Managerial Economics
2022, Vol. 23, No. 2, pp. 131–152
https://doi.org/10.7494/manage.2022.23.2.131

Joanna Hernik*

The usefulness of big data in creating innovations. The example of Google Trends

1. Introduction

Over the past century, innovation has become an important management issue, sometimes even referred to as the 'religion of the 20th-century entrepreneurs' (Salter, Alexy, 2014, p. 27–28). When the competition is fierce and market challenges more difficult, innovation in particular is considered a success factor for enterprises, entire industries, and even countries (Akbari et al., 2021). It undoubtedly affects the standard of living of both current and future generations. Implementing innovations – understood as the process of developing and introducing something new, innovative, or advanced with the intention of creating value or benefits (Hisrich, Kearney, 2014) – is a multidimensional process (Baregheh et al., 2009). Moreover, innovation is not something that happens by itself; therefore, it should be a systematic activity that may be learned and practiced. Thus, entrepreneurs must deliberately seek sources of innovation and their symptoms in the environment – this boosts the chances for successful innovation (Drucker, Maciariello, 2014; Shah et al., 2015). It is important from several perspectives, such as creating vacancies, surviving on the market, improving people's standard of living, and building economic growth. Therefore, it may be said that it is not capital and labor that create innovation, but that capital and labor result from innovations.

The speed and pace of innovation slow down when the potential combinations of factors are being used up. When this happens, one needs to look beyond the current framework, which nowadays has become easier than ever thanks to the Internet. Currently, the 5G Internet, the Internet of Things and the digitization trends generate more data than ever before, which is why the term 'big

* The West Pomeranian University of Technology in Szczecin, Faculty of Economics, e-mail: jhernik@zut.edu.pl. ORCID ID: 0000-0001-8895-6039.

data' – i.e. huge amounts of data – has emerged. According to a recent study by the International Data Corporation (IDC), the world generated or replicated approximately 64.2 zettabytes of data in 2020, with a projected cumulative annual growth rate of 23% by 2025 (IDC Blog, 2021). Nowadays, personal mobile devices and corporate data centers process more and more information with every passing minute, which is additionally accelerated by remote work (Zhang, 2021). As a consequence, we generate an increasing number of various types of data that can be used by enterprises to develop. A huge amount of data is collected by search engines. According to estimates, the Google search engine, which is dominant on the market, receives approx. 63,000 queries per second, which amounts to 3.5 billion searches per day. Undoubtedly, the data based on queries definitely belongs to the big data category (https://blog.hubspot.com/marketing/google-search-statistics). It should be mentioned that at the end of 2021, Google had 92% of the market, while its nearest competitor, Bing, had less than 3%. It may therefore be concluded that the vast majority of Internet searches are carried out with Google (Google traffic overview, Dec 2021).

The purpose of this article is to prove the suitability of the big data concept for creating and implementing product innovations. The concept of implementing new ideas will be discussed here. The process should start with generating and evaluating these ideas – and this stage is the focus point of this article. The empirical material was data obtained from Google Trends, i.e. data generated by the Google search engine. The author's hypothesis is that the usefulness of the data generated by Google Trends depends on the way the query is entered, which means that the end-user perspective is necessary. The article is structured as follows: presentation of a literature overview on big data and creating innovation; discussion of the research methodology; and a subsequent discussion on the usefulness of the obtained data and conclusions. To the best of the author's knowledge, there have been no published articles to date discussing the differences between the way a query is entered and various periods of analysis.

2. Literature review and conceptual framework

Innovations are the result of invention, i.e. a thought focused on some needs. Innovation itself is rarely an 'absolutely new discovery'; usually it is a creative use of factors and solutions that already existed. It's worth recalling the theorem of Shumpeter (1934), who said that innovation was a 'novel combination' of new or existing knowledge, resources, equipment, and other available factors. Consequently, ideas are created and put into practice (Haberler, 1950; Mazur, Malkowski, 2021).

The nature, sources, and determinants of innovation have occupied researchers for decades. Interestingly, innovation was initially applied to various fields, such as religion, as exemplified by the publication 'Episcopal Innovation; or the test of modern orthodoxy' by Church of England, Diocese of Peterborough in 1820. In 'Gentlemen's Magazine' no. 96 from 1804, one could read about innovations in architecture. In the same period, there were also works related to economic activity, such as 'A Periodical work, Exclusively devoted to Agriculture, and Rural Affairs' from 1805, where readers could learn that the commercial way of charging interest on what is new and uncertain is a common 'financial innovation'. For at least 200 years, innovation has attracted the attention of researchers and practitioners, including businesspeople.

Innovations began to gain popularity in the publications of the 1960s, such as Arnfield's 'Promoting innovation' (1966) or Hayhurst's 'The dynamics of innovation' (1968), which mostly referred to technical changes, but also postulated that innovation should not end with the producer, but be continued by the distributor, and even the consumer, who could take part in product testing. Lack of knowledge about the nature and dynamics of innovation was also seen as a reason for high failure rate of new market entities (Hayhurst, 1968). These issues have remained valid to this day.

Of course, innovations may vary in terms of character. Neirotti and Pesce (2019) wrote that looking from the perspective of the direction of impact, one can distinguish 'inward' process innovations within the company (they improve the effectiveness and efficiency of activities, and therefore relate to technology); there may also be product innovations offered to recipients (they change the company's offer, so they relate to the market). The aforementioned Schumpeter wrote about innovations in the form of a new market, as well as new suppliers, and new forms of organization. A different approach was presented by the Christensen model, which differentiate two types of innovation: incremental innovation and breakthrough innovation (Christensen, 1997; Christensen et al., 2015; Gobble, 2016). Innovations may include functional areas of the enterprise, e.g. marketing. They then concern a product, price, promotion or distribution, and are aimed at meeting the needs of recipients better (Persaud et al., 2021). As may be seen, there are many ways to approach the concept of innovation.

