SSARSC

Scientific Society of Advanced Research and Social Change
SSARSC International Journal of Information and Communication Technology
Volume 1 Issue 1, January-June 2018, ISSN 2581 - 5873

Alternative Search Engines

A timeline analysis based on webometric indicators

Bernd Markscheffel
Institut für Wirtschaftsinformatik (Business Information Systems Engineering)
Department of Economic Sciences and Media, Technische Universität Ilmenau
Ilmenau, Germany
bernd.markscheffel@tu-ilmenau.de

Abstract— **In this paper we analyze the market of Alternative Search Engines. Basis of this approach was the COLLNET 2011 paper which describes the use of Webometric indicators to classify and to rank Alternative Search Engines [1]. We combine web usage indicators and reputation based indicators to get an objective and comprehensive picture of the search engines in their special segments. The survey was conducted two times in 2015 and 2016, so we were able to derive statements about the development within the separate categories and about the dynamic character of a special segment.**

Keywords—Webometrics; Alternative Search Engines (ASE)

I. INTRODUCTION

Search engines such as Google, and Bing are prominent tools to search for information about users concerns on the Internet. The search engine market is dominated by two Big Players. Google and Bing as universal search engines with a market share of 86 % - Google in USA and 93% Google in Europe [2] and a couple of domestic search engines in selected countries South Korea: Naver 77% [3], [4] Russia: Yandex 60% [5], China: Baidu 81% [2], Czech Republic: Seznam 37% [6] have reached a respectable market share in their special regions. A chance for competitors are Vertical- or Alternative Search Engines (ASE) which are providing options to search for special document types, specific topics or time-sensitive information [7], [8]. In this paper the dynamic development of the ASE market is investigated because of the large number of ASE and their rapidly changing range. As a basis, we use our study from 2011 [1] where a ranking of ASEs within selected categories was created. The main indicators for the ranking are derived from a webometric analysis where we combined web link structure indicators, web usage and web technology indicators [1], [9], [10]. The investigation was repeated two times (2015 and 2016) and it is planned to continue this yearly cycle to discover not only the best ASE in its category. Moreover, it will help do figure out the big picture of the dynamic of an Alternative Search Engine category. So, we will be able to answer questions like:

- Are there competitors in the market beside the big internet companies?
- Who are the main players in its specific segment?
- Which category has the most dynamic character?

Basis of this approach was the COLLNET 2011 paper which describes the use of Webometric indicators to classify and to rank Alternative Search Engines [1]. As search engines have become an essential tool for searching for information on the web many alternative search services are specialized in finding topic- or media-specific search results. By creating a ranking of these ASE within selected categories we are able present an overview of the ASE which are currently available. With the help of webometric indicators the ASE were compared and the most popular ASE of the respective categories were determined.

II. METHODOLOGY

A. *Define the Categories*

We have adopted the classification approach of 2011 for the categorization of the ASE [1]. It has given good results and it reflects the market of vertical search engines sufficiently. So, we will conduct the comparison of the ASE world in the following categories:

- image search engines,
- video search engines,
- audio search engines,
- question & answer services,
- social bookmarking services
- blog search engines,
- people search engines,
- science search engines.

B. *Determine the universal set*

In a second step, we had to update the universal set of ASE. The main criteria for a search engine to put in our classification are unmodified. A search engine of our universal set has to

- be available and functional,
- fit in one of the selected categories,
- use methods of his own to utilize their own or an external search index,
- be without a restriction regarding topic or country (except the restrictions given by the categories) and
- offer its service without a registration or charges for the user (except science search engines) [1].

We had to check and update the basic numbers of ASE. We did in in the same way like in our first study. We have checked several search engine lists (e.g. 20search.com, Phil Bradley's SE-list…) [11] and added and completed it with the results of Google queries and types of related:URL queries. We had to clean the data, eliminate spelling errors, remove doublets and finally we got our updated universal set for each category of ASE.

