

# KEYWORD ANALYSIS WITH USING STATISTICAL METHODS

Dominika Krasňanská<sup>1</sup>   
Mária Vojtková<sup>2</sup> 

DOI: <https://doi.org/10.31410/ITEMA.2020.251>

---

**Abstract:** *There are currently more than a billion websites worldwide. In so many websites, everyone wants to be visible to search engines through the keywords that people search for. The article deals with the process of creating keywords, through which we can identify the intention of the searcher. The process of creating keywords consists of several steps, namely the collection of keywords, subsequent cleaning of keywords, their categorization and the last step is the interpretation of keywords. The paper focuses mainly on the categorization of keywords, which we obtain through the use of statistical methods, which includes a method of visualizing relationships between keywords by determining the strength of the association between words called concept linking or term map.*

**Keywords:** *Keywords, Keyword categories, Statistical methods.*

---

## INTRODUCTION

The Internet originated in the early 1960s. Initially it served as a communication medium for the military, later it was used by universities to share scientific knowledge, but its use was very difficult. The change came thanks to Tim Bernerser Lee, an employee of the European Organization for Nuclear Research (CERN), who simplified working with the Internet for the average user by creating the World Wide Web in 1990 (Jurčík, 2017).

The World Wide Web (WWW) is the official term for the part of the Internet where information is in the form of web pages called documents. Each document has its own specific address, literally a Uniform Resource Locator (URL), that allows it to be found and displayed in programs called a web browser (Berners Lee, et al., 1994).

At present, the Internet does not only serve to mediate communication and entertainment, but the possibilities of the web space have begun to be used mainly by entrepreneurs, whether to promote products or services, or directly on trading by the web, where they look for potential customers. This procedure can be implemented only in cooperation with the customers themselves (on the Internet, these are visitors to the selected website). It is the visitors who try to satisfy their needs via the Internet by searching for various products and services. Initially, the visitor was only oriented among dozens of websites, which did not present any complications. There are currently more than a billion websites in the world. In so many websites, everyone wants to be visible to search engines through the keywords that people search for. Keywords are terms that people enter into search engines. Based on the keyword, relevant web pages appearing that address the issue for the keyword you entered. Knowing the

---

<sup>1</sup> University of Economics in Bratislava, Faculty of Economic and Informatics, Dolnozemska cesta 1, Slovakia

<sup>2</sup> University of Economics in Bratislava, Faculty of Economic and Informatics, Dolnozemska cesta 1, Slovakia

keywords is especially important for the site owner, because it is through the keywords that visitors come to the site, who can then perform a conversion on the site<sup>3</sup>.

## **THE IMPORTANCE OF KEYWORD ANALYSIS**

Keyword analysis is a way to gain insight into market behavior. This is a detailed analysis of words and phrases that are relevant to the selected area based on the focus of the site. (Binka, 2019).

Keyword analysis makes it easier to get to know your customers. The purpose of keyword analysis is to identify the intent of the searcher through the keywords. This means identifying words or phrases that the visitor enters into the full-text search engine. Performing keyword analysis is important for every site owner. It is the first step in understanding user behavior. However, applying the results found in the keyword analysis is more important, because without incorporating the results, all the work is useless. Only when the keyword analysis findings are incorporated can the web come to the forefront of search results. Keywords need to be incorporated within the site, because it is through keywords that visitors come to the site.

## **PROCESS OF CREATING KEYWORDS**

In the world of the Internet, a keyword is a word or phrase that is entered into a search engine by an Internet visitor, such as Google, as a simplistic term for finding a comprehensive solution. Keywords are the deciding factor in displaying a website's bidding on a search engine. After entering a keyword into a search engine, the search engine will show us the search result, i.e. a list of websites that match the content of the keyword.

The position of each website in the list of search results is determined by the search engine's ranking, so the smaller the position number, the better the website is ranked for that keyword. It is natural that every website strives to achieve the best result in its area of interest. Ideally, it ranks first in search results after entering keywords that are relevant to the industry. However, there is only one first position, so we also consider the position in the top ten results to be a good result. The top ten means that the webpage appears right on the first page of search results in the search engine.

Why is position so important? Because more website traffic means more leads for the site owner, which in turn increases conversions. If a website ranks higher in search results, traffic will be higher. Search engine search analysis speaks for itself. The searcher usually clicks on the link (website) from the first page that the search engine offers him after entering the keyword. In general, therefore, a smaller number of website positions will ensure more traffic. For this reason, knowing the keywords for any website is the starting point for gaining potential visitors. Keyword collection is the first step in keyword analysis.

The second step is to clean up your keywords. In the keyword cleanup process, duplicate keywords need to be removed in the first phase, then zero-search keywords and irrelevant keywords need to be removed. The principle of the third step, keyword categorization, is to

---

<sup>3</sup> The conversion is the required action that the visitor will perform on the website, it can be the execution of the order, registration on the website, filling in the contact form, filling in the survey, registration in the newsletter, etc. What will be considered a conversion on a particular site is set by the site owner individually according to their own business goals. In the e-shop, the conversion in most cases is considered to be an online order, filling in a contact form, filling in a survey, registering for a newsletter, etc.

categorize keywords into the categories we create. In each keyword analysis, the categories differ, depending on the area analyzed. For products, it can be a division of keywords into categories such as price, color, quality, type of product. The individual categories in which we classify keywords should be as specific as possible and clearly distinguishable from the rest of the categories. Keyword categorization is one of the most important parts of the whole analysis process. The reason is precisely that the created categories serve as a basis for creating the information architecture of the entire website. It is the information architecture of the website that significantly influences the position of the website in search results.

The practical part of the paper is mainly focused on the categorization of keywords using statistical methods. The last step in keyword analysis is to interpret the results. That means an explanation of the procedures used and the conclusions we have reached.

### CONCEPT LINKING

Creating a link between concepts and data is one of the most important and difficult steps in the research process (Angot, Milano, 2001). To understand the relationships between words based on the common occurrence of words in a document, there is a graph called concept links or term map (Chakraborty, Pagolu, Garla, 2014). Concept linking or term map is a term for a graph that shows the specific strength of the association of words on the edge with a word in the middle. The association is shown on the graph by lