

Received May 5, 2020, accepted May 26, 2020, date of publication June 1, 2020, date of current version June 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999160

Direct Answers in Google Search Results

ARTUR STRZELECKI^{ID} AND PAULINA RUTECKA^{ID}

Department of Informatics, University of Economics in Katowice, 40-287 Katowice, Poland

Corresponding author: Artur Strzelecki (artur.strzelecki@ue.katowice.pl)

This work was supported by the Ministry of Science and Higher Education of the Republic of Poland for the University of Economics in Katowice for 2020.

ABSTRACT This study aims to explore Google's new "featured snippet" (a direct answer that appears in web search results) to better understand the method of extracting and displaying such answers above regular ranked results. The motivation for the study was to investigate form, structure, and relation to query, content, and domain of the direct answer. This study utilized a dataset of 743,798 keywords and the displayed direct answers. It also included web search data for every direct answer, e.g. resulting domain, full URL address and its original ranking, content of the direct answer, and competition (measured as the number of monthly searches for every query). The main finding concerns the form of construction of keywords used in the website content. Keywords should be built in the form of short, two-to-four-word sentences comprising the subject and its attribute. Using relative pronouns, articles, and prepositions, as well as using questions as queries, can help to properly define a query and display the best direct answer. The dataset is relatively small compared to the volume of searches made daily. No other factors were extracted from the URL, and all data concerned the Google search engine only. Implications for webmasters include: keywords used in website content should be close to grammar forms used in queries; use keywords in URLs; write a comprehensive introduction section on the webpage; and use tables and lists HTML markup. The main objective of this study was to examine the structure of the user's queries which caused in appearing of direct answers. It is comprehensive and based on real data used by web search engine users. The data used to support the findings have been deposited in the Zenodo repository (<https://doi.org/10.5281/zenodo.3541092>).

INDEX TERMS Keyword search, metasearch, search engines, search methods, search problems, web search.

I. INTRODUCTION

Google is now the most important web search engine. Google engineers are constantly developing the search algorithm to make it the most consistent with mobile standards and the most voice search adapted. The first extended snippets were introduced by Google in 2012. Their goal was to show the most important information on the search engine results page (SERP). In addition to the snippets like a knowledge graph or multimedia carousel, which have been popular for several years, direct answers are becoming an increasingly common snippet.

The direct answer snippet, also called the featured snippet, was introduced by Google in 2016. It appears as a succinct response in the form of a paragraph of text, list, or table. The direct answer gives an immediate response to the user's query. This kind of snippet is also suitable for voice search. Voice search is becoming more popular by introducing it not only on Google Home devices but also as Google's (voice) Assistant on mobile phones. The Assistant initially was deployed

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

in May 2016. Direct answer snippets are placed on the top of the search results page. They are also designed to be read by Google's Assistant.

The motivation for the present study is to explore the new snippet. The new snippet is displayed in a form of a direct answer. This direct answer appears in web search results. We would like to better understand what we can learn from the method of extracting and displaying direct answers in web search results. Much research has been performed on regular snippets and this field is well explored [1], [2]. However, direct answers, extracted from the content of the page, represent a new field that has not yet been comprehensively studied.

Thus, the current gap in the literature centers on the lack of research on direct answer snippets. There have been several studies on regular or rich snippets, but all of these are placed among 10 regular results [3]–[5]. Direct answers represent a completely new method of presenting results in SERPs, and this study aims to explore this topic by analyzing retrieved data from SERPs.

The objective of this study is to analyze a substantial volume of queries, along with search results data. This data

contains a list of 743,798 keywords, resulting in direct answers, and their parameters, such as original position, type, source, length, content, and language. Based on the above discussions, the following research questions related to direct answers are proposed.

First, users search in very different ways, ranging from entering very simple queries containing only one word to searching for whole phrases containing many words. Google, however, accepts a maximum of 32 words in a query [6]; the rest are ignored. In this regard, this paper aims to understand what length/number of keywords direct answers are more often displayed:

RQ1. What is the expected length of keywords for triggering direct answers?

Second, users search using simple grammar forms, like groups of nouns and adjectives, nouns and verbs, or more complex forms such as using whole sentences with the context. This paper aims to understand what kind of grammar forms used in queries trigger direct answers more often:

RQ2. What grammar forms are significant in direct answers?

Finally, webmasters can create pages and URL addresses in a friendly form. Friendliness means that the URL is built from words. This paper aims to understand if there is a connection between words used in queries and words used in URLs:

RQ3. How important are keywords in URL and answer content in choosing websites as reliable sources for direct answers?

The advantage of this study is that it uses real search data. By having a set of 743,798 queries, this study was able to collect the same number of direct answers. It also collected supporting data such as the position of results, where direct answers were extracted from, and the source of origin, e.g. the website's domain name. This study makes several observations regarding current direct answers appearing in search results.

The main finding concerns the form of construction of keywords used in the content of the website. Keywords should be built in the form of short, two-to-four-word sentences consisting of the subject (in this case, a noun, which unambiguously defines what is asked) and the attribute for this word. Using relative pronouns, articles, and prepositions can be helpful in properly defining a query and displaying the best answer as a featured snippet. To obtain a featured snippet answer, it is also useful to use question sentences as queries. When using simple queries containing only a noun, "entities" (e.g. a Knowledge Graph) are more likely to be returned by the search engine.

The paper is organized as follows. Section 2 contains a review of the relevant literature on search engine results and snippets, their different types, and research conducted on them. Section 3 includes the method and material for data retrieval and processing, while section 4 presents the data and quantitative results. In section 5 the authors discuss the contribution of the research. In section 6 the authors conclude its limitations and, finally, draw conclusions

about the results and propose possible future research avenues.

II. LITERATURE REVIEW

Basic search engine results page is combined data of unique keywords, positions, and URL results. According to the concept of search engine visibility described in [7], the visibility of websites in search engines comes from algorithms that rank and order them according to calculated ranking positions. The original concept [8] of ranking for the Google search engine is named PageRank, after one of Google's founders. PageRank was invented and published in 1998 [9]. This concept takes into account incoming links, and based on volume and quality, ranking positions for websites and corresponding keywords are estimated [10]. Currently, web search engines use different ranking factors for websites to determine their position on a results page.

Today this topic is attracting more attention [11] and can be divided into onsite and offsite factors [12]. Onsite factors are domain-, website-, and page-related [13]. Search engines take into account different elements found in the source code of a webpage such as title, headings, descriptions, time of last update, mobile design, and structured data for rich snippets [14]. Offsite factors are link-related [15], user action-related [16], special rules-related [17], brand-related [18], and spam-related [19]. Based on these factors, the search engine creates and displays results, as answers to the questions asked [20]. Regular results are presented in the form of a number of snippets presented in the order. It is usually ten snippets presented on one page. Snippet is built from title, URL address short text extracted from website [21].

