Web Search Statistics: Identifying and Satisfying User Needs

Kathrin Kohlert (Federal Statistical Office of Germany, Destatis)

kathrin.kohlert@destatis.de

1.   Introduction

For the dissemination of statistics binding quality criteria apply. These are criteria which generally apply for all statistical offices of the European Statistical System, since they are integral part of the European Statistical Code of Practice. Among others, important quality criteria for the dissemination of statistics are timeliness, relevance and accessibility [1]. Accessibility implies finding information as well as understanding it. These criteria are essential if one wants to follow Steve Krug's most important usability principle "Don't make me think" [2].

If a website offers up-to-date, relevant, easy and fast accessible information it can be assumed that users are satisfied. To satisfy the users' needs is a primary aim of all statistical offices. But how do we know, what is relevant to our users? Often people follow their intuition, when trying to develop user-friendly products. Sometimes their intuition is right, sometimes it is not. However, statisticians should base their actions not on intuition, but on reliable data.

2.   Methods

Log file based search analytics are one data source which offers information about user behaviour. The log files contain information about every access to the internal search engine on a certain website. Based on the available data (log files) it is possible to analyze which terms users enter into the internal search engine and on which page they enter them. Thus, these statistics do not tell you who is trying to find information or if users find what they need in general. But they show what search terms users are looking for on a certain page of a website [3].

With the help of excel macros the data from the log files can be analyzed in a very flexible way and can be used to answer different questions, such as:

- What are the search terms (queries) users enter most frequently?
- On which page of the website exactly do they enter a certain term?

- How does the number of search queries for a certain term change in the course of time?
- At what time did a user start his search? (This information can be used to identify high frequent access by bots.)

3. Examples

The web search statistics offer crucial information about the needs of users. They provide an insight into how users tick and help us to understand them better: The web search statistics show which topics are especially relevant to users at the moment and where users need an easier access to information. At Destatis, we use the insights gained not only to improve the results of our internal search engine, but also to improve our website in general and to adapt the content to the users' needs.

3.1. Relevance

In the summer of 2015, when the number of refugees coming to Europe increased, the web search statistics showed that more and more people were looking for information about refugees online – it seems that this topic was especially relevant to them. For example, in August 2015, the term "refugees" itself was the fourth most common search term on the Destatis website, whereas it was ranked 331 in August 2014.

Table 1. Refugees: search terms (complete website)

August 2014		August 2015	
Position	**Search term**	**Position**	**Search term**
12	migration background	4	refugees
20	migration	13	asylum
49	foreigner	14	asylum seeker
63	migrant	16	migration
78	immigration	19	migration background
188	asylum seeker	51	immigration
209	asylum	72	foreigner
331	refugees	75	migrant

By looking at the number of search terms concerning refugees over time, one can see that even more people were looking for data about refugees on the Destatis website in September 2015. In December, there was a strong decline, which is in general typical for this time of the year. In January, interest in data about refugees increased again. Finally, over the past few months, the number of search queries remained roughly between 1,000 and 1,500.
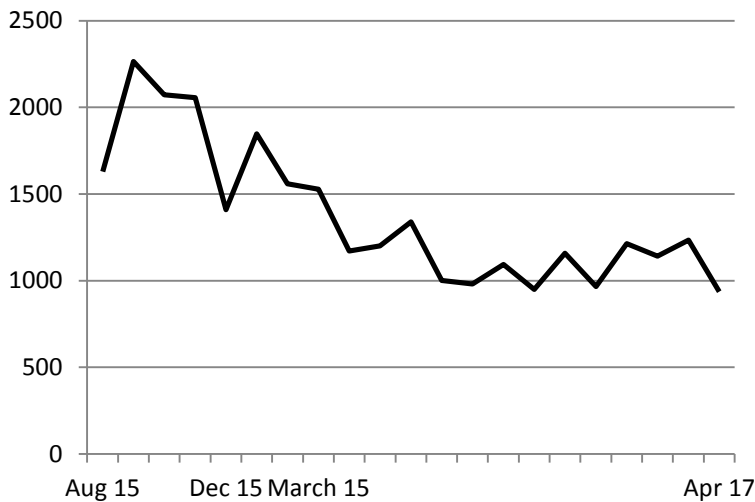
Figure 1. Refugees: search queries over time (complete website)

Because of the high number of people looking for data about refugees on the Destatis website, a special page on refugees was published in December 2015. In March 2016, a number of FAQ were added.

3.2. Accessibility

With the help of the web search statistics one cannot only find out which information is especially relevant to users, but also identify, on which pages the access to certain information has to be improved. The results of a systematic analysis of the web search statistics showed a lot of different starting points for improving the accessibility: Sometimes, users would not find the information they were looking for, because they looked in another area of the Destatis website. In other cases, users looked in exactly the right place but did not find the information, because it was not presented prominently enough or because the language used was only suitable for experts.

As an example table 2 shows the top ten search terms users entered into the internal search engine in 2015 whilst browsing the health section of the Destatis website. Most frequently they searched for "suicide".

Table 2. Health: top ten search terms 2015

Position	Term	Occurrence	Position	Term	Occurrence
1	suicide	456	6	cancer	180
2	dementia	417	7	measles	172
3	alcohol	285	8	smoke	165
4	diabetes	190	9	depression	162
5	stroke	180	10	obesity	153

Most often, users entered the term "suicide" into the internal search engine on the theme page "causes of death". In 2014, two tables with information about suicides were created to satisfy the users' needs in a better way. However, the web search statistics show that some users seem to have problems with finding these tables. To improve the access to this information, the number of suicides was included into a more prominent table with key figures, which users can see at first sight.

4.   Conclusion

To satisfy the users' needs is a primary aim of all statistical offices. Having the web search statistics based on log files at hand, one does not have to follow one's intuition any more when developing user friendly products. The web search statistics offer reliable data that give hints about the best way to improve the content on a website: They show which information is especially relevant to users and indicate where the access to information has to be improved.

But the method also has its limitations. Although it is possible to estimate access via bots – machine access cannot be identified on a reliable basis. Furthermore, the web search statistics only show the search terms users enter, but do not tell us exactly what his or her problem is.

The combination of the web search statistics with data of web analytic tools (for example PIWIK) that offer content tracking and other features would increase the knowledge gain on user behavior.

References

[1] Eurostat (2011), European Statistics Code of Practice, http://ec.europa.eu/eurostat [27 October 2016].

[2] S. Krug (2014), Don't make me think! Web & Mobile Usability: Das intuitive Web, mitp, 3rd edition, p. 11.

[3] J. (G.) Redish (2014), Letting go of the words: writing web content that works, Morgan Kaufmann, 2nd edition, p. 29.