As mentioned before, the creation of innovations is a multidimensional process. Innovation can be generated by employees, engaged customers (especially dissatisfied ones), by benchmarking, and finally – by observing competition. One of the main sources of innovation, as Walder et al. (2006) wrote, are competitors and consumers. The latter may be asked about their expectations and needs using various types of market research. However, it is more difficult to check the innovativeness of competitors before their proposals appear on the market, and

even then, it requires constant tracking of economic information and analyses. On the other hand, enterprises that want to develop have to introduce market innovations. This is due to several reasons – each product sooner or later gets old (as shown in the product life cycle), consumer expectations change, and lastly, new products are introduced by competition (Walder et al., 2006). The need to implement innovations seems indisputable (Mikalef et al., 2019).

The key question to be asked is whether there is any system for acquiring and implementing innovation in the company? Furthermore, if it exists, does it involve external stakeholders – and which ones? No stakeholder group should be underestimated, as each may be a source of new ideas in a dynamic environment. The best product ideas take into account the perspectives of multiple stakeholders and teams, as well as market and customer research (wherever possible). Moreover, a lot depends on management methods and the nature of leadership in the company. Leadership, as Grošelj et al. (2021) wrote, is considered one of the key factors in innovative behaviors at work.

Introducing innovations into the market makes sense when the company has a chance to gain a financial and comparative advantage. On one hand, it requires a search for new ideas, while on the other, an analysis of their profitability. Therefore, every new idea has its source, then it gets developed and evaluated, and finally implemented onto the market. It seems that one of the sources of new ideas may be big data, because modern technologies that collect data have already passed the era of fermentation and development, and nowadays are mature enough to be treated as reliable sources of data and inspiration (Capurro et al., 2021; Sanasi et al., 2021). However, it is worth noting that the current analyzes of the use of big data in business activities are ambiguous and should be approached with appropriate caution, as with any other market research results (Capurro et al., 2021).

This article uses data from Google Trends – a service provided by Google. As already mentioned, the Google search engine is used by 92% of Internet users, which gives the company a central role in gathering market data (Carrière-Swallow, Labbé, 2011; Capurro et al., 2021). Google's infrastructure is just as impressive – it includes hundreds of thousands of servers (estimated to exceed 450,000) spread over thousands of clusters in dozens of data centers around the world (Carr, 2006). When someone types keywords (queries) into Google, the search engine compares them to the index to determine the best matches, and displays links to them along with the relevant cached snippets from web documents. For all of this to work, Google needs to store and analyze a significant proportion of all web content, which is both technically and economically challenging. Yet, on the other hand, it allows the company to collect information known as big data, which is so large and complex that traditional computers are unable to process

it (Carr, 2006). In other words, big data means larger, more complex data sets, especially from new sources. These data sets are so extensive that traditional computing software simply cannot manage them (What is Big Data?). In case of such large collections, large amounts of low-density unstructured data are processed, and the data is generated in milliseconds thanks to networked servers.

Google Trends lists the frequency with which a specific search term is searched in several languages from various regions of the world. To facilitate the comparison of terms, Google normalizes the search data, which means that the search results are normalized to the time and location of the query. The process involves two steps: first, each data point is divided by the sum of the location searches and the time range selected by the user, and secondly, the resulting numbers are scaled from 0 to 100, based on the proportion of the topic to all searches across all topics.

The value of big data becomes clear when we understand that all traditional market data that underpins future decision-making is historical, which should be understood as 'partially out of date' (Carrière-Swallow, Labbé, 2011). In case of data from Google Trends, we know that it is based on the micro-user data; it contains information about a large sample of Internet users (which is a much larger sample than the research agencies could investigate), and is released at high frequency and regular intervals, so it is highly up-to-date (Carrière-Swallow, Labbé, 2011). Big data seems to have become an important form of capital today.

3. Development of hypotheses

Everyone has probably heard about big data, but apart from the high technology industry, few companies use it in their daily activities (Capurro et al., 2021). On the other hand, scientists emphasize the increasing role of big data in better understanding customer needs and in the processes of innovative companies (Nambisan et al., 2017; Mikalef et al., 2019). Therefore, in this article, the author hypothesizes that the usefulness of big data for creating innovation depends largely on the way the query is formulated (here the author refers to the Google search engine and the Google Trends platform). It is crucial to adopt the mindset of the end user – the consumer – and enter queries the way an Internet user would. This corresponds to the vision of creating product or marketing innovations mentioned earlier.

H1: The usefulness of big data for creating innovation depends largely on the way the query is formulated.

Since the area of this study still has many unknowns, the hypothesis adopted herein is non-directional and has an exploratory nature. By using this approach,

the authors wish to show how companies could use big data and what information supporting the innovation process they might obtain. To date, the research of other authors has often focused on companies in digital technology sectors (Liu et al., 2016; Jun et al., 2018), while the authors of this article show the universal possibilities of using big data, regardless of the industry or the level of technological development of a company.

4. Methodology

The aim of this study is to prove that big data analysis in Google Trends may be a useful tool for searching for – and implementing – innovations. The author also wants to prove that the obtained results differ depending on what query is entered and what analysis period is selected. The research method used in this study is content analysis, i.e. the study of the content of messages posted on the Internet and any other written sources and documents.