Recent works on snippets are on snippets length, text reuse, and evaluating quality. Snippets are investigated in terms of measuring the gaze behavior of web users who interact with SERP's that contain plain and rich snippets, and observe the impact of both types of snippets on the web search experience [22]. Concerning snippet length, short snippets on mobile devices of one line are considered to provide too little information about the result, so that search performance and subjective measures are negatively affected regardless of query type. Long snippets of five lines lead to better performance than short and medium snippets [3]. Maxwell *et al.* [2] tested conditions where the change in information gain from snippets was the greatest. Four different scenarios were tested with 1) title only; 2) title plus one snippet; 3) title plus two snippets; and 4) title plus four snippets. Search engine users broadly prefer longer result snippets, as they are perceived to be more informative. In most cases, search engines for mobile devices present two or three lines of snippet for each result link. Maxwell *et al.* [2] suggest that long snippets provide a better search experience on desktop screens, but this may not be true for mobile devices because of the smaller screen. According to Kim *et al.* [4] users with long snippets on mobile devices exhibit longer search times with no better search accuracy. This is caused by the longer reading time, frequent scrolling with bigger viewport movements, and greater time

consumption for searching and reading one result. The overall findings suggest that, unlike desktop users, mobile users are best served by snippets of two to three lines.

Snippets in the form of tables that are extracted from web pages are a key component of search features such as tabular featured snippets from Google and Bing. Descriptive titles provide crucial context for interpreting these tables. One approach is to produce titles by selecting existing text snippets associated with the table. Different approach is accomplished by extracting many text snippets that have potentially relevant information to the table, encoding them into an input sequence, and using both copy and generation mechanisms to balance relevance and readability of the generated title [23]. However, Chen *et al.* [1] suggested another approach, instead of the current model of “reuse snippets”, snippet generation is in the form of paraphrase.

A significant area on research in snippet concerns qualitative studies. Lurie and Mustafaraj (2018) [24] investigated the effects of Google SERP in evaluating the credibility of online news sources. They noticed that the knowledge graph, the freshness of top stories, the panel of recent tweets, or a verified Twitter account are parts of the SERP that are used to assess the credibility of the source. Web search snippets in terms of quality evaluation are subject to credibility judgments. The same short snippets provide diverse informational cues and how these cues can be interpreted differently depending on the user and his or her background [25]. Snippets have different levels of readability and word complexity, however readability of snippets in Google and Bing search engines mismatched with the reading comprehension of children age 11 to 13 [26]. Google direct answers' quality was evaluated. It was found that Google provided significantly higher-quality answers to person-related questions than to thing-related, event-related, and organization-related questions. Google also provided significantly higher-quality answers to where- questions than to who-, what- and how-questions [27]. Snippets are also influencing partisanship. Generally, they amplify partisanship, and this effect is robust across different types of webpages, query topics, and partisan queries [28].

Search engines display different types of snippets on their SERPs. Snippets fit into five categories of differing presentations: normal snippets; rich snippets; Google News; entity types; and featured snippets [14]. Users may get information from the SERP directly, may or may not click through to read each resulting webpage, and may not even have the option of clicking if the query is very specific [29]. Users express their interest in different types of snippets as clicks, attention, and satisfaction on SERPs [30].

A. NORMAL SNIPPETS

Normal snippets are displayed for typical, regular organic results. The prototype web search engines displayed normal snippets as two lines of description placed below the title and URL of the result [5]. Recently it has been noticed that commercial web search engines are testing changes in the

length of normal snippets. This change is observable both on desktop and mobile versions [3], [4]. Varying relatively few words, and even their location, within a snippet can have a significant influence on the clickthrough of a snippet [31]. Normal snippets in previous works have mainly been evaluated as being informative enough for users of web search engines [25]. Recent research has been conducted on different age groups to see how normal snippets are perceived by younger and older search engine users [26]. Regular snippets are also a source of data for explaining unknown terms through automatic translation [32]. Normal snippets are also able to influence users; for example, political partisanship snippets can amplify partisanship and influence undecided voters [28].

B. RICH SNIPPETS

Rich snippets are based on the structured data dictionary schema.org [33]. Google, Microsoft, Yahoo, and Yandex founded this common project, schema.org, and are interpreting structured data included in RDFa, Microdata, and JSON notation [34]. Rich snippets, interpreted through structured data, are displayed together with normal snippets [35]. Search engines show results with structured data, created on schema.org, on product availability, price and condition, recipes, reviews, jobs, music, video, etc. Rich snippets are considered an important element of SERPs, especially when examining results placed at the bottom of SERPs [22].

C. NEWS SNIPPETS

News snippets are displayed from online newspapers' headlines through the Google News service. These snippets are part of this vertical news aggregator and are provided completely automatically [36]. Google News automatically displays results as a snippet, together with images for results in countries where Google News is available [37]. Online newspapers have reacted to these snippets in different ways. Recently, in Germany and Spain, Google News has been restricted based on claims that this practice of displaying snippets of news releases violates news publishers' copyright [38]. Hence, news snippets are considered to be driving changes in European law [39]. In literature, there are several suggestions for solving this possible violation; for example, a plan for ancillary copyright has been proposed by creating original snippets [40].

D. ENTITIES

This fourth category comprises snippets created from entities. Entities in Google are known as “Knowledge Graph” (launched in 2012) and in Bing are known as Satori (introduced in the same year) [41]. These entities are constructed objects and concepts, including people, books, events, movies, places, science, arts, etc. [42]. These entity databases, e.g. Google Knowledge Graph, are considered by users as an important part of web search results [43]. In a recent study, however, results have indicated that there are widespread inconsistencies in the coverage and quality of the information included in the Knowledge Graph [24].

E. FEATURED SNIPPETS

Featured snippets represent the latest observed improvement in SERPs. The search engine retrieves pieces of information from web pages and displays them in the answer box, above organic results, together with a source URL and the title of the page [44]. Google automatically determines that a page contains an answer to the user's query and presents the result as a featured snippet. The other known names for featured snippets are answer boxes or direct answers. The goal of direct answers is to deliver results for a query without the need to visit the page presented in the search engine [45]. Featured snippets are displayed in three different forms, i.e. paragraphs [27], tables [23], or lists, either ordered or unordered [46].

In the beginning of featured snippets' introduction into SERPs, Miklosik *et al.* [46] examined the different possible factors for presenting a piece of information from a particular webpage in featured snippets. Featured snippets were named by Google as "answer boxes" and Google noticed that several prerequisites needed to be met for a website to be included in the featured snippet box: high ranking in the SERP; multiple keyword inclusion in the webpage's content; different locations for keywords like headings, title, URL, paragraphs, images' alternative descriptions, and links; and structured content in the form of ordered or unordered lists.

Recently, [47] examined 163,412 keywords in Polish that resulted in SERPs with featured snippets. Analysis of the data showed that the appearance of a featured snippet is closely related to the question form of the phrase (the occurrence of pronouns) or the occurrence of words in the phrase specifying an attribute that has a specific value, considered as the answer to the query. It was also observed that almost half of the featured snippets (48%) were taken from the result in the first ranking position.

