.

Search engines provide a rather holistic view of the ever increasing online scholarly information resources and information sharing platforms (Case & Given, 2016). **The** adoption of search engines by scholars **in order to satisfy** a wide range of their information needs is gaining popularity (Luh *et al.*, 2016; Umenhofer, 2019). Publishers, on the other hand, have realized that high visibility in search engines brings about readers and customers; inevitably, they develop strategies **to gain** visibility and discoverability (Lee *et al.*, 2016).

Search engine visibility is measured through **the** ranking position and the number of keywords and pages **that** are visible in the search engine results page (Dickinson & Smit, 2015). Hence, publishers’ visibility comparison **refers** to the number of keywords, position and number of items that are visible in the search engine result page (Dickinson & Smit, 2016). The comparison will disclose what is the search market share of each compared publisher (French & Fagan, 2019). Based on this comparison further analysis of internet strategies in STM publishing, marketing and promotion can be **also** set. Visibility in search engines is always **a** subject to **the** algorithm that **sorts** and sets **the** ranking **of the** results based on the **content’s type, the** metadata, **the** models of content creation (Killoran, 2013; Miklosik *et al.*, 2019; Zhang & Dimitroff, 2005) or **hides** results due to different reasons (Strzelecki, 2019).

This work focuses on **Google’s** search engine visibility data for five Scientific, Technical and Medical (STM) Publishers (Okerson, 1996; Teixeira da Silva & Dobránszki, 2016): Elsevier,

Emerald Publishing, Springer, Taylor & Francis and John Wiley & Sons. Data is retrieved through two commercial online tools specialized in retrieving and saving the data of the domain's visibility in the following search engines: SEMrush ("SEMrush - Online Visibility Management Platform" n.d.) and Ahrefs ("Ahrefs - SEO Tools & Resources To Grow Your Search Traffic," n.d.). SEMrush and Ahrefs except preserving basic visibility, import additional data and develop their own visibility metrics. Some predictions on the visibility of STM publishers are done by the use of Google Trends (Poulos *et al.*, 2017). This dataset is prepared based on Google search engine results page. Since Google is currently the most popular search engine ("Statista - Global market share of search engines 2010-2019", n.d.), the collected data should be of better quality and more precise than the ones from other search engines.

Background knowledge

Discoverability is certainly a key issue in publishing (Phillips, 2014) and is connected with information evaluation (Greco, 2013: 3): *"Every book is a new product. Readers discover their preferences and spread information, both positive and negative, via the information cascade"*. Readers discover scientific material according to their level of subject experience and their information literacy skills and competences. Discovering scholar information goes beyond and is related to the searching material presupposing the updated services for conferences, events, publication trends and ongoing publication opportunities. Scholars often discover and evaluate information through keywords and advanced searching in search engines. In that context, STM publishers explore (and have to further explore) and develop the novel information and communication technologies so as to better inform, promote the published material, and indeed penetrate into new scholar audience (Hunter, 2007).

In the STM publishing, readers have more than one role at the same time, by being authors, reviewers, editors, members of the scientific committee or the editorial board, as translators and members (often active) of scholar communities (Johnson *et al.*, 2018; Banou, 2017). This is considered to be an advantage for the publishers who collaborate with them (Zhu & Kelley, 2015). Scholars need to be up to date and find the required information in one place using a specific search engine (Michaels, 2015). In that framework, readers have access, discover, use and share scholarly information through online platforms that support publishing services for the discovery, access and delivery of electronic resources (Somerville & Conrad, 2013). Therefore, publishers monitor their products' discovery through the collaboration with search

engines such as Google, Yahoo and Bing in order to improve their visibility to scholarly community, making modifications whenever necessary (Somerville *et al.*, 2012).

Major STM publishers are inevitably the protagonists. *“Anyone who assumes that book publishing firms will be pushed aside because of the growth of the self-publishing authors or operations just does not understand the innovative strategies and operational structures that have been crafted in the last few years by a cluster of major publishers ...The industry’s great traditions, and its unyielding, and at times undisciplined, quest for perfection enables it to handle effectively depressions, recessions, technological convergence, war and social upheavals”* (Greco *et al.*, 2007: xii, xiv). STM publishers play a leading role in the publishing chain as key partners of the scholarly community, developing and often establishing activities and methods such as registration, certification, formalization, improvisation, dissemination, preservation, and use of scholarly information (Ware & Mabe, 2015). That leading role of STM publishers needs to be discoverable. This is achieved through the establishment of collaborations between all those involved in the publishing chain (publishers, scholarly community, agents, libraries, bookstores, communities of readers, social media etc.). In that context, the development of online services by search engines such as Google, Yahoo, Bing, etc. (Somerville & Conrad, 2014; Zhu & Kelley, 2015) has to be studied.