4.1. Research setting

The study consisted of entering 6 queries (Q1-Q6) into Google Trends (https://trends.google.com) related to the search for new areas for innovation. The user always gets 25 results in so called the TOP category, or fewer if there is insufficient data. The search results are the 25 most frequently entered phrases in the Google search engine, associated with a given query (i.e., if a user enters 'innovations', they will see in the first place 'New Innovations', which is the name of an American company selling medical research software). It is worth noting that these 25 items are based on millions of searches (unfortunately, it is not known exactly how many). Nothing can be said about the group of respondents either as no such data is available. The only clue describing the group of surveyed users is the information as to which Internet users use the Google search engine.

As already mentioned, the study consisted of entering 6 queries – queries Q1 and Q2 were general in nature, and unrelated to any everyday problem. Queries Q3, Q4, Q5 and Q6 related to problems or needs. All inquiries were created solely for this study – the author selected ones associated with introducing changes or searching for innovations.

The key assumption is that when entering search queries into Google Trends, one should adopt the perspective of the end user, the consumer. Therefore, a hypothesis was made (H1) that by entering Q3-Q6 (the end-user perspective), one may get more data on problems and needs, which could be used to create new products or services.

H1: By using the consumer's perspective when formulating a query in Google Trends, one can get more data on problems and needs, therefore it is more useful for creating innovations.

In this study, therefore, a query was entered 6 times for 3 different periods each (5 years, 1 year, 30 days). A total of 18 x 25 results were obtained (with one exception – when there was not enough data for the 30-day period and no results were shown). As a result, 425 variables found in this study were the maximum – there can't be any more since this is the limit to the number that the Google Trends app can display.

The queries entered into Google Trends, which yielded the discussed results (the operation was repeated 3 times for 3 different periods), were as follows:

- Q1 – innovations,
- Q2 – new product,
- Q3 – a problem with,
- Q4 – solve a problem,
- Q5 – how to,
- Q6 – how to easy.

The results for Q1-Q6 were a list of 25 expressions related to the given query. First, the obtained results were analyzed in terms of the content – did the results show a product/service related to the query? Was a specific company name displayed? Was a particular need/problem visible in a specific area? Content that could not be classified as belonging to any of the previously formulated groups was placed under the category 'other' and was eventually excluded from the statistical calculations. This research procedure is a meta-analysis in nature, and is called a 'keyword frequency analysis'. Then, basic descriptive statistics were calculated, and Wilcoxon and Friedman's tests were used, aimed at examining the existence of dependencies and verifying the hypothesis.

So far, companies have commissioned such studies to research agencies, but in case of global companies or universal problems, big data may be used – which is the essence of this article. It is known that the use of big data has its limitations, which will be elaborated on in the Discussion, but the authors wish to show its advantages or opportunities (Capurro et al., 2021).

4.2. Data and data collection

When typing a search term into the Google Trends, e.g. 'innovations', one needs to select a region of interest, the time span analysis, and a category, e.g. 'finance'. One can also leave these parameters as default, i.e. the search term

would be analyzed for the whole world, for the full period (12 years) and for all categories. Google Trends will primarily show graphical data – in Figure 1 it can be observed that the greatest interest in innovations took place in December 2017 but the generated data may also be downloaded in Excel format to analyze it further. This was the approach used in this study, except that the search was performed for periods shorter than 12 years (namely: 5 years, 1 year, and 30 days). Data was collected exactly on December 31, 2021; it is presented in Table 1.

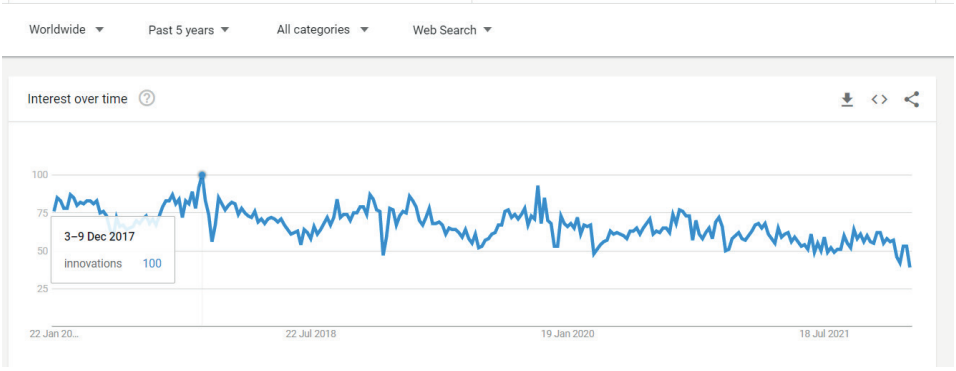

Figure 1. Graphical search results for the term 'innovations' in Google Trends for the period 2017–2021

Source: https://Trends.google.com/Trends/explore?date=today%205-y&q=innovations

Table 1

First collection of data – total results for three measurements

A query	Q1 – innova-tions	Q2 – new product	Q3 – a problem with	Q4 – solve a problem	Q5 – how to	Q6 – how to easy
Product (service) event	18	1	30	9	26	4
Name of a company	25	3	12	3	5	2
Needs in a specific area	11	2	5	7	37	43
Other	54	6	47	19	68	49
Average	27	3	23,5	9,5	34	24,5

As can be seen in Table 1, the first two columns (general) are not related to any problem, while the next four columns (end user) contain phrases which are typical of users looking for a solution to their problem. The authors assumed that the second group of inquiries would yield more results, which quickly was discovered to be untrue. However, on closer inspection, it was found that all the results for the query 'innovation' (i.e. 25 items) simply contained a company name. This type of results may be very useful for tracking competitors' activity; however, it is less useful for finding unmet needs and creating product innovations.