Not every query entered in the search engine returns content in the featured snippet form. The observation of search engine algorithm history shows that the featured snippet form will be the most expected in the Google development process. In connection with the query, it gives a result similar to the conversation in natural language [27] and the snippet content is possible to be read at loud and it's suitable for voice search responses [47].

Researches on the way of entering a query to the search engine that causes the return of specific results in Google have led to the typology of queries. Base typology distinguishes three types of queries: informational queries, navigational queries, and transactional queries [48]. Based on later research, it was extended by commercial and local query types [49]. This division is currently in use.

A new technique based on a neural network for natural language processing (NLP) - BERT (Bidirectional Encoder Representations from Transformers), introduced by Google in 2018 and described in 2019, changes the approach to reading user input queries. BERT is analyzing words immediately before or after keywords [50], including skipped in the classic meaning of NLP stopwords [51].

In the study authors decided to analyze the queries, in terms of their length, grammatical structure, and content of keywords in the URL, which led to the return of featured snippet in the context of a new query understanding technique. Research can be a starting point for typology update or creation of a new one, aimed at identifying query types that generate snippets.

III. METHODS AND MATERIALS

The dataset was collected using Senuto (<https://www.senuto.com>). Senuto is an online tool that extracts data on websites' visibility from the Google search engine. Senuto crawls data from Google daily, based on its queries list, and saves results, along with the ranking position. Senuto currently has a database of 20 million queries. Each query is entered at least once per month into a Polish-localized Google search engine and a list of the top 50 results is returned.

A dataset was acquired in September 2019. The date range of the dataset is covering the period from 1 September to 15 September. As the basic structure comprises only keyword, position, and page URL, the authors asked Senuto's owners to extend the crawling procedure tool to be able to extract more data from SERPs. The goal of this research is to examine direct answers; therefore an extended structure for extracting results was prepared. Our direct answer extraction algorithm is presented as Algorithm 1. It was designed based on the approach proposed as Machine Reading Comprehension for feature snippets extraction [52].

Algorithm 1 Direct Answers Extraction

```

Require: keywords
rawDirectAnswers = []
for all keywords do
  if directAnswerPattern exist in current.SERP then
    rawDirectAnswers.append (extractDirectAnswerPattern
      (currentSERP))
  end if
end for
for all rawDirectAnswers do
  refineRawDirectAnswers (currentrawDirectAnswer)
end for
features = findFeatures(refineRawDirectAnswers)
return features

```

The first step is that the entry of the algorithm is a list of keywords. The primary list of keywords is collected from search engine suggest and autocompletion tools. The list is extended by the snowball mechanism. The more keywords collected, the more suggestions and autocompletion. The next step identifies direct answer patterns. Fortunately, the search engine uses a pattern to list and clarify the direct answer. A direct answer pattern consists of four parts: the content of the direct answer, the original position of URL, from which the direct answer is displayed in SERP, the full URL of a direct answer, and its type. From the full URL we filter

TABLE 1. Statistics for direct answers' snippet length comparison for top six languages.

	English	Polish	German	Spanish	Italian	French
Upper whisker	424.00	439.00	422.00	417.00	424.00	424.00
3rd quartile	307.00	302.00	303.00	308.00	306.00	299.50
Median	277.00	264.00	268.00	274.00	265.00	253.00
Mean	260.78	248.02	250.15	256.01	249.39	232.86
1st quartile	229.00	206.00	211.00	221.00	206.00	174.00
Lower whisker	112.00	62.00	73.00	91.00	57.00	20.00
No. of data points	365538	335115	19432	10090	4799	4976

domain and main domain name. In the return features of the snippets are collected.

In the extended visibility structure, along with a basic set of data, the following additional elements were collected:

1. Keyword (also called a query): the term entered to search engine interface. Keyword length is from one term up to 32 terms in the case of Google. Other search engines accept other maximum numbers of terms included in one query.
2. Number of monthly searches: the number of searches for this keyword in the Polish-localized version of Google. Data were imported through an API from Google Ads.
3. Featured domain: a specific part of the URL between the protocol (HTTP or HTTPS) and the first slash.
4. Featured main domain: the normalized domain name. It does not contain any prefixes before the name, such a subdomain.
5. Featured position: the ranking position of the result displaying the same URL that is displayed in direct answer. Direct answers are extracted only from pages in positions 1 to 10 from the SERP's first page.
6. Featured type: defines dimensions of displayed feature snippet (paragraph, list, and table). All three types were found in the dataset.
7. Featured URL: the final URL of the result that the piece of information is displayed from.
8. Content: the exact piece of information extracted from the direct answer box and normalized to text only by removing the HTML code.
9. Content length: length of the normalized text of the direct answer measured in number of characters.

The dataset with the extended visibility structure had 743,798 crawled keywords that resulted in direct answers on SERPs [(the authors have deposited this dataset as an open access upload at Zenodo (<https://doi.org/10.5281/zenodo.3541092>))].

Before proceeding further with these data, the authors noticed that keywords and direct answers were in many different languages. The most common languages were Polish and English. The authors used a function in Google Sheets (*detectlanguage*) to set an additional parameter for each direct answer's language. This function detects the language of a text in a cell. In cases where a direct answer contains two languages, the one that is chosen is the one in which the function has the most confidence. After detecting the language of all the direct answers' content, results revealed 94 different languages. The top six languages had more than 4,000 results. The authors also observed that the descriptive statistics for each language were slightly different (Table 1.)

At this stage, the authors decided to choose only one language for further analysis. The natural choice was English as this language we had the highest number of snippets and English is the most common language.

IV. RESULTS

This section details prepared data about direct answers crawled from the Google web search engine and provides the results after analysis of the downloaded data, followed by the discussion of the results.

The analyzed dataset contained 365,538 results in English. The results in English were a priority dataset for this work. In the data analysis, special attention was paid to the length of queries; repetition of words in the construction of queries divided into parts of speech; an occurrence of question words at the beginning or elsewhere in the query; form of returned results (paragraph, table, list); comparing the words from the query to the words used in the snippet content and the source URL; source page position in SERP; top source domains and credibility analysis (owner and webpage type). To enhance readability, the analysis has been divided into three parts related to the type of data. The first part presents data related to the query that caused the presentation of a specific featured snippet. The second part presents the results of the featured

TABLE 2. Length of query (words) and occurrence.

Length of query (words)	1	2	3	4	5	6	7	8	9	10
Number of queries	4,004	97,699	136,012	80,967	31,156	9,992	3,590	1,388	513	217

TABLE 3. Volume of the parts of speech and specific words in queries.