Search engine optimization and implementation is undeniably one of the STM publishers’ challenges. After a historical SEO (Search Engine Optimization) research, Umenhofer (2019) has showed that the three most important SEO elements include keywords, title tags and meta tag descriptions. Large STM publishing companies (such as Elsevier, Emerald Publishing, Springer, Taylor & Francis, John Wiley & Sons) have the advantage (in comparison to small, independent publishers) of specific *ad hoc* departments, of highly qualified staff, marketing opportunities and experience. Promotion strategies (that include search visibility) are set and experts working at the publishing chain (within the publishing company) can better advice the authors and certainly gain the data and metadata required. Metadata provided by scholars and/or publishers, like document description and keywords, are used by Google in building its ranking. However, metadata accuracy is improved through vocabulary management solutions (Dictionary of Publishing & Printing, 2006) which are semantics-based metadata repositories that are developed by the publishers and are embedded in their publishing process. These are used for automated code and data generation to be employed across various scholarly communication platforms and search engines.

Meanwhile, established academic publishers increase their resource investments in order to measure and optimize scholarly content for web discovery and to produce online products for search engine optimization (Somerville & Conrad, 2014). This concerns search engine optimization for mainstream search engines such as Google, Yahoo, Bing, etc. For their part, publisher platform providers are intended for possession of easily found and well-presented online content designing services for content discovery, while academic publishers produce discovery tools that anticipate the research workflow of scholarly community. In that context, this shaped scholarly environment extends the boundaries for discoverability of scholarly information as well as its dissemination, visibility and usage on the web (Somerville & Conrad, 2013).

As online scholarly information is more than important in a reader's life, most readers use to be informed by online information resources, such as the publishers' web pages, online bookstores and social media. The traditional "word of mouth" has become mostly online and Google search engine takes a leading role as many readers tend to follow the hierarchy appeared on the page (Beckwith, 2003). Google, despite not being able to index the entire web, like all the rest of search engines (e.g., Yahoo, Bing, etc.), is the most frequently used search engine (Norris *et al.*, 2008). It is effective and successful taking also into consideration Google scholar which focuses exclusively on scholarly information (Markland, 2006). According to the survey of Jamali & Asadi (2010), Google general search engine is the most used tool from scientists to the discovery of scholarly information. In particular, scientists become more aware of the quantity of searchable scholarly information by a general search engine, relying on Google for scholarly literature discoverability. Thus, Google as the most popular search engine for students (Griffiths and Brophy, 2005) and academic scientists (Hemminger *et al.*, 2007), has an impact on scholarly information discovery. Google scholar initially indexes documents and displays them in search results based on a specially developed automated algorithm. However, after the initial indexing, Google attempts to benefit from human hand indexing of documents, mainly by the authors, which potentially add value and improve the indexing. Furthermore, Google, based on human judgment, adapts the searching algorithm in areas such as science, finance, medical, legal and other important scholarly information areas (Strzelecki, 2020). The study of Hariri (2011) assessed the effectiveness of Google search engine's relevance ranking, by comparing users' views with the system's assessment of relevance.

The following sections present the methodological framework as well as the results of the survey about Google search engine visibility data for five STM publishers: Elsevier, Emerald Publishing, Springer, Taylor & Francis and John Wiley & Sons. The basic visibility of a domain name in search engines can be evaluated based on the number of keywords that are entered to search engines; they are triggering results with: a. researched domain name, b. ranking position of URLs that belong to researched domain name, and c. sum of pages from research domain that are available in the search engine results page.