C. *Select meaningful indicators*

In our first approach [1] we have used the WIF [12]_and Google PageRank [13] as reputation based indicators and Alexa Traffic Rank [14] as a web usage indicator.

So, the second main task was the analysis and the selection of meaningful indicators. It was necessary to reflect the current bibliometrics and webometrics literature, collect indicator candidates and finally we had to discuss and select the new indicators for the comparison. We have segmented the indicator set – according to the benchmark set of our first study – in two semantic categories:

A) Web usage based indicators and in

B) Web structure or reputation based indicators.

A) Web usage indicators

We have discussed several solutions for a powerful indicator set. Finally, we used Alexa (www.alexa.com) and Similar Web (www.similarweb.com - an Israel based information technology company with its key competence in market intelligence, web analytics, data mining and business intelligence) as source for a more profound indicator set. Every source offers a number of measures. We have used the most comprehensive indicators which are available in both in Alexa and Similar Web. So we combined the web usage analysis indicator set as follows:

- Rank is a number which is calculated for each web site and is proportional to its web usage value;
- Bounce Rate is the percentage of visitors who enter the site and then leave ("bounce") rather than continuing on to view other pages within the same site;
- PageViews/Visitor is the total number of pageviews divided by the total number of unique visitors for the same period and
- Average Time is the average time a user spent on the site [15].

For each indicator we use the indicator value to calculate a rank within the ASE category. Then we calculated the average rank for each indicator from both sources, and finally we calculate an overall average rank value which represents the Web Usage Rank of the special ASE.

TABLE II. WEB USAGE INDICATORS SAMPLE DATA CUT OUT FOR AUDIO SEARCH ENGINES AND ALEXA AS SOURCE

Website	Source: Alexa							
	Rank		*Bounce Rate*		*PageViews/ Visitor*		*Average Time*	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
Air MP3	282.094	17	64,30%	23	2,50	15	01:24	22
ArtistServer	445.183	19	46,40%	17	3,20	11	03:04	10
BeeMP3	9.489	8	27,90%	4	5,54	2	03:33	7
FindSounds	97.572	13	33,60%	6	1,98	19	01:43	19
Grooveshark	942	4	38,30%	10	1,05	22	05:18	3
HulkShare	3.962	6	40,60%	11	2,94	14	04:03	6
Last.fm	13.085	9	45,00%	15	3,96	6	04:13	5
LivePlasma	919.349	22	50,00%	20	1,00	23	02:09	16
…	…	…	…					

B) Web structure indicators

In according to our previous study we use the

- Page Rank and a special derivate of the
- Web Impact Factor (WIF).

With the help of MOZ' tool Open Site Explorer (MOZ is an inbound marketing company based in US and well known for its SEO tools and MOZ Analytics - www.moz.org) we calculated these additional structure centered indicators:

- Root Domains as the number of other sites that link to your page/site [16] and
- Domain Authority, which is a score (on a 100-point scale) developed by MOZ, that predicts how well a website will rank on search engines. One can use Domain Authority when comparing one site to another or tracking the "strength" of a website over time. MOZ calculated this metric by combining all of the other link metrics—linking root domains, number of total links, MozRank, MozTrust, etc.—into a single score [16].

As already described above we calculate for each indicator a rank within the ASE category and finally we calculate an overall average rank value which represents the Web structure rank of the special ASE. The average value of the Web usage rank and the Web structure rank gives the finally ranking order of the ASEs within their category.

III. DISCUSSION AND RESULTS

The findings of our research conducted in the second survey are represented by the rankings shown in Table 2.

Table 3 shows the results of the study which were carried out one year later, with exactly the same methodology and exactly the same indicators.