Adjectives		Nouns		Question words	
Word	Volume	Word	Volume	Word	Volume
vs	17564	games	3765	review	7674
to	12846	Windows	3715	meaning	7116
in	11693	Android	3112	definition	4578
of	11109	Mac	1920	history	4528
for	6494	Google	1913	price	3883
a	5586	Linux	1761	specs	3151
and	4887	iPhone	2173	quotes	3362
how	4746	Sony	1566	code	2866
the	4422	pdf	1522	name	2564
on	3840	Samsung	1519	biography	2552

snippet content analysis in terms of the length and type of data presented. The third part concerns the source domains from which the featured snippet content was taken.

A. QUERY ANALYSIS

Table 2 presents the results of the query length analysis. In the studied dataset, the most common were three-word queries: 136,012 results, of which 37.2% were in English. Less often appeared queries consisting of two (97,699 results, 26.7% of all) or four words (80,967 results, 22.15% of all). About 8.52% of the set consisted of five words (31,156 results). Very long (7–10 words) and very short (one word) queries were rare. The length of queries may suggest that they were built in the form of an interrogative sentence or a phrase comprising a noun and a question word.

The way in which the most popular queries were built was additionally studied after splitting the queries into separate words. The occurrence of each of the appearing words was counted, without paying attention to the position in the query. The analysis of the 50 keywords appearing most frequently allowed the authors to distinguish specific groups. The 10 most common words in this group are shown in Table 3.

Other popular words, not classified into any of the groups, included free and best.

Analysis of Table 3 shows that users often use various parts of speech (e.g. relative pronouns, articles, prepositions) to build queries. The 10 most common are listed in Table 4, column 1. These words were found in the entire collection 83,187 times. It is worth emphasizing that these words did not occur in single-word phrases. Furthermore, not every word that appeared in the phrases was analyzed. Users also used words in queries that were marked as questions (see Table 4, column 3). These were words representing various parts of speech, but their common denominator was being an attribute of something searched, e.g. price, definition, meaning, etc. The 10 words most frequently used in queries defined as attributes appeared 42,274 times in the entire set of analyzed keywords. Nouns (22,966 cases within the top 10 most used words in this group) appearing in the query were often brands name (e.g. Windows, Android, Mac). The most common nouns in the surveyed dataset were words related to the IT industry. We have analyzed words, that usually were considered as stopwords, however, since October 2019, Google can consider the full context of a word by looking at the words that come before and after it [50].

TABLE 4. Interrogative sentence question words and linking verbs in query phrase.

Beginning of the phrase		Any other position of the phrase	
Word	Occurrence	Word	Occurrence
how	2,627	is	2,381
what	962	how	2,119
is	804	what	674
who	387	do	607
where	133	does	489
why	125	are	184
do	97	has	184
does	93	have	161
when	68	who	126
will	68	where	110

To investigate whether the users typing the queries paid attention to the grammatical form of the queries (e.g. forming them like natural language), a group of interrogative sentence question words and linking verbs was distinguished. The results of this analysis were divided into two parts: the number of question words and linking verbs appearing at the start of the sentence; and the number of question words and linking verbs not appearing at the start of the sentence. Table 4 shows the 10 most used question words and linking verbs. However, the analysis focused on such question words as how, what, is, who, where, why, do, does, when, will, if, which, whose, are, was, did, have, has, had, and were. The question words and linking verbs at the beginning of the query appeared 5,517 times, compared to 7,571 times not at the beginning of the sentence. In total, question words appeared 13,088 times.

B. CONTENT ANALYSIS

Figure 1 presents the results of the analysis of the occurrence of the individual words that were used to build the key phrases. The “snippet content” bar shows the percentage of the individual words present in the content of the snippet. The “URL” bar shows the percentage of the individual words in the URL that was the source of the snippet. To carry out this analysis, keywords were divided into columns (each word in a separate column). In the next step, a condition for each column was created that checked whether the word was in the URL or snippet content, returning a “True” or “False” value, depending on the result. Because of the variety of

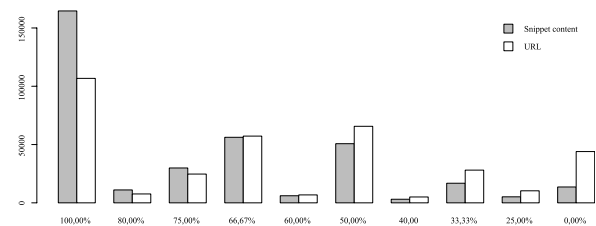


FIGURE 1. The percentage of individual words used to create the query present in the content of the snippet or URL.

the keywords’ length, the results of empty tables returning “True” were rejected. In 45.03% of snippet content, all the individual words from the query occurred. For the URLs, the figure was 29.21%. Snippets that contained at least a half of the words used in the query constituted 87.16% of the analyzed dataset. For the URLs, the figure was 73.53%.

In the analyzed dataset, every form of featured snippet was presented: (paragraph, list, and table), with paragraphs being the most common. The results in Table 5 also show the average number of phrase searches for which a specific featured snippet form was displayed. The average number of searches was similar for every type of snippet and ranged from 20.55 for paragraphs to 21.95 for lists. Also, the average position of the page considered as the snippet source in the search results was similar; however, it could be observed that the more specific the form of the data presentation, the higher the average position was. The best positions were taken by the URLs constituting the sources for tables (2.54), and the worst for paragraphs (2.78).

TABLE 5. Displayed result types.

Result type	Volume	Percentage of total	Monthly average of searches	Average position of the source page	Average volume of words in snippet	Max. words in the content	Min. words in the content
Paragraph	276,345	75.60	20.55	2.78	42.13	90	1
List	77,857	21.30	21.95	2.74	50.27	107	4
Table	11,336	3.10	20.84	2.54	30.97	88	4

TABLE 6. Ranking positions of page URL.

Ranking position of the page	Number of occurrences	Percentage of total
1	134,682	36.84
2	66,305	18.14
3	54,152	14.81
4	43,886	12.01
5	35,352	9.67
6	11,429	3.13
7	7,411	2.03
8	5,370	1.47
9	3,997	1.09
10	2,954	0.81

According to results in Table 1, the average length of snippet was 260 characters, with the median being 277 characters. Segmenting this into snippet types revealed that the average volume of words in content differed for each type: 30 words for tables; 42 words for paragraphs; and 50 for lists. This shows that the online paragraph must be short, consisting of 50 words or less.

C. DOMAIN ANALYSIS

Table 6 presents data showing the position of the URL source on the SERP from which the information was used to display featured snippets. The data in Table 6 shows the position of web pages in the Google SERP from which the featured snippets were created. After a specific query was entered in the search engine, a snippet occurred that was formed from pages in positions 1 to 10 position from the SERP content. The source of featured snippets were usually pages that came first (134,682 results, which is 36.84% of the analyzed set)

in search results for a specific query. The lower the position of the page in the search results, the less often it was used by Google to create the featured snippet. In 91.47% of cases, the Google search engine used as the source of the featured snippet pages in positions from 1 to 5. Pages with positions from 2 to 5 were the source for 54.63% of featured snippets (position 2, 18.14%; position 3, 14.81%; position 4, 12.01%; position 5, 9.67%). Pages with positions from 6 to 10 were used less frequently than in 5% of cases in the entire analyzed dataset.