Methods

The method of collecting data from a search engine is named scraping (Marres & Weltevrede, 2015). Usually major search engines in their terms of service do not allow data scraping. However, it is impossible for a search engine to recognize, whether scraping is done very gently or in its normal user search behavior. Users are using search engines, dozens of times a day and only if the search engine recognizes a different traffic from user's network, it can ask for a captcha in order to solve and prove that the entered queries are not automated. Google does not share any download or export methods for SERPs, neither provides an API to exporting search results. The only way to obtain data is to scrape them directly from SERPs. Scraping Google is against their TOS.

For data scraping there are several available tools, like scripts written in Python, which allows results' scraping from Google's search engine ("google PyPI," n.d.; "hyperion-gray / googlespider," n.d.; "Scrapeulous. - Search Engine Scraping" n.d.) or online tools such as SEMrush and Ahrefs which allows registered users to use their scraped data. These tools use the scraping method in a large scale to obtain data from Google. Data is retrieved through two commercial online tools: SEMrush ("SEMrush - Online Visibility Management Platform," n.d.) and Ahrefs ("Ahrefs - SEO Tools & Resources To Grow Your Search Traffic," n.d.). Both tools are specialized in retrieving and saving the data of the domain's visibility in search engines. Basic search engine visibility is the combined data of a unique keyword, position and URL result.

SEMrush and Ahrefs except they preserve basic visibility, import additional data and develop their own visibility metrics as well. We used this data set to compare search engine visibility of these five STM publishers. They compete in the same scientific press area. Some predictions of the STM publishers' visibility are done with the use of Google Trends (Poulos

et al., 2017). This dataset is prepared based on Google search engine results page. Since Google is currently the **most** popular search engine (“Statista - Global market share of search engines 2010-2019” n.d.), collected data should be **of** better quality and more precise than from other search engines.

We started by collecting keywords connected with STM publishers. **The** above mentioned tools are commercial, so the purchased plan allowed us to download limited data. From SEMrush we obtained 30.000 keywords for each STM publisher domain, and from Ahrefs we obtained 10.000 keywords along with additional parameters for each keyword. **Commonly analyzed** parameters include: keyword, position, URL, search volume and traffic. From both tools, the obtained dataset was sorted in descending order by traffic volume. This means that we get the most **of the** performing keywords in terms of search engine visibility. Data **on** keywords was collected in period between **the** 15th of April and **the** 29th of May 2019 by both sources. **These data was downloaded** on **the** 29th of May 2019. Thereafter, our analysis included total URL visibility. Hence, the visibility on **the** URL level for each STM published was estimated. SEMrush allowed **downloading** 5.000 URLs along with **the** traffic volume **of** each URL and Ahrefs allowed us to download 10.000 URLs with **the** traffic volume **as well**. **The** traffic volume for **the** URL **combines** traffic for every keyword that will show results of a particular URL in Google search engine. This dataset contains data **from** the United States, which is available under www.google.com and the limit of the displayed results from one query is set to 50. The dataset contains only regular results and only from official domains of STM publishers, including subdomains.

Results

The dataset for the first keyword study is a set of 10 CSV files separated with semicolon. Each file has a similar structure containing five columns: keyword, position, URL search volume and traffic. SEMrush and Ahrefs put **also** into search visibility other parameters displayed in columns. However, **the five mentioned above** are common for both tools. The keyword column contains keywords, also known as queries. In each row the results **of** one unique keyword **are displayed**. **A** keyword may contain one or more terms. The upper limit is usually set by **the** search engines, e.g. **the** upper limit for terms in one query set by Google is twenty-five. The position column contains the results in numeric values ranging from 1 to 50, describing the position of results for a previously entered keyword. The position is estimated only in organic results listed in Google search engine results page (SERP). The position is not

measured outside **ten (10)** regular blue links. The SERP has also additional types of results called snippets. Snippets can enrich regular results, and there **are separate featured news and entity types snippets**. If the result is in **the** first position, **it has** sometimes additional site links, placed below **the** first result, with more results from the same domain name. They are also counted as **the** first position for the same keyword, but with another URL.

The URL column contains URL addresses: **The** URL address is presented in SERP as a result of entering a keyword into a search engine. URLs in each file are only from one domain name. Domain names studied for visibility in Google search engine for five STM publishers are:

- elsevier.com for Elsevier,
- emeraldinsight.com for Emerald Publishing
- springer.com for Springer,
- tandfonline.com for Taylor & Francis,
- wiley.com for John Wiley & Sons.