TABLE II. RESULTS FOR THE SEVERAL ASE CATEGORIES IN 2015

	Image	Video	Audio	Social	Q&A	Blog	Science	People
1	Google Image Search	YouTube	SoundCloud	reddit	Ask.com	Google Blogsearch	Google Scholar	Facebook
2	Flickr	Vimeo	SoundClick	Pinterest	Yahoo! Answers	Blog Catalog	Microsoft Academic Search	LinkedIn
3	Yahoo! Image Search	Dailymotion	Grooveshark	StumbleUpon	Answers.com	Blogarama	BASE	Yahoo
4	deviantART	hulu	Pandora	Tumblr	Quora	IceRocket	allacademic	Spokeo
5	Getty Images	myVideo	BeeMP3	fark	ChaCha	BOTW	CiteULike	Intelius
6	Bing Images	Break	Midomi	Digg	AllExperts	Technorati	Science.gov	Pipl
7	SmugMug	PBS	Last.fm	Delicious	Mahalo	Bloggernity	INFOMINE	ZabaSearch
8	morgueFile	liveleak	mp3Skull	Technorati	WikiAnswers	Bloglinks	CiteSeerX	Wink
9	Photobucket	Blip	MixCloud	Newsvine	AnswerBag	Bloglines	refseek	Yoname
10	Corbis Images	Metacafe	FindSounds	Slashdot	FunAdvice	Topix	WorldWide Science	PeekYou

TABLE III. RESULTS FOR THE SEVERAL ASE CATEGORIES IN 2016

	Image	Video	Audio	Social	Q&A	Blog	Science	People
1	Flickr	YouTube	Last.fm	reddit	Ask.com	Blog Catalog	Google Scholar	Facebook
2	Yahoo! Image Search	Hulu	SoundCloud	Twitter	Answers.com	Topix	OAIster	VK
3	deviantART	Bing Video	freesound	Pinterest	Yahoo! Answers	blog-collection	JSTOR	Odnoklassniki
4	shutterstock	Dailymotion	SoundClick	fark	ask.fm	Blogarama	WOS	LinkedIn
5	Fotolia	vimeo	Freemusic-archive	farkFolkd	zhihu	IceRocket	WorldCat	Yahoo
6	Bing Images	Google Video	Midomi	Tumblr	Quora	ask	Microsoft Academic Search	whitepages
7	pixabayg	tudou	jamendo	Stumbleupon	WikiAnswers	Blogflux	EbscoHost	Spokeo
8	SmugMu	Break	Freeplay-music	Digg	zhidao	bloglog	Mendeley	Zabasearch
9	Getty Images	liveleak	soundjay	Colvia	gutefrage	plazoo	SpringerLink	Pipl
10	Photobucket	MySpace Video	shazam	Delicious	askielly	alltop	BASE	beenverifide

TABLE IV. DETAILED VIEW FOR VIDEO SEARCH ENGINES DEVELOPMENT

Rank	2010	Value
1	YouTube	1,7
2	MySpace Video	5,3
3	Google Video	5,3
4	Dailymotion	6,7
5	AOL Video	6,7
6	Vimeo	9,0
7	Veoh	9,3
8	Truveo Video Search	10,0
9	Metacafe	10,3
10	Megavideo	11,7

Rank	2015	Value
1	YouTube	2,4
2	Vimeo	5,9
3	Dailymotion	6,4
4	hulu	7,1
5	myVideo	8,1
6	Break	8,2
7	PBS	8,3
8	liveleak	9,6
9	Blip	9,6
10	Metacafe	9,7

Rank	2016	Value
1	YouTube	1,6
2	hulu	8,1
3	Bing Video	9,4
4	Dailymotion	10,6
5	Vimeo	10,9
6	Google Video	11
7	tudou	11,7
8	Break	12,2
9	liveleak	13,3
10	MySpace Video	15,3

23

As discussed in the basis paper [1] the interpretation of our ranking is limited by the calculation of the total values and by potential inaccuracies of the indicators. The total values were calculated by summing up the equally weighted ordinal scaled values of the indicators. The distances and quotient from ordinal scaled values cannot be interpreted. Hence, the total values can provide only the ranking order within the categories. That's why we present only the ranking within the individual ASE categories and draw conclusions only when the interpretation is unambiguous.