The Pearson correlation coefficient between the source page ranking position in the SERP and the number of occurrences of snippets from pages with this ranking position is -0.883 ($p < 0.001$). The negative correlation value indicates that the decrease in the search results is accompanied by a decrease in the number of pages that were used to build snippets. The negative correlation also shows that the numbering position is reversed: the lower the number, the better the result

TABLE 7. Top 20 source domains for featured snippets.

Lp	Main domain	Average position of the source page for list	Average position of the source page for paragraph	Average position of the source page for table	List results from a domain	Paragraph results from a domain	Table results from a domain	Average position for all	Total results from a domain
1	wikipedia.org	2.35	2.09	2.34	2,122	50526	1977	2.11	54,625
2	fandom.com	1.80	2.10	1.83	1,142	3160	181	2.01	4,483
3	quora.com	2.93	2.64	n/d	341	4101	0	2.66	4,442
4	youtube.com	1.70	2.34	1.20	23	3792	15	2.33	3,830
5	nih.gov	2.24	1.79	2.06	42	3756	32	1.8	3,830
6	healthline.com	1.90	2.18	1.88	628	1885	8	2.11	2,521
7	investopedia.com	2.48	2.18	n/d	29	2320	0	2.19	2,349
8	thefreedictionary.com	1.58	2.56	1.55	89	1779	420	2.34	2,288
9	merriam-webster.com	3.14	2.82	n/d	128	2043	0	2.84	2,171
10	wikihow.com	2.79	2.69	4.00	1,457	616	1	2.76	2,074
11	amazon.com	4.13	3.80	3.94	391	1466	183	3.88	2,040
12	lifewire.com	2.59	2.45	1.91	567	1245	106	2.47	1,918
13	goodreads.com	1.60	2.11	3.00	1,345	480	1	1.73	1,826
14	britannica.com	3.13	2.83	1.00	24	1792	1	2.83	1,817
15	yourdictionary.com	2.32	2.36	1.86	155	1,523	7	2.35	1,685
16	cnet.com	2.80	3.73	3.47	495	1,062	15	3.43	1,572
17	stackoverflow.com	2.26	2.04	n/d	208	1,298	0	2.08	1,506
18	techtarget.com	4.00	3.65	1.00	15	1,420	1	3.65	1,436
19	microsoft.com	2.19	2.18	1.91	392	1,008	33	2.17	1,433
20	imdb.com	1.85	3.26	2.45	493	826	56	2.72	1,375

in the search results. The correlation between these values is very strong.

Table 7 presents the top 20 source domains for featured snippets. It also contains detailed information about the number of featured snippets created, broken down by type featured snippets (paragraph, table, list), and the average

position of the domain in the search results, broken down into featured snippets (paragraph, table, list). In the entire dataset of 365,538 featured snippets, 52,727 domains were used. Pages from the top 20 domains were the source of 99,221 featured snippets, which represents 27.14% of the entire analyzed dataset.

The most featured snippets were created based on information from the wikipedia.org domain: 54,625 results from this domain, representing 14.94% of all analyzed results. Wikipedia.org is recognized and trusted in the category of online encyclopedias created on a social basis. The second and third positions in the list of the top 20 domains were fandom.com (4,483 results) and quora.com (4,442 results). The goal of these two services is to present answers to frequently asked questions. They both are based on a Q/A formula. In fourth position was youtube.com (3,830 results, average URL position 2.33), which presents movies and music. The fifth position was the government website devoted to health, nih.gov (3,830 results, average URL position 1.8), created by the US Department of Health & Human Services. The remaining domains mainly comprised websites about health and IT, the main types of which were online dictionary, encyclopedia, or Q/A formula.

Among the top 20 domains there were also pages about books, movies, games, and one e-commerce site: amazon.com. The Amazon site was in the top 20 possibly due to the large number of product descriptions that can be the source of featured snippets. Besides, amazon.com is highly trusted by consumers and is located high in search results. Of the total number of 52,727 URLs used to build featured snippets in the analyzed dataset, the URLs that appeared only once numbered 30,857, representing 58.52% of all domains. It is worth noting that over 42% of domains were a source for featured snippets more than once.

V. DISCUSSION

The purpose of the featured snippets is to present the answer to the user's question with no need to visit a specific webpage. There is a trend in which Google tries to accumulate various technological solutions under its brand. In this way, Google creates a closed technological ecosystem. In exchange for the use of content from a specific website as part of featured snippet, Google highlights where the information was collected from. In this way, it ensures the authors' copyright, while recognizing them and their brand as experts.

The second reason for creating featured snippets is being the growth of voice queries. For this solution, the possibility to choose only one answer allows it to be read by a voice synthesizer. Keywords that caused the appearance of featured snippets were no longer than 10 words. This guarantees to choose the best answer for the query. Almost 30% of queries that featured snippets were built in a grammatical form (interrogative sentence, use of prepositions, or hyphens).

This confirms the trend observed by the authors about the growing interest in voice searches. This trend is characterized by users utilizing queries similar to every-day natural language. Featured snippets, in this case, are similar to natural answers, which the search engine uses for communication with a user. Overlapping keywords from a query with the keywords in snippet content or URL address is very important. The more the user's query coincides with the phrase used

on the page, the more likely this page will be used to build featured snippets.

Featured snippets were created for queries with different average monthly searches. The monthly search average was not a determinant for Google to build a featured snippet for a specific query. More than 25% of featured snippets were built based on 20 top domains with high positions in search results: 2.523 on average. These websites are mostly dictionaries, encyclopedias, Q/A forms, or pages about IT, health, books, and movies. Over 42% of domains were a source of more than one snippet. The sources of featured snippets were websites with a high rank in the search results, and these top positions in the search results are domains that meet the requirements of Google algorithms.

This paper has presented an analysis of the dataset containing keywords along with direct answer results in Google web searches. After using the *detectlanguage* function to divide the initial dataset of 743,798 keywords into languages, it was found that half the data were in English. The authors, therefore, proceeded to analyze only this part of the dataset. The analysis covered three dimensions: queries, content, and domains. The findings indicate that the Google search engine is being developed in the direction of displaying the exact answer for the query directly from the search results page. Google does not decrease the importance of regular results, but makes the most valuable page URL from the top 10 results stand out.

There is no automatic inclusion and no schema that can be added to websites for them to be considered for use in direct answers. That is why the findings on a queries level can help webmasters to prepare better content on their websites for inclusion as snippets.

The analysis of the data shows that the appearance of a featured snippet is closely related to the question form of the query (the occurrence of pronouns) or the occurrence of words in the query specifying an attribute that has a specific value, considered as the answer to the query, e.g. product price. Frequently appearing queries contain adjectives in the superlative form, e.g. the highest peak, the largest city. All these queries have unambiguous and undeniable value.

Google uses as a source for direct answers only websites with search-engine rankings, i.e. they meet the criteria of Google's ranking factors. Meeting these criteria defines the website as containing content of a high value that is popular among users. This means that the website enjoys the trust of both the search engine algorithm and the users themselves. This makes it very likely that the information in the direct answer is correct and accurate. The same websites, such as wikis or dictionaries, are often cited.

The most common type of direct answer is the result in the form of a paragraph. This kind of result is the most legible and at the same time the most convenient to be read by the voice search Assistant. It potentially allows the computer to use a voice response using a speech synthesizer. Results in the form of a list or table appear less frequently. The form of a list usually appears in cookery recipes (ingredients for preparing

a dish) or in health and medical area, when symptoms of a disease are mentioned. The form of the table appears most frequently for queries regarding prices or other values related to financial products (such as taxes, loans, and insurance) or other data originally presented as a table.

Direct answers often occur along with an accompanying image. Images draw the user's attention by illustrating the presented result, as well as linking to the page URL from which the data were used to create the direct answer. By occupying the highest position in ranking results, it is expected that direct answer has a higher click-through ratio than results below, and domains displayed in this position gain expert-opinion status. Often, direct answers are called zero position results [44].

RQ1 was answered using the results of the query length analysis. Keywords in the analyzed dataset contained only 1 to 10 individual words, despite the search engine accepting 32 different words in a query. Keywords longer than 10 words did not generate featured snippets. The most often used length of the query was two to four words.

Regarding *RQ2*, analyzing the data revealed that, to generate a featured snippet, the query should be built as a grammatically correct sentence. The use of the question form in a sentence or using words like relative pronouns, articles, or prepositions often resulted in obtaining featured snippets compared to using short terms. Queries containing a question about the attribute (price, definition, review, etc.) returned answers in the form of featured snippets.

The answer to *RQ3* is that keywords in the URL and answer content are important. In 73.53% of cases, more than half the words from the queries were in the URL, and in 87.16% of cases, more than a half the query words were in the answer content. The proper placement of keywords in friendly URLs or site content increases the likelihood that the page will be used as the source of featured snippets. Through this research, therefore, it is possible to define webmaster guidelines. These guidelines contain not only technical suggestions, but also tips on how to properly write and format texts on a website.

VI. CONCLUSION

This section contains practical implications of the results, followed by limitations and further research.

A. PRACTICAL IMPLICATIONS

This study also yields several direct implications for website owners or webmasters. First, keywords inserted into the content of the webpage should be as close in grammatical form as possible to those used in queries asked by users. Second, it is recommended to put keywords in the URL. Nowadays, URLs are often written in a user-friendly form, containing words separated by a dash separator. The study confirmed that around 70% of URLs contained keywords from the query. Putting the right keywords in the page URL will help it to be considered as a source for direct answers. Third, when preparing content for the website, the introduction section should not only provide a brief description, but also contain a

comprehensive answer to the query. This increases the chance of this part of the website being used as a direct answer.

Fourth, webmasters can use tables in HTML markup. Properly marked HTML tables with table headings appeared in direct answers. The tables in the present study contained a maximum of four data rows and one header row; if the source data had more rows, information about how many rows were left was displayed. These tables had a maximum of three columns; although the source data could have many more columns, they were reduced to a maximum of three columns. The header line was bold and the table title came from the HTML header immediately preceding the table on the source page. Next to the table, there may or may not be a photo illustrating the result. The same rule applies to lists in HTML. The website owner can use ordered or unordered lists to present content on the page. The search engine was able to extract this list and use it as a direct answer. Fifth, it is recommended that webmasters should follow guidelines prepared by Google for website owners. Webmasters must follow general guidelines concerning discoverability, readability, and quality issues [53]. Sixth, the average number of searches for a query does not determine whether direct answers will be displayed or not; it was observable that direct answers exist for queries with a very low number of searches.

B. LIMITATIONS

This study has several limitations. First, the set of data comprised only 743,798 keywords and direct answers, with only 365,538 of them in English. This dataset is relatively small compared to the volume of searches made daily by search engine users, which reached in 2016, 63,000 searches per second (5.5 billion searches per day) [54]. Second, all the data concerns the Google search engine only, which is the dominant search engine in most countries [55]. The authors realize that results containing featured snippets are observed in other search engines, like Bing, but its current market share is very low. Exploring data from the Google search engine, this study used popular keywords and results that are used by most of the internet user population.

Third, this study did not introduce any other factors related to the featured URL content. These factors could be extracted values of different HTML tags like title, headings, ALT descriptions, meta description, etc. Fourth, when analyzing the individual results, the authors noticed that there were differences between displaying snippets using the same query in a voice search (Google Assistant) and a browser search. The results for the same query could lead to snippets for a voice search but not for a browser search.

C. FURTHER RESEARCH

Future research should be conducted to investigate the factors affecting the display of direct answer from specific websites in the featured snippets area. Another direction of future studies is to analyze the content of the featured snippet in other languages, not only in English. A comparison of snippet

length for six languages showed that descriptive statistics for snippets are different for different languages.

Methods for data analysis can be applied to other search engines. The presented approach can be generalized. Although this study focus is on the Google search engine, because of its dominant position among other search engines, a similar approach is possible to do with Bing or Baidu. Both of them present extend answers to queries.

Currently, no other vertical search, except web and voice, lead to feature snippet presentation. Image or video searches do not return a feature snippet. However, Google can return a random image against the featured snippet if considers it helpful. Yet, this study does not cover this image part.

VII. DATASET

The data used to support the findings of this study have been deposited in the Zenodo repository (<https://doi.org/10.5281/zenodo.3541092>).