Data collected for the above **mentioned** five domain names is not limited to the main domain names but is **also** collected for every subdomain set on the main domain. This approach allows **the** collection of results for subdomains. This was required since STM publishers use subdomains for their publishing platforms. They publish on subdomain online libraries and records with access to subscribers.

Search volume is the volume of monthly keyword searching. This piece of data is imported by both tools from Google Ads platform through API. Search volume allows traffic estimation for a particular keyword. Traffic is an internal metric developed by these both tools, however, **each one** did this independently. Traffic estimation is based on position, search volume and Projected Click Trough Ratio (PCTR) for **a specific** keyword. PCTR includes the assumption that a user will click this particular result with an estimated probability ranging from 1% to 100%. PCTR is not revealed publicly, however, both tools employed publish estimated search traffic for each keyword. Search traffic differs **from** the same keyword included in datasets coming from both tools. It differs because they use different PCTR, scrape keywords and results on different days and may add additional, not **publicly** known factors **in order** to estimate **the** traffic. Thereafter **the** traffic volume for each URL was analyzed. This involves

the URL and a total estimated traffic for this URL from all keywords, pointing in results to that URL.

[Insert Table 1 near here]

Table 1 includes data on STM visibility on keywords level in Google search engine. Data is obtained independently from the two sources, i.e. Ahrefs and SEMrush. The column that estimated traffic shows the sum of the traffic by examining a set of keywords: 10k keywords from Ahrefs and 30k keywords from SEMrush per domain. Thus, the normalized value is then calculated as follows:

$$\text{Normalized value} = \frac{\text{estimated traffic}}{\text{best estimated traffic}} \times 100$$

[Insert Figure 1 near here]

In Table 2 the descriptive data on STM visibility on page level in Google search engine are portrayed. For this data two independent sources are employed. The Column that estimated traffic shows the traffic's sum from an examined set of pages: 10k pages from Ahrefs and 5k keywords from SEMrush per domain.

[Insert Table 2 near here]

Discussion

Scholars need to locate efficiently and effectively trusted online content in order to satisfy their information needs for research and other scholarly activities (Somerville *et al.*, 2012). The rise of Google and Google Scholar has significantly affected the way the research

community behaves towards discovery and accessibility of scholarly information (Zhu & Kelley, 2015). STM publishers gradually redefine and redevelop their strategies by including in their online platforms advanced services for all the aspects of the publishing chain including paper submission, evaluation and review, editing and publication (Guédon *et al.*, 2019). In this regard, it is quite important for STM publishers to be aware of Google visibility results on both keywords and page levels and, thereafter, inform their online publishing platforms, databases, abstracting and indexing services etc. (Somerville & Conrad, 2014).

[Insert Figure 2 near here]

Google's information searching is, for many, synonymous to the use of the internet (Case & Given, 2016). Indeed, the rise of Google search engine presents both opportunities and challenges for libraries, publishers, and service providers (Somerville & Conrad, 2014). The ubiquity of scholarly information **searched** on Google has led to a discussion regarding STM online platforms. Figure 2 positions publisher's Google visibility in the publishing cycle developed by Johnson *et al.*, (2018). Google is extensively used by the research communities all over the world. Researchers (staff and students of all disciplines and levels) are increasingly relying on Google **in order to** find scholarly literature as a unique source itself (Jamali and Asadi, 2009; Wu and Chen, 2012). Another aspect of **Google's** extensive use has to do with the required online information literacy skills (Mayr and Walter, 2007). Indeed, Google provides convenience and **requires** very basic online searching skills; while other platforms require more advanced information seeking skills such as resources/databases selection, evaluation and efficient search strategies (Jamali and Asadi, 2009). Google is a user-friendly search engine that can give scholars a quick view of some of the scholarly information on a topic, and some of the related issues addressed (Kesselman and Watstein, 2005). However, sometimes Google results may be confusing; as often **that** a single paper is listed several times coming from different sources as well as different versions of the same paper might also exist. Publishers' databases do provide advanced search tools allowing scholars to focus on the specific aspects of their research.