For the purposes of this study, the most important thing was to look for signs of needs, therefore the category 'other' was excluded from further analysis. Moreover, the results of the analysis related to the 'Company name' were commented on. Further analysis encompassed results from Tables 2–4. (It is worth recalling that the maximum number of results displayed by Google Trends for each query is 25; if in Tables 1–4 the sum for a given query is greater than 25, it means that the search result contained more than one category, i.e. for example, both a product and a specific need).

Table 2

The results obtained for the period of 5 years (2017–2021)

A query	Q1 – innovations	Q2 – new product	Q3 – a problem with	Q4 – solve a problem	Q5 – how to	Q6 – how to easy
Product (service) Event	6	0	20	5	6	1
Name of a company	7	1	5	1	1	1
Needs in a specific area	6	0	0	15	15	22

Table 3

The results obtained for the period of 1 year (2021)

A query	Q1 – innovations	Q2 – new product	Q3 – a problem with	Q4 – solve a problem	Q5 – how to	Q6 – how to easy
Product (service) Event	7	1	9	4	9	0
Name of a company	7	2	5	2	3	1
Needs in a specific area	5	1	5	2	3	1

Table 4

The results obtained for the period of 1 month (December 2021)

A query	Q1 – innova-tions	Q2 – new product	Q3 – a problem with	Q4 – solve a problem	Q5 – how to	Q6 – how to easy
Product/service/ event	5	0	1	unavail-able	11	3
Name of a company	11	0	2	unavail-able	1	0
Needs in a specific area	0	1	0	unavail-able	19	20

The main observation to be made at this point is that although the numbers in both tables are relatively small, they are based on huge amounts of searches (possibly even millions). It seems, therefore, worth analyzing them in more depth.

5. Results analysis

In order to determine whether the adopted hypothesis may be confirmed, the authors calculated basic descriptive statistics (mean, standard deviation, median, minimum and maximum), as well as:

1) they checked with the Wilcoxon test whether there were statistically significant differences between the 'end user' and 'general' groups for specific time periods;
2) they used Friedman's test to check whether there were statistically significant differences between the three time periods related to 'end user' and 'general'.

The statistical significance was set at $p < 0.05$. Statistical analysis was performed using the IBM SPSS Statistics 25 suite.

The descriptive statistics are presented graphically in Figures 2, 3 and 4. They illustrate the potential of the data obtained, and allow for the first conclusions to be drawn.

In case of the 5-year period the figure (Fig. 2) is based on the following data:

- for 'end user' – mean M = 7.67, median Me = 8, standard deviation SD = 5.51, minimum Min = 2, and maximum Max = 13;
- for 'general' – mean M = 3.33, median Me = 3, standard deviation SD = 0.58, minimum Min = 3, and maximum Max = 4.

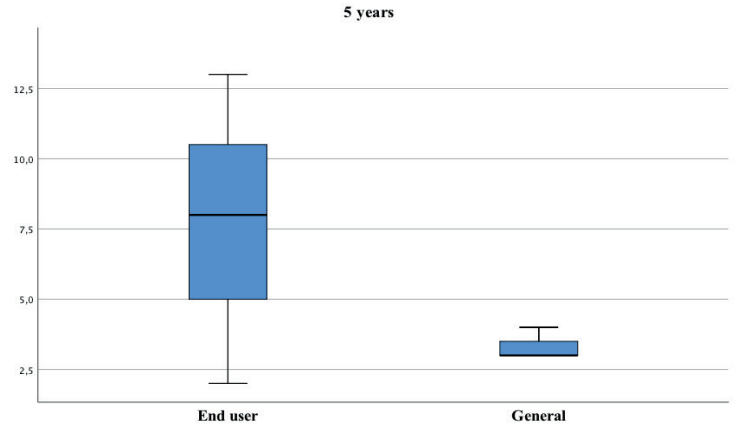

Figure 2. Differences in basic statistical characteristics for 2017–2021

Source: own study

The figure for 1-year period (2021) (Fig. 3) is based on the following data:

- for 'end user' – mean M = 3.67, median Me = 2.75, standard deviation SD = 1.59, minimum Min = 2.75, and maximum Max = 5.5;
- for 'general' – mean M = 3.83, median Me = 4, standard deviation SD = 0.76, minimum Min = 3, and maximum Max = 4.5.

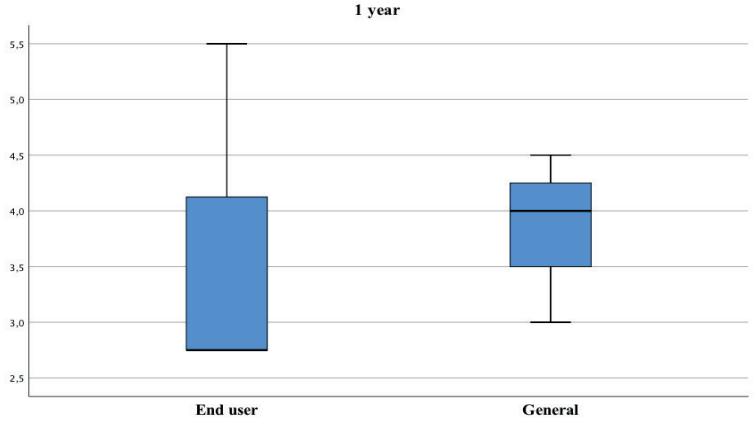

Figure 3. Differences in basic statistical characteristics for a 1-year period

Source: own study

Figure 4 illustrates the data for 1-month period (December 2021). The figure is based on the following data:

- for 'end user' – mean M = 6.25, median Me = 5, standard deviation SD = 6.22, minimum Min = 0.75, and maximum Max = 13;
- for 'general' – mean M = 2.83, median Me = 2.5, standard deviation SD = 2.52, minimum Min = 0.5, and maximum Max = 5.5.