We do not wish to comment the ranking order for every individual category in detail (e.g. Table IV illustrates as an example the closer view on the dynamic of the video search engine category), but we can determine some interesting facts concerning the big picture of the development in the ASE timeline from 2010 – 2016..

It is interesting to see that the big internet companies like:

- Yahoo! with Image Search, Flickr, Answers, People Search;
- Microsoft with Bing Images and Academic Search and first and foremost
- Google with Image Search, YouTube, Blogsearch, and Google Scholar

are able to transfer their experience in the universal or horizontal search area to the vertical search engine domain and were able to play here a dominant role.

E.g. Google is dominating the market of video search engines with YouTube since our first investigation and receives not only the best value of all examined ASE but has also the largest gap to the next competitor in 2016.

Due to the multiple year study we are able to make statements about the dynamic character for a special search engine category. Table 5 and 6 show the big picture of the surveys. Blue are the new candidates, green have improved and pink have declined its position. When we use the freshness ratio (FR - percentage of new ASE in the top ten) as an indicator for the dynamic of an ASE category or in other words as a statement for the chance to establish a company in a niche segment of Search Engines we can point out following results.

- The overall FR-value is nearly constant with 41% in 2015 and 43% in 2016, that means not only in the 5 year gap but also from 2015 to 2016 almost half of the of the ASE are new in the top ten which is a clear indication that ASE is a highly dynamic market segment.
- The top positions in 2016 are established candidates. It will be interesting to see, if they can defeat its position in the following surveys.

If we have a closer lock to the several categories we will mention the following insights.

- Audio Search engines have the highest average FR (90% in 2015, also the yearly change with 60 % in 2016 is high). This is the category with the highest dynamic structure. So the chance for competitors to establish a convenient market position in this segment is high. But on the other hand the FR declines over time, which can be interpreted as a consolidation tendency.
- The market for Image Search engines seems the most stable segment with an average FR=30%.
- Blog Search engines and Science Search engines have the largest FR growth, that can be interpreted as a signal for an increasingly change in this special segments.

TABLE V. BIG PICTURE OF THE 2015 SURVEY

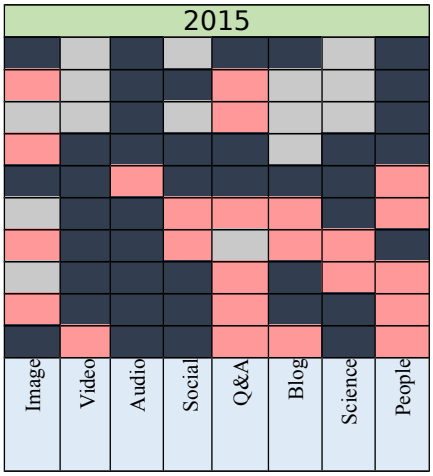
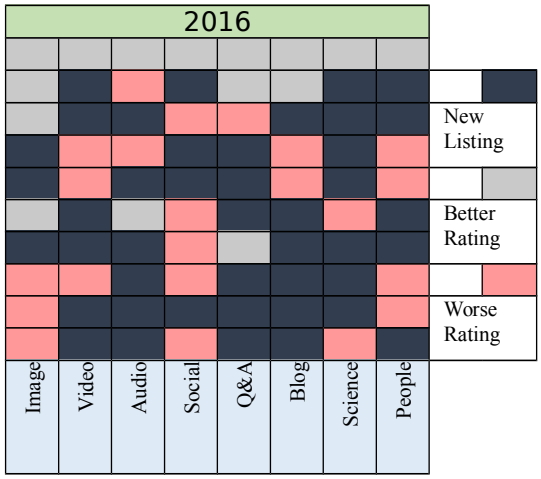


TABLE III. BIG PICTURE OF THE 2016 SURVEY



IV. SUMMARY AND FUTURE WORK

With our ASE surveys we are able to discuss the development in the sector of vertical search engines. The use of webometric indicators, the combination of web usage indicators and reputation based indicators and especially the use of the new indicator set, based on SEO reflexions, helps us

to get an improved and more objective picture of the different segments. The timeline analysis allows us to create a big picture of every segment and to analyze the potential of the dynamic structure of a specific category. Later surveys will encourage the timeline-oriented interpretation of Alternative Search Engines.