Banou (2017) discusses the publishing chain as information publishing chain-circuit **that is** enlightening the emerging role of information technologies. **Characteristically**, the ungraded

role of the reader (who is in STM publishing also a member of the scholar community often with a variety of roles: reviewer, editor, author, reader, professor who orders books and writes book reviews, etc.) and the ever central role of the publishers which have managed to maintain their influence and power **is discussed**. More specifically, Banou (2017) argues that STM publishing companies face challenges and exploit opportunities that derive from information and communication technologies; but their strategies go deeper exploiting and taking advantage of the readers' information seeking behavior, literary skills and evaluation of information. As Clark and Phillips (2014: 21) write: *"In a world of abundance, the publishers offer a vital service in selecting authors and developing their content to meet readers' needs. They manage the author's brands and focus readers on the books they have selected. That service is worth paying for when time is scarce. To attempt another definition: the publishing process may be described as managing the scarcity of good authors and content to drive profitability"*; Chi (2014: 346) further suggest that *"it's no longer enough to provide information, we need to help our users find the right information in the right context at the right time. Now, we need to provide answers, not just information"*. In that framework, publishers aspire to reach their audience and penetrate into new audiences by exploiting Google (and other search engines) as well. We **also** have to note that one of the aims of the publishers nowadays might be a "relationship based on trust" with their readers who will then regularly use their online services for being informed and engaged (Banou, 2017; Phillips, 2014).

Indeed, large STM publishers, apart from information providing and promotion, focus on developing and empowering **readers'** engagement through a number of ways among which may recognize their scholar online communities. In a changing world, where authors and readers have a variety of options, "trust" is an important factor in attracting and encouraging readers to use publisher's services. Furthermore, publishers try to create cultures of information seeking and sharing by developing specific services among which we may recognize online reading and project/subject groups/communities. Departments in the publishing houses have also been added enriching the publishing chain. In that context, **Google's** visibility is undoubtedly a great opportunity; STM publishers that appear first at Google searching have **the** advantage **to proceed** to the next steps of information trust, searching, sharing, evaluating etc. and certainly developing "loyal" audiences and further encouraging **readers'** participation. **Scholars due to the time pressure and other situational constraints tend to visit the first publishers that appear in a search machine; according to**

Umenhofer (2019: 259), *"An independent book publisher would be keen on understanding that if they can get their name to the top of, or near to the top of, one of the three major search engines (Google, Bing, and Yahoo!), then they can increase their website traffic substantially"*. The level of publishers visibility in search engines apparently is a strong predictor for information searching habits, scholarly practices, communication patterns and citation behaviour of scholars of various disciplines (Case and Given, 2016). On the other hand, habitual scholarly information seeking might encounter a trade-off between the depth and the breadth of the scholarly information seeking.

Conclusion

Considering the impact of publishers' efforts for discoverability and visibility on scholarly information seeking one might reflect on McLuhan's contemplations through Lapham (1997: xii) fascinating note *"content follows form, and the insurgent [information] technologies give rise to new structures of feeling and thought"*, provided in the introduction of *"Understanding Media: the Extensions of Man"* (McLuhan, 1997). Publishers' constantly introduce new information technologies which in turn redevelop readers'/customers' behaviour, interaction and expressions. Over the last decade the literature emphasizes the impact of online content and online technologies and the internet on the information seeking habits and practices of scholars working in most, if not all disciplines and study areas (Case and Given, 2016). More work is required in order to understand these phenomena and to negotiate the structure and the values of the publishing industry for the benefit of the community and scholars' information needs satisfaction. Undoubtedly, search engine optimization constitutes a challenge and an opportunity for STM publishers regarding their promotion, marketing and audience development strategies. In an everyday changing publishing world, promotion methods seem to be at a constant crossroad where publishers have to better exploit information and communication opportunities, the internet and in that context Google search engine so as to reach existing and future audiences. More and more people trust search engines for discovering titles, being informed and having access to context. Considering STM publishing, it is noteworthy that the access to content and the providing of a membership is of major importance. Specific departments at the publishing companies focused on marketing, information technologies, readers' development and they take care of it. As discoverability is among the key issues of the publishing industry, the methods, the ways and the competition of Google search engine visibility will be of value for

future researches focusing probably on local, small publishing markets as well as on investigating specific issues.