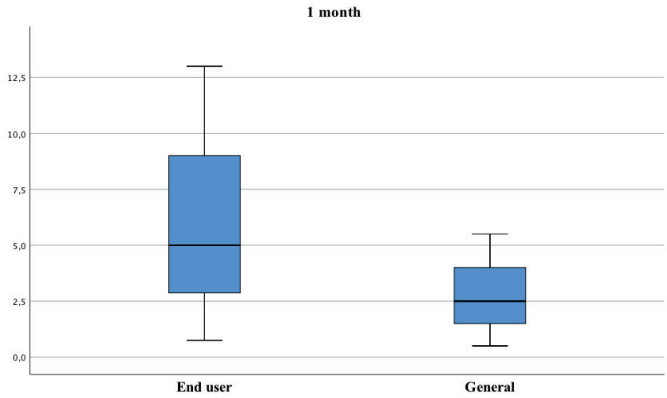

Figure 4. Differences in basic statistical characteristics for a 1-month period

Source: own study

It's known that the median is a better measure for sets with extreme cases, therefore we shall focus on this value. The comparison of all medians is presented in Figure 5.

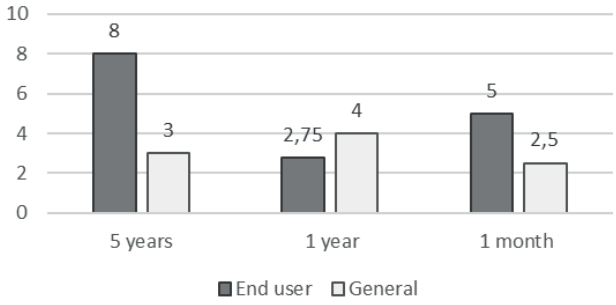

Figure 5. Comparison of the median value obtained for all measurements

Source: own study

Despite the visible differences in the median values, the results of statistical tests (Wilcoxon's test for dependent groups) indicate no statistically significant differences between the individual periods. The test results are as follows: for 5 years – $Z = 1.07$, $p = 0.29$; for 1 year – $Z = 0.54$, $p = 0.59$; for 1 month – $Z = 0.54$, $p = 0.59$.

Using the same data, the question was inverted and it was checked whether there were statistically significant differences between the 3 time periods in the categories 'general' and 'end user' (analysis made with Friedman's test). Figures 6 and 7 illustrate this issue.

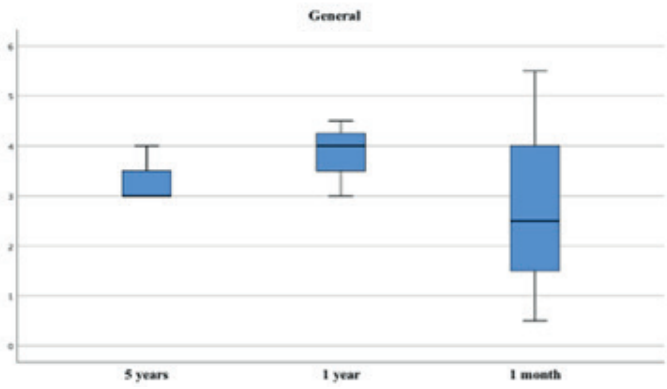

Figure 6. Results obtained for 'General' in particular time spans

Source: own study

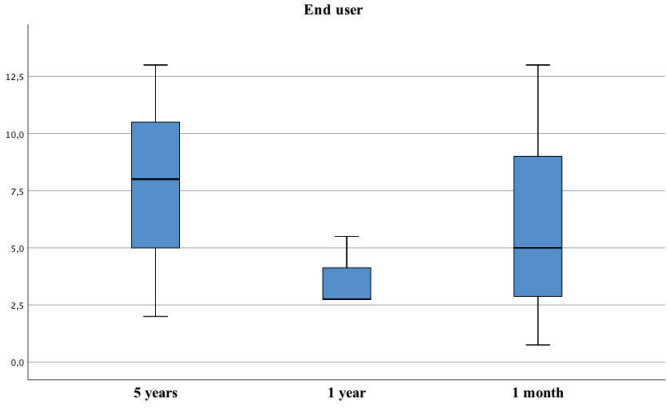

Figure 7. Results obtained for 'End user' in particular time spans

Source: own study

In this study, the authors focused on seeking options that would yield the best results for creating product innovations. By verifying whether there were significant differences between the 3 time spans in the 'General' and 'End user' categories, it may be stated that in case of 'General' queries, the data useful for creating innovations was obtained for a period of 1 month. However, the scope of this data is still worse than for the 'End user' type of queries. For 'General' queries, the average number of useful results in each period fluctuated between 3 and 4, and the highest value – as already mentioned – was obtained for the 1-month analysis. In case of 'End user' queries, the worst average result was obtained for 1 year (3.67), and the best – for 5 years (7.67). For both the 5-year and 1-month period, the maximum amount of data obtained was 13 – compared to 5.5 for 1 year. The median was the highest for the 5-year period. It is worth emphasizing that the best result for the 5-year period for 'End user' was twice as high as the best result in the 'General' category.

Despite significant differences visible in Figures 6 and 7, the results of the analysis for 'general' do not allow for the conclusion that these differences were statistically significant ($\chi^2_F(1) = 1.27$; $p = 0.53$). The same is valid for 'end user' – no statistically significant differences were observed either ($\chi^2_F(2) = 1.64$; $p = 0.44$). This means that, at least from the statistical point of view, the results for both groups and all periods were equivalent.

6. Discussion

In recent years, there has been a debate for and against the use of big data for market decisions. There are studies, such as Carrière-Swallow and Labbé (2011) 'Nowcasting with Google Trends in an emerging market', which confirm that the information from Google Trends allows one to predict trends and phenomena with great accuracy. Similar conclusions were presented by Askitas and Zimmermann (2009), who analyzed the problem of unemployment. On the other hand, there is evidence that predictions and analyzes based on Google Trends fail. For example, Rovetta (2021) emphasized that Google Trends did not provide the right amount of data for the calculations, despite being big data. A different opinion, however, was presented by Medeiros and Pires (2021), who claimed that in order to use Google Trends, one needs to have certain knowledge when entering queries and then analyzing the answers.

Globalization and digitization are currently the main drivers of change (Das et al., 2018) – they can be feared because they change the way market entities operate, but they may also be used to develop the organization. It may be difficult to create breakthrough or radical innovations these days, but introducing changes, as emphasized by Nagano et al. (2016) and Gatignon et al. (2002), is

simply an imperative of our times. Hence, there's a need to look for new sources of innovation.

As already mentioned, innovation is not something that happens by itself, but is a structured or systematic process that requires discipline; it may be learned and practiced. It is also worth being proactive and looking for non-standard sources of innovation – hence the authors' proposal to look at big data and Google Trends. According to the Oracle website (https://www.oracle.com/emea/big-data/what-is-big-data), in relation to big data, the user will have to process large amounts of low-density unstructured data. Its value may be unknown, so only structuring and analysis will help better understand and use it in order to comprehend changes and predict the future (Jun et al., 2018). This is what the authors did by checking the data available in Google Trends.

As Jun et al. (2014) have noted, Google Trends undoubtedly offers an almost instantaneous reflection of the needs, desires, requirements and interests of users; plus, the tool is easy to use and provides various options for comparison. In this study the authors assumed that big data analysis in Google Trends may be a useful tool for searching for – and implementing – innovations. They also wanted to prove that the obtained results differed depending on what query was entered and what analysis period was selected. Unfortunately, the adopted concept of statistical analysis did not prove the hypothesis that the usefulness of big data for creating innovation largely depends on the way the query is formulated. However, it must be emphasized that it was not possible to confirm significant statistical differences for different ways of formulating the query, which means that all the obtained results were equivalent. Perhaps other analyzes should be carried out, ones better suited to the specifics of the study, and possibly more data is needed.

On the other hand, basic statistical measures such as mean, standard deviation, median, minimum and maximum values clearly show differences in the obtained results and encourage the hypothesis to be maintained. In this study, when the method of asking the question was the variable, the obtained results were as follows:

1) for the 5-year period (2017–2021), for the 'general' query, the average was 3.33 of the useful result, and for the 'end user' query, the average was 7.67, which is twice as high; moreover, the median for 'end user' was three times higher;
2) for the 1-year period, the results for both approaches were similar – no significant differences were found (the mean for 'general' was 3.83, and for 'end user' 3.67);
3) for the 1-month period, for the 'general' query, the average was 2.83, while for the 'end user', the average was 6.25, meaning twice as much data was obtained; the median was also twice as high.

Therefore, despite the lack of statistical significance, it may be concluded that if the end-user point of view was adopted and queries over the 5-year period analyzed, one would get more useful data than for other entries (here called 'general') as well as other analysis periods. It should also be noted that in the Q1 query – 'innovations' in the 'general' group, there were company names which included the word 'innovations' in the name. Such data was not excluded from the analysis, but if it had been done, the search values in the 'general' group would have been even smaller.

By focusing not on the type of query entered, but on the comparison of the analysis periods, it may be stated that the queries for 1 year in the 'general' category do have some value, but they are still worse than the results of the 'end user' category. From the perspective of the usefulness of data in the implementation of innovations, the best among the three analyzed periods was the 5-year period, followed by the 1-month period.

Much better results for 'end user' queries allowed the authors to adopt the preliminary research hypothesis that the usability of big data is influenced by how one formulates the query. However, as already mentioned, further statistical analyzes need to be carried out on more data.

7. Conclusions

Thanks to modern technologies, we generate more and more amounts of various types of data that may be used for the development of organizations. A huge amount of data is collected by search engines, which leads to the creation of big data. In this study the authors assumed that big data analysis in Google Trends may be a useful tool for searching for – and implementing – innovations. They also wanted to prove that the obtained results differed depending on what query was entered and what analysis period was selected. Over 400 responses were received in the form of phrases typed by Internet users all over the world, and the results of the conducted analysis allowed for the identification of differences and the selection of specific ways of entering queries and certain periods. The authors believe that the obtained results prove sufficiently that the end-user perspective should be adopted as it gives more useful information relating to needs and problems. Undoubtedly, such information is necessary to successfully implement market innovations.

7.1. Theoretical implications

The basic descriptive statistics highlight differences in favor of the hypothesis but it has not been statistically verified. From the theoretical point of view, other

methods of conducting the analysis should be considered, and perhaps more data should also be collected. The authors notice a potential for further considerations and research on the use of big data. It is also worth adding that hypotheses are made on the basis of theoretical premises, and the theory involving the use of big data in various areas of knowledge has not yet crystallized, which places this work in the domain of exploratory studies.

7.2. Practical implications

Discovering human needs and searching for answers to them is not only the domain of entrepreneurs, therefore this study may have a fairly broad practical applications. By adopting the general assumptions, i.e. ones that do not refer to specific products or industries, the authors showed that the presented path may be recreated by both entrepreneurs and creators of political programs, as well as leaders of non-governmental organizations.