V. ACKNOWLEDGMENTS

The data collection was conducted as part of the project "ASE: Five years later" [17] from the master classes in Information Retrieval. I want this solution to be understood as one possible example how to sensitize an IR-class for Informetrics issues.

References

[1] B. Markscheffel and B. Eine, "The top 10 alternative search engines (ASE) within selected categories ranked by webometric indicators," Proceedings of the 7. International Conference on Webometrics, Informetrics and Scientometrics & 12 Collnet Meeting. September 20 – 23, 2011, Istanbul, Turkey, pp. 362 – 372.

[2] Statista, "Google rules the west, Baidu top in China," Retrieved February 12, 2017 from: https://infographic.statista.com/normal/chartoftheday_5472_google_rules_the_west_baidu_top_in_china_n.jpg

[3] C. Duerden, "Advertising on naver: How is it different to Google?," Retrieved August 17, 2017 from: http://www.geniegoals.co.uk/blog/advertising-on-naver

[4] E. S. Ha, "Let's Google something? Naver in Korea," Retrieved February 12, 2017 from: https://korcan50years.com/2016/06/26/overshadowing-of-google-in-korea-portal-website-naver/

[5] B. Mcgonigle, "Search engine market share in russia," Retrieved February 12, 2017 from: http://russiansearchmarketing.com/search-engine-in-russia/

[6] J. Vidim, "The share of search engines Google and Seznam on the Czech internet," Retrieved February 12, 2017 from: http://www.evisions.cz/blog-2016-06-20-the-share-of-search-engines-google-and-seznam-on-the-czech-internet

[7] J. Gelernter, "At the limits of Google: Specialized search engines," Retrieved August 6, 2016 from: http://www.allbusiness.com/technology/software-services-applications-search-engines/10603593-1.html.

[8] D. Lewandowski, "Spezialsuchmaschinen," In D. Lewandowski (Ed.), Handbuch Internet-Suchmaschinen (pp. 53-69). Heidelberg: Aka Verlag.

[9] L. Björneborn and P. Ingwersen, "Toward a basic framework for webometrics," Journal of the American Society for Information Science and Technology, 55 (14), 2005, pp. 1216-1227.

[10] Thelwall, M., Vaughan, L. & Björneborn, L.. Webometrics. Annual Review of Information Science and Technology, 39, 2005, pp. 81-135.

[11] P. Bradley, "Which search engine when?" Retrieved February 12, 2017 from: http://www.philb.com/whichengine.htm

[12] .P. Ingwersen, "The calculation of web impact factors", Journal of Documentation, Vol. 54 Issue: 2, 1998, pp.236-243,

[13] .L. Page, S. Brin, R. Motwani, T. Winograd, "The PageRank citation ranking: Bringing order to the web," Retrieved September 10, 2010 from: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf.

[14] Alexa Internet, "About the alexa traffic rankings ". Retrieved August 11, 2016 from: http://www.alexa.com/help/traffic-learn-more.

[15] SimilarWeb, Glossary. Retrieved February 12, 2017 from: https://www.similarweb.com/knowledgebase/glossary/

[16] MOZ (2017). *Domain Authority*. Retrieved February 12, 2017 from: https://moz.com/learn/seo/domain-authority

[17] B. Markscheffel, "Education in Informetrics: An Approach in an Information Retrieval Context - or - Alternative Search Engines: Five Years Later." In Proceedings of the 12th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & 17th COLLNET Meeting in Nancy, France, 12 to 15 December 2016