References

- Ahrefs (n.d.), "SEO Tools & Resources To Grow Your Search Traffic", available at: <https://ahrefs.com/> (accessed 15 January 2020).
- Banou, Ch. (2017), *Re-inventing the Book. Challenges from the Past for the Publishing Industry*, Elsevier - Chandos Publishing, Cambridge - Oxford.
- Beckwith, K. (2003), "Googled: the quest for visibility on the Internet", *Learned Publishing*, Vol. 16 No. 4, pp. 277-283.
- Case, D.O. and Given, L.M. (2016), *Looking for information: A survey of research on information seeking, needs and behavior*, 4th ed., Emerald Group Publishing Limited, Bingley.
- Chi, Y.Y. (2014) "The e-volution of publishing: challenges and opportunities in the digital age", *Publishing Research Quarterly*, Vol. 30 No. 4, pp. 344-351.
- Clark, G. and Phillips, A. (2014), *Inside Book Publishing*, 4th ed., Routledge, London and New York, NY.
- Dickinson, Z. and Smit, M. (2015), "Being where the people are: the challenges and benefits of search engine visibility for public libraries", *Library Hi Tech News*, Vol. 32 No. 10, pp. 11-15.
- Dickinson, Z. and Smit, M. (2016), "Canadian public libraries and search engines: barriers to visibility", *Aslib Journal of Information Management*, Vol. 68 No. 5, pp. 589-606.
- Dictionary of Publishing and Printing (1926), 3rd ed., A & C Black Publishers Ltd, London, p. 169.**
- Frederiksen, L. (2015), "Exploring discovery", *Public Services Quarterly*, Vol. 11 No. 1, pp. 43-47.
- French, R.B. and Fagan, J.C. (2019), "The Visibility of Authority Records, Researcher Identifiers, Academic Social Networking Profiles, and Related Faculty Publications in Search Engine Results", *Journal of Web Librarianship*, Vol. 13 No. 2, pp. 156-197.
- google (n.d.), "PyPI", available at: <https://pypi.org/project/google/> (accessed 15 January 2020).
- Greco, A.N., Rodriguez, C. and Wharton, R. (2007), *The Culture and Commerce of Publishing in the 21st century*, Stanford University Press, Stanford.
- Greco, A.N. (2013), *The Book Publishing Industry*, 2nd ed., Routledge, London and New

York, NY.

Griffiths, J.R. and Brophy, P. (2005), "Student searching behavior and the web: use of academic resources and Google", *Library Trends*, Vol. 53 No. 4, pp. 539-54.

Guédon, J.C., Kramer, B., Laakso, M., Schmidt, B., Šimukovič, E., Hansen, J., ... and Patterson, M. (2019), *Future of scholarly publishing and scholarly communication*, Report of the Expert Group to the European Commission, Luxembourg, Publications Office of the European Union.

Hariri, N. (2011), "Relevance ranking on Google: are top ranked results really considered more relevant by the users?", *Online Information Review*, Vol. 35 No. 4, pp. 598-610.

Hemminger, B.M., Lu, D., Vaughan, K.T.L. and Adams, S.J. (2007), "Information-seeking behavior of academic scientists", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 14, pp. 2205-2225.

Hunter, K., Virkler, S. and Sidi, R. (2007), "Disruptive technologies: taking STM publishing into the next era", *Serials*, Vol. 20 No. 1, pp. 50-53.

hyperion-gray / googlespider (n.d.), "GitLab", available at: <https://gitlab.com/hyperion-gray/googlespider> (accessed 15 January 2020).

Jamali, H.R. and Asadi, S. (2010), "Google and the scholar: the role of Google in scientists' information-seeking behaviour", *Online Information Review*, Vol. 34 No. 2, pp. 282-294.

Johnson, R., Watkinson, A. and Mabe, M. (2018), *The STM Report: An overview of scientific and scholarly publishing*, International Association of Scientific, Technical and Medical Publishers, Hague.

Kesselman, M. and Watstein, S.B. (2005), "Google Scholar™ and libraries: point/counterpoint", *Reference Services Review*, Vol. 33 No. 4, pp. 380-387.

Killoran, J.B. (2013), "How to Use Search Engine Optimization Techniques to Increase Website Visibility", *IEEE Transactions on Professional Communication*, Vol. 56 No. 1, pp. 50-66.

Lapham, L.H. (1997), "Introduction to the MIT Press Edition. The Eternal Now", in McLuhan, M., *Understanding Media: the Extensions of Man*, MIT Press, Cambridge-MA and London, pp. ix-xxiii.