Additionally, it may be stated that by analyzing Google Trends, the authors noticed market opportunities, new needs, and current problems – all this may be a source of innovation, understood as new products, new services, or new approaches to problems. Interestingly enough, one may also obtain insights on the successes or failures of competitors, which is very valuable from a management point of view. So, one can say that an access to big data thanks to Google Trends enables better decision-making.

7.3. Limitations and future studies

It must be remembered that Google Trends is an imperfect tool, mainly because it does not take into account the irrational behaviors of people and does not display many connections. Moreover, the obtained data is highly averaged and it is often difficult to relate it to specific, local solutions. It must also be kept in mind that the smallest change of the query in Goggle Trends changes the obtained results, so it does matter whether we enter 'innovation' or 'innovations'. In addition, two people entering the same query, but in different countries, may obtain different results. This means that the data presented in this article, as well as all the data obtained from Google Trends, is difficult to verify, but it is also a field for further research on the possibilities of using big data.

In conclusion, it may be said that the performed meta-analysis (i.e. the analysis of the frequency of keywords) could be used to examine changes in consumer behaviors and identify new areas for innovation. On the other hand, however, it does not show many connections and dependencies, so it is worth combining keyword analyzes with other data collection methods.

References

[1] Akbari, M., Khodayari, M., Khaleghi, A., Danesh, M. and Padash, H. (2021) 'Technological innovation research in the last six decades: a bibliometric analysis', *European Journal of Innovation Management*, vol. 24, pp. 1806–1831, https://doi.org/10.1108/EJIM-05–2020–0166.

[2] Arnfield, R.V. (1966) 'Promoting innovation – the role of the industrial liaison service', *Education + Training*, vol. 8, pp. 250–252, https://doi.org/10.1108/eb015723.

[3] Askitas, N. and Zimmermann, K.F. (2009) 'Google econometrics and unemployment forecasting', IZA Discussion Papers, No. 4201, Institute for the Study of Labor (IZA), Bonn, [Online], Available: http://nbn-resolving.de/urn:nbn:de:101:1–20090615167 [2 Jan 2022].

[4] Baregheh, A., Rowley, J. and Sambrook, S. (2009) 'Towards a multidisciplinary definition of innovation', *Management Decision*, vol. 47, pp. 1323–1339, https://doi.org/ 10.1108/00251740910984578.

[5] Capurro, R., Fiorentino, R., Garzella, S. and Giudici, A. (2021) 'Big data analytics in innovation processes: which forms of dynamic capabilities should be developed and how to embrace digitization?', *European Journal of Innovation Management*, vol. 25, pp. 273–294, https://doi.org/10.1108/EJIM-05–2021–0256.

[6] Carr, D.F. (2006) 'How Google Works', [Online], Available: https://www.academia.edu/5916601/ Rate_This_Article_How_Google_Works [8 Jan 2022].

[7] Carrière-Swallow, Y. and Labbé, F. (2011) 'Nowcasting with Google trends in an emerging market', *Journal of Forecasting*, vol. 32, pp. 289–298.

[8] Christensen, C.M. (1997) The innovator's dilemma: when new technologies cause great firms to fail, Boston: Harvard Business School Press.

[9] Christensen, C.M., Raynor, M.E. and McDonald, R. (2015) 'What is Disruptive innovation?', *Harvard Business Review*, [Online], Available: https://hbr.org/2015/12/what-is-disruptive-innovation [2 Jan 2022].

[10] Das, P., Verburg, R., Verbraeck, A. and Bonebakker, L. (2018) 'Barriers to innovation within large financial services firms: An in-depth study into disruptive and radical innovation projects at a bank', *European Journal of Innovation Management*, vol. 21, pp. 96–112, https://doi.org/10.1108/EJIM-03–2017–0028.

[11] Drucker, P. and Maciariello, J. (2014) Innovation and Entrepreneurship, London: Routledge, https://doi.org/10.4324/9781315747453.

[12] Gatignon, H., Tushman, M.L., Smith, W. and Anderson, P. (2002) 'A structural approach to assessing innovation: construct development on innovation locus, type, and characteristics', *Management Science*, vol. 48, pp. 1103–1122.