REFERENCES

- [1] W.-F. Chen, M. Hagen, B. Stein, and M. Potthast, "A user study on snippet generation: Text reuse vs. paraphrases," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 1033–1036, doi: [10.1145/3209978.3210149](https://doi.org/10.1145/3209978.3210149).
- [2] D. Maxwell, L. Azzopardi, and Y. Moshfeghi, "A study of snippet length and informativeness," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 135–144, doi: [10.1145/3077136.3080824](https://doi.org/10.1145/3077136.3080824).
- [3] J. Sachse, "The influence of snippet length on user behavior in mobile Web search," *Aslib J. Inf. Manage.*, vol. 71, no. 3, pp. 325–343, May 2019, doi: [10.1108/AJIM-07-2018-0182](https://doi.org/10.1108/AJIM-07-2018-0182).
- [4] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon, "What snippet size is needed in mobile Web search?" in *Proc. Conf. Hum. Inf. Interact. Retr. (CHIIR)*, Mar. 2017, pp. 97–106, doi: [10.1145/3020165.3020173](https://doi.org/10.1145/3020165.3020173).
- [5] D. Lewandowski, "The retrieval effectiveness of Web search engines: Considering results descriptions," *J. Documentation*, vol. 64, no. 6, pp. 915–937, Oct. 2008, doi: [10.1108/00220410810912451](https://doi.org/10.1108/00220410810912451).
- [6] P. Stox. (2017). 20 of Google's Limits You May Not Know Exist. SearchEngineLand. Accessed: Apr. 30, 2020. [Online]. Available: <https://searchengineland.com/20-googles-limits-may-not-know-exist-281387>
- [7] A. Strzelecki, "Google Web and image search visibility data for online store," *Data*, vol. 4, no. 3, p. 125, Aug. 2019, doi: [10.3390/data4030125](https://doi.org/10.3390/data4030125).
- [8] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, pp. 107–117, Apr. 1998, doi: [10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X).
- [9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," World Wide Web Internet Web Inf. Syst., Stanford InfoLab., Stanford, CA, USA, Tech. Rep., 1998.
- [10] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999, doi: [10.1145/324133.324140](https://doi.org/10.1145/324133.324140).
- [11] W. Serrano, "Neural networks in big data and Web search," *Data*, vol. 4, no. 1, p. 7, Dec. 2018, doi: [10.3390/data4010007](https://doi.org/10.3390/data4010007).
- [12] C. Ziakis, M. Vlachopoulou, T. Kyrkoudis, and M. Karagkiozidou, "Important factors for improving Google search rank," *Future Internet*, vol. 11, no. 2, p. 32, Jan. 2019, doi: [10.3390/fi11020032](https://doi.org/10.3390/fi11020032).
- [13] M. P. Evans, "Analysing Google rankings through search engine optimization data," *Internet Res.*, vol. 17, no. 1, pp. 21–37, Feb. 2007, doi: [10.1108/10662240710730470](https://doi.org/10.1108/10662240710730470).
- [14] A. Strzelecki and P. Rutecka, "The snippets taxonomy in Web search engines," in *Perspectives in Business Informatics Research*, vol. 365, M. Pańkowska and K. Sandkuhl, Eds. Champaign, IL, USA: Springer, 2019, pp. 177–188.
- [15] S. Zhang and N. Cabbage, "Search engine optimization: Comparison of link building and social sharing," *J. Comput. Inf. Syst.*, vol. 57, no. 2, pp. 148–159, Apr. 2017, doi: [10.1080/08874417.2016.1183447](https://doi.org/10.1080/08874417.2016.1183447).
- [16] E. Agichtein, E. Brill, and S. Dumais, "Improving Web search ranking by incorporating user behavior information," *ACM SIGIR Forum*, vol. 52, no. 2, pp. 11–18, Jan. 2019, doi: [10.1145/3308774.3308778](https://doi.org/10.1145/3308774.3308778).
- [17] A. Strzelecki, "Website removal from search engines due to copyright violation," *Aslib J. Inf. Manage.*, vol. 71, no. 1, pp. 54–71, Jan. 2019, doi: [10.1108/AJIM-05-2018-0108](https://doi.org/10.1108/AJIM-05-2018-0108).
- [18] J. P. Dotson, R. R. Fan, E. M. Feit, J. D. Oldham, and Y.-H. Yeh, "Brand attitudes and search engine queries," *J. Interact. Marketing*, vol. 37, pp. 105–116, Feb. 2017, doi: [10.1016/j.intmar.2016.10.002](https://doi.org/10.1016/j.intmar.2016.10.002).
- [19] M. Egele, C. Kolbitsch, and C. Platzer, "Removing Web spam links from search engine results," *J. Comput. Virol.*, vol. 7, no. 1, pp. 51–62, Feb. 2011, doi: [10.1007/s11416-009-0132-6](https://doi.org/10.1007/s11416-009-0132-6).
- [20] V. N. Gudivada, D. Rao, and J. Paris, "Understanding search-engine optimization," *Computer*, vol. 48, no. 10, pp. 43–52, Oct. 2015, doi: [10.1109/MC.2015.297](https://doi.org/10.1109/MC.2015.297).
- [21] J. B. Killoran, "How to use search engine optimization techniques to increase Website visibility," *IEEE Trans. Prof. Commun.*, vol. 56, no. 1, pp. 50–66, Mar. 2013, doi: [10.1109/TPC.2012.2237255](https://doi.org/10.1109/TPC.2012.2237255).
- [22] M.-C. Marcos, F. Gavin, and I. Arapakis, "Effect of snippets on user experience in Web search," in *Proc. XVI Int. Conf. Human Comput. Interact.-Interacción*, 2015, pp. 47:1-47:8, doi: [10.1145/2829875.2829916](https://doi.org/10.1145/2829875.2829916).
- [23] B. Hancock, H. Lee, and C. Yu, "Generating titles for Web tables," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 638–647, doi: [10.1145/3308558.3313399](https://doi.org/10.1145/3308558.3313399).
- [24] E. Lurie and E. Mustafaraj, "Investigating the effects of Google's search engine result page in evaluating the credibility of online news sources," in *Proc. 10th ACM Conf. Web Sci.*, May 2018, pp. 107–116, doi: [10.1145/3201064.3201095](https://doi.org/10.1145/3201064.3201095).
- [25] M. Kattenbeck and D. Elswiler, "Understanding credibility judgements for Web search snippets," *Aslib J. Inf. Manage.*, vol. 71, no. 3, pp. 368–391, May 2019, doi: [10.1108/AJIM-07-2018-0181](https://doi.org/10.1108/AJIM-07-2018-0181).
- [26] D. Bilal and L.-M. Huang, "Readability and word complexity of SERPs snippets and Web pages on children's search queries," *Aslib J. Inf. Manage.*, vol. 71, no. 2, pp. 241–259, Mar. 2019, doi: [10.1108/AJIM-05-2018-0124](https://doi.org/10.1108/AJIM-05-2018-0124).
- [27] Y. Zhao, J. Zhang, X. Xia, and T. Le, "Evaluation of Google question-answering quality," *Library Hi Tech*, vol. 37, no. 2, pp. 312–328, Jun. 2019, doi: [10.1108/LHT-10-2017-0218](https://doi.org/10.1108/LHT-10-2017-0218).
- [28] D. Hu, S. Jiang, R. E. Robertson, and C. Wilson, "Auditing the partisanship of Google search snippets," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 693–704, doi: [10.1145/3308558.3313654](https://doi.org/10.1145/3308558.3313654).
- [29] P. Thomas, A. Moffat, P. Bailey, F. Scholer, and N. Craswell, "Better effectiveness metrics for SERPs, cards, and rankings," in *Proc. 23rd Australas. Document Comput. Symp. (ADCS)*, 2018, pp. 1–8, doi: [10.1145/3291992.3292002](https://doi.org/10.1145/3291992.3292002).
- [30] A. Chuklin and M. de Rijke, "Incorporating clicks, attention and satisfaction into a search engine result page evaluation model," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 175–184, doi: [10.1145/2983323.2983829](https://doi.org/10.1145/2983323.2983829).
- [31] M. A. Islam, R. Srikant, and S. Basu, "Micro-browsing models for search snippets," in *Proc. IEEE 35th Int. Conf. Data Eng. (ICDE)*, Apr. 2019, pp. 1904–1909, doi: [10.1109/ICDE.2019.00206](https://doi.org/10.1109/ICDE.2019.00206).
- [32] B. Li and J. Yao, "Study on unknown term translation mining from Google snippets," *Information*, vol. 10, no. 9, p. 267, Aug. 2019, doi: [10.3390/info10090267](https://doi.org/10.3390/info10090267).
- [33] W. Hop, S. Lachner, F. Frasinicar, and R. De Virgilio, "Automatic Web page annotation with Google Rich Snippets," in *On the Move to Meaningful Internet Systems (Lecture Notes in Computer Science)*, R. Meersman, T. Dillon, and P. Herrero, Eds. Berlin, Germany: Springer, 2010, pp. 957–974.
- [34] T. Steiner, M. Hausenblas, and R. Troncy, "How Google is using linked data today and vision for tomorrow," in *Proc. Linked Data Future Internet Future Internet Assem. (FIA)*, 2010, pp. 1–10.
- [35] A. Khalili and S. Auer, "WYSIWYM authoring of structured content based on Schema.org," in *Web Information Systems Engineering—WISE*, X. Lin, Y. Manolopoulos, D. Srivastava, and G. Huang, Eds. Berlin, Germany: Springer, 2013, pp. 425–438.
- [36] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: Scalable online collaborative filtering," in *Proc. 16th Int. Conf. World Wide Web (WWW)*, 2007, pp. 271–280, doi: [10.1145/1242572.1242610](https://doi.org/10.1145/1242572.1242610).
- [37] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proc. 15th Int. Conf. Intell. Interfaces (IUI)*, 2010, pp. 31–40, doi: [10.1145/1719970.1719976](https://doi.org/10.1145/1719970.1719976).