Lee, S., Jang, W., Lee, E. and Oh, S.G. (2016), "Search engine optimization: A case study using the bibliographies of LG Science Land in Korea", *Library Hi Tech*, Vol. 34 No. 2, pp. 197-206.

Lloyd, S. (2008), "A Book Publisher's Manifesto for the Twenty-first Century: How Traditional Publishers Can Position Themselves in the Changing Media Flows of a

- Networked Era", *Library Trends*, Vol. 57 No. 1, pp 30-42.
- Luh, C.J., Yang, S.A. and Huang, T.L.D. (2016), "Estimating Google's search engine ranking function from a search engine optimization perspective", *Online Information Review*, Vol. 40 No. 2, pp. 239-255.
- Marres, N. and Weltevrede, E. (2015), "Scraping the social? Issues in real-time research", *Journal of Cultural Economy*, Vol. 6 No. 3, pp. 313-335.
- Mayr, P. and Walter, A.K. (2007), "An exploratory study of Google Scholar", *Online Information Review*, Vol. 31 No. 6, pp. 814-830.
- Michaels, K. (2015), "The evolving challenges and opportunities in Global publishing", *Publishing Research Quarterly*, Vol. 31 No. 1, pp. 1-8.
- Markland, M. (2006), "Institutional repositories in the UK: what can the Google user find there?", *Journal of Librarianship and Information Science*, Vol. 38 No. 4, pp. 221-228.
- McLuhan, M. (1997), *Understanding Media: the Extensions of Man*, MIT Press, Cambridge-MA and London.
- Miklosik, A., Evans, N., Zak, S. and Lipianska, J. (2019), "A framework for constructing optimisation models to increase the visibility of organizations' information in search engines", *Information Research*, Vol. 24 No. 1, paper 808.
- Miller, L. (2007), *Reluctant Capitalists. Bookselling and the Culture of Consumption*, The University of Chicago Press, Chicago and London.
- Norris, M., Oppenheim, C. and Rowland, F. (2008), "Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR", *Online Information Review*, Vol. 32 No 6, pp. 709-715.
- Okerson, A. (1996), "What academic libraries need in electronic content licenses: Presentation to the STM library relations committee, STM annual general meeting, October 1, 1996 ", *Serials Review*, Vol. 22 No. 4, pp. 65-69.
- Phillips, A. (2014), *Turning the Page. The Evolution of the Book*, Routledge, London and New York.
- Poulos, M., Papavlasopoulos, S., Kostagiolas, P. and Kapidakis, S. (2017), "Prediction of the Popularity from Google Trends Using Stationary Control: The Case of STM Publishers", in *2017 Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, IEEE, pp. 159-163.
- Rzepa, H.S. and Murray-Rust, P. (2001), "A new publishing paradigm: STM articles as part of the semantic web", *Learned Publishing*, Vol. 14 No. 3, pp. 177-182.
- Scrapeulous. (n.d.), "Search Engine Scraping", available at: <https://scrapeulous.com/>