[13] Gobble, M.A.M. (2016) 'Defining Disruptive Innovation', *Research-Technology Management*, vol. 59, pp. 66–71, https://www.doi.org/10.1080/08956308.2016.1185347.
[14] Grošelj, M., Černe, M., Penger, S. and Grah, B. (2021) 'Authentic and transformational leadership and innovative work behaviour: the moderating role of psychological empowerment', *European Journal of Innovation Management*, vol. 24, pp. 677–706, https://doi.org/10.1108/EJIM-10–2019–0294.
[15] Haberler, G. (1950) 'Joseph Alois Schumpeter 1883–1950', *The Quarterly Journal of Economics*, vol. 64, No. 3, pp. 333–372.
[16] Hayhurst, R. (1968) 'The dynamics of innovation', *European Journal of Marketing*, vol. 2, pp. 131–138, https://doi.org/10.1108/EUM0000000005248.
[17] Hisrich, R.D. and Kearney, C. (2014) Managing Innovation and Entrepreneurship, SAGE Publications.
[18] IDC Blog (2021) 'Future of Industry Ecosystems: Shared Data and Insights', 6.01.2021, [Online], Available: https://blogs.idc.com/2021/01/06/future-of-industry-ecosystems-shared-data-and-insights/ [2 Jan 2022].
[19] Jun, S.-P., Park, D.-H. and Yeom, J. (2014) 'The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference', *Technological Forecasting and Social Change*, vol. 86, pp. 237–253.
[20] Jun, S.-P., Yoo, H.S. and Choi, S. (2018) 'Ten years of research change using Google Trends: From the perspective of big data utilizations and applications', Technological Forecasting and Social Change, vol. 130, pp. 69–87, https://doi.org/10.1016/j.techfore.2017.11.009.
[21] Liu, J., Li, J., Li, W. and Wu, J. (2016) 'Rethinking big data: A review on the data quality and usage issues', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 134–142, https://doi.org/10.1016/j.isprsjprs.2015.11.006.
[22] Mazur, R. and Malkowski, A. (2021) 'Significance of employee innovation for knowledge economy based on the case of industrial companies located in West Pomerania', *Zeszyty Naukowe. Organizacja i Zarządzanie*, nr 152, pp. 155–168.
[23] Medeiros, M.C. and Pires, H.F. (2021) 'The Proper Use of Google Trends in Forecasting Models', https://doi.org/10.48550/arXiv.2104.03065.
[24] Mikalef, P., Boura, M., Lekakos, G. and Krogstie, J. (2019) 'Big data analytics and firm performance: Findings from a mixed-method approach', *Journal of Business Research*, vol. 98, pp. 261–276.
[25] Nagano, M.S., Stefanovitz, J.P. and Guimaraes, T. (2016) 'Assessing some important factors to reduce obstacles in product innovation', *International Journal of Innovation Management*, vol. 20, pp. 1–29, https://doi.org/10.1142/S1363919616500225.
[26] Nambisan, S., Lyytinen, K., Majchrzak, A. and Song, M. (2017) 'Digital Innovation Management: reinventing innovation management research in a digital world', *Mis Quarterly*, vol. 41, pp. 223–238.

[27] Neirotti, P. and Pesce, D. (2019) 'ICT-based innovation and its competitive outcome: the role of information intensity', *European Journal of Innovation Management*, vol. 22, pp. 383–404, doi: 10.1108/EJIM-02–2018–0039.
[28] Persaud, A., Wang, S. and Schillo, S.R. (2021) 'Assessing industry differences in marketing innovation using multi-level modeling', *Journal of Business & Industrial Marketing*, vol. 36, pp. 1371–1388, https://doi.org/10.1108/JBIM-12–2019–0532.
[29] Rovetta, A. (2021) 'Reliability of Google Trends: Analysis of the Limits and Potential of Web Infoveillance During COVID-19 Pandemic and for Future Research', *Frontiers in Research Metrics and Analytics*, https://doi.org/10.3389/frma.2021.670226.
[30] Salter, A. and Alexy, O. (2014) 'The nature of innovation', in Dodgson, M., Gann, D.M. and Phillips, N. (ed.) The Oxford Handbook of Innovation Management, OUP Oxford.
[31] Sanasi, S., Trabucchi, D., Pellizzoni, E. and Buganza, T. (2021) 'The evolution of meanings: an empirical analysis of the social media industry', *European Journal of Innovation Management*, vol. 25, pp. 97–121, https://doi.org/10.1108/EJIM-09–2020–0363.
[32] Schumpeter, J.A. (1934) The Theory of Economic Development, Cambridge, MA: Harvard University Press.
[33] Shah, R., Gao, Z. and Mittal, H. (2015) 'Innovation, Entrepreneurship, and the Economy in the US, China, and India', Amsterdam: Elsevier, https://doi.org/10.1016/C2014–0-01381-0.
[34] Walder, B., Weiermair, K. and Pérez, A.S. (2006) Innovation and Product Development in Tourism: Creating Sustainable Competitive Advantage, Erich Schmidt Verlag.
[35] 'What is Big Data?' [Online], Available: https://www.oracle.com/big-data/what-is-big-data/ [9 Jan 2022].
[36] Wilson, J. (1820) 'Episcopal Innovation: or, the test of modern orthodoxy, in eighty-seven questions, imposed as articles of faith upon candidates for licences and for holy orders in the Diocese of Peterborough; with a distinct answer to each question, and general reflections. Relative to their illegal structure and pernicious tendency', Church of England, Diocese of Peterborough, London: L.B. Seeley.
[37] Zhang, K. (2021) 'High-performance computing has become crucial to competitive advantage – in every industry', *Fortune*, 15 December, [Online] Available: https://fortune.com/2021/12/15/high-performance-computing-ai-5g-data-analytics-tsmc/ [9 Jan 2022].

Summary

A huge amount of data is collected by search engines. According to estimates, the Google search engine, which is dominant on the market, receives billions of search requests daily. Of particular note is that a large part of the collected data is available through the Google Trends service. As a consequence, various types of data can be used by enterprises for their development but they often do not take advantage of this opportunity. Therefore, the purpose of this article is to prove the suitability of the big data concept for creating and implementing product innovations, using the example of Google Trends. Discovering human needs and searching for answers to them is not only the domain of entrepreneurs, therefore this study may have a fairly broad practical applications. By adopting general assumptions, i.e. ones that do not refer to specific products or industries, the author has shown that the presented path may be recreated by both entrepreneurs and creators of political programs, as well as leaders of non-governmental organizations who need to implement innovations. The results revealed the selection of specific ways of entering queries in Google Trends and certain periods of analysis which are the most useful for creating innovations. Descriptive statistics (such as median) clearly show that the results typed in Google Trends are better when taken from a user perspective and can be used to create innovations. Despite substantial differences, the results do not allow for the conclusion that these differences were statistically significant. Thus, preliminary data supports the hypothesis, but more research is needed.

JEL codes: C44, D81, M1, O31

Keywords: *better decisions*, *identifying needs*, *keyword analysis*, *product innovation*, *sources of ideas*