- [38] J. Calzada and R. Gil, "What do news aggregators do? Evidence from Google News in Spain and Germany," *Mark. Sci.*, vol. 39, no. 1, pp. 134–167, Jan. 2020, doi: [10.1287/mksc.2019.1150](https://doi.org/10.1287/mksc.2019.1150).
- [39] A. Bridy, "The price of closing the 'value gap': How the music industry hacked EU copyright reform," *Vanderbilt J. Entertainment Technol. Law*, vol. 22, pp. 323–358, Jun. 2019, doi: [10.2139/ssrn.3412249](https://doi.org/10.2139/ssrn.3412249).
- [40] M. Potthast, W.-F. Chen, M. Hagen, and B. Stein, "A plan for ancillary copyright: Original snippets," in *Proc. 2nd Int. Workshop Recent Trends News Inf. Retr. (ECIR)*, 2018, pp. 3–5.
- [41] A. Uyar and F. M. Aliyu, "Evaluating search features of Google knowledge graph and bing satori," *Online Inf. Rev.*, vol. 39, no. 2, pp. 197–213, Apr. 2015, doi: [10.1108/OIR-10-2014-0257](https://doi.org/10.1108/OIR-10-2014-0257).
- [42] A. Singhal. (2012). *Introducing the Knowledge Graph: Things, Not Strings*. Official Google Blog. Accessed: Apr. 30, 2020. [Online]. Available: <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>
- [43] K. Juel Vang, "Ethics of Google's knowledge graph: Some considerations," *J. Inf., Commun. Ethics Soc.*, vol. 11, no. 4, pp. 245–260, Nov. 2013, doi: [10.1108/JICES-08-2013-0028](https://doi.org/10.1108/JICES-08-2013-0028).
- [44] C. C. Wakefield, "Achieving position 0: Optimising your content to rank in Google's answer box," *J. Brand Strateg.*, vol. 7, no. 4, pp. 326–336, 2019.
- [45] A. Miklosik, "Search engine marketing strategies: Google answer box-related search visibility factors," in *Handbook of Research on Entrepreneurship and Marketing for Global Reach in the Digital Economy*, L. Carvalho and P. Isaías, Eds. Hershey, PA, USA: IGI Global, 2019, pp. 463–485.
- [46] A. Miklosik, I. Hlavaty, F. Dano, and P. Cervenka, "Google answer box keyword-related analysis a case study," *Eur. J. Sci. Theol.*, vol. 12, no. 5, pp. 185–194, 2016.
- [47] A. Strzelecki and P. Rutecka, "Featured snippets results in Google Web search: An exploratory study," in *Marketing and Smart Technologies*, vol. 167, Á. Rocha, J. L. Reis, M. K. Peter, and Z. Bogdanović, Eds. Singapore: Springer, 2020, pp. 9–18.
- [48] A. Broder, "A taxonomy of Web search," *ACM SIGIR Forum*, vol. 36, no. 2, pp. 3–10, Sep. 2002, doi: [10.1145/792550.792552](https://doi.org/10.1145/792550.792552).
- [49] D. Lewandowski, J. Drechsler, and S. Mach, "Deriving query intents from Web search engine queries," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 9, pp. 1773–1788, Sep. 2012, doi: [10.1002/asi.22706](https://doi.org/10.1002/asi.22706).
- [50] P. Nayak. (2019). *Understanding Searches Better Than Ever Before*. Google. Accessed: Apr. 30, 2020. [Online]. Available: <https://blog.google/products/search/search-language-understanding-bert>
- [51] A. Rajaraman and J. D. Ullman, "Data mining," in *Mining of Massive Datasets*. Cambridge: Cambridge, U.K.: Cambridge Univ. Press, 2011, pp. 1–17.
- [52] C. Zhang, X. Zhang, and H. Wang, "A machine reading comprehension-based approach for featured snippet extraction," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1416–1421, doi: [10.1109/ICDM.2018.00195](https://doi.org/10.1109/ICDM.2018.00195).
- [53] *General Guidelines*, Google, Menlo Park, CA, USA, 2019, pp. 1–166.
- [54] D. Sullivan. (2016). *Google Now Handles at Least 2 Trillion Searches Per Year*. SearchEngineLand. Accessed: Apr. 30, 2020. [Online]. Available: <https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247>
- [55] Statista. (2019). *Worldwide Desktop Market Share of Leading Search Engines From January 2010 to April 2019*. Accessed: Nov. 4, 2019. [Online]. Available: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>

ARTUR STRZELECKI received the Ph.D. degree in management from the University of Economics in Katowice, Poland, in 2013. He is currently an Assistant Professor with the Department of Informatics, University of Economics in Katowice. His research interests include search engines and e-commerce.

PAULINA RUTECKA is currently pursuing the Ph.D. degree with the Faculty of Informatics and Communication, University of Economics in Katowice, Poland.

...