- (accessed 15 January 2020).
- SEMrush (n.d.), "Online Visibility Management Platform", available at: <https://www.SEMrush.com/> (accessed 15 January 2020).
- Somerville, M.M., Schader, B.J. and Sack, J. (2012), *Improving the discoverability of scholarly content in the Twenty-First Century: Collaboration opportunities for librarians, publishers, and vendors, A SAGE White Paper*, SAGE Publications, Los Angeles, CA.
- Somerville, M.M. and Conrad, L.Y. (2013), "Discoverability challenges and collaboration opportunities within the scholarly communications ecosystem: A SAGE White Paper update", *Collaborative Librarianship*, Vol 5 No. 1, paper 4.
- Somerville, M.M. and Conrad, L.Y. (2014), *Collaborative improvements in the discoverability of scholarly content: Accomplishments, aspirations, and opportunities, A SAGE White Paper*, SAGE Publications, Los Angeles, CA.
- Statista (n.d.), "Global market share of search engines 2010-2019", available at: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/> (accessed 15 January 2020).
- Striphas, T. (2009), *The Late Age of Print. Everyday book culture from consumerism to control*, Columbia University Press, New York, NY.
- Strzelecki, A. (2019), "Website removal from search engines due to copyright violation", *Aslib Journal of Information Management*, Vol. 71 No.1, pp. 54-71.
- Strzelecki, A. (2020), "Google Medical Update: Why Is the Search Engine Decreasing Visibility of Health and Medical Information Websites?" *International Journal of Environmental Research and Public Health*, Vol. 17 No. 4, paper 1160.
- Teixeira da Silva, J.A. and Dobránszki, J. (2016), "How Authorship is Defined by Multiple Publishing Organizations and STM Publishers", *Accountability in Research*, Vol. 23 No. 2, pp. 97-122.
- Thompson, J.B. (2010), *Merchants of Culture. The Publishing Business in the Twenty-First Century*, Polity Press, Cambridge.
- Umenhofer, L. (2019), "Gaining Ground: Search Engine Optimization and Its Implementation on an Indie Book Press", *Publishing Research Quarterly*, Vol. 35 No. 2, pp. 258-273.
- Ware, M. (2006), *Scientific publishing in transition: An overview of current developments*, Mark Ware Consulting, Bristol.
- Ware, M. and Mabe, M. (2015), *The STM report: An overview of scientific and scholarly journal publishing*, International Association of Scientific, Technical and Medical

Publishers, Hague.

Wu, M.D. and Chen, S.C. (2014), "Graduate students appreciate Google Scholar, but still find use for libraries", *The Electronic Library*, Vol. 32 No. 3, pp. 375-389.

Zhang, J. and Dimitroff, A. (2005), "The impact of webpage content characteristics on webpage visibility in search engine results (Part I)", *Information Processing & Management*, Vol. 41 No. 3, pp. 665-690.

Zhu, J. and Kelley, J. (2015), "Collaborating to reduce content gaps in discovery: What publishers, discovery service providers, and libraries can do to close the gaps", *Science & Technology Libraries*, Vol. 34 No. 4, pp. 315-328.

Table 1. Results of estimated traffic for STM publishers on keywords level

	Ahrefs			SEMrush		
	Keywords	est. Traffic	Norm. value	Keywords	est. traffic	Norm. value
Elsevier	10.000	1.174.836	100	30.000	1.059.501	100
Wiley	10.000	903.618	76,9	30.000	843.370	79,6
Springer	10.000	576.713	49,0	30.000	644.278	60,8
Taylor	10.000	411.038	34,9	30.000	429.988	40,5
Emerald	10.000	117.192	9,9	30.000	83.009	7,8

Figure 1. Normalized value of keywords traffic estimates for all STM publishers

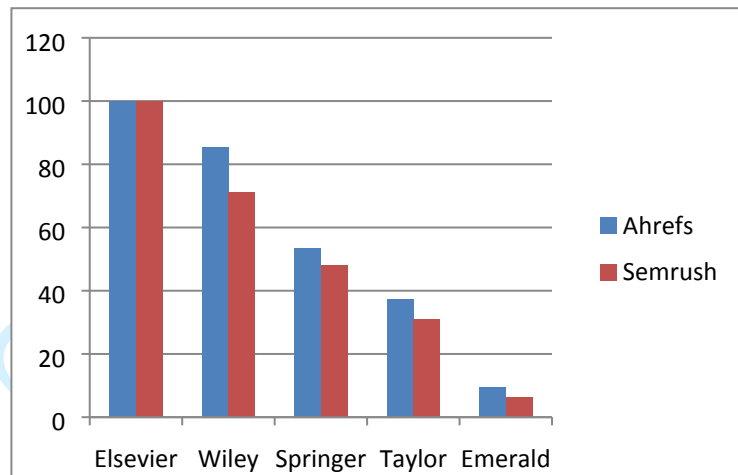


Table 2. Results of estimated traffic for STM publishers on URL level

	Ahrefs			SEMrush		
	Pages	Est. traffic	Norm. value	Pages	Est. traffic	Norm. value
Elsevier	10.000	1.294.834	100	5.000	1.070.953	100
Wiley	10.000	1.105.911	85,4	5.000	760.964	71,0
Springer	10.000	693.343	53,5	5.000	517.072	48,2
Taylor	10.000	481.993	37,2	5.000	331.683	31,0
Emerald	10.000	124.862	9,6	5.000	66.626	6,2

[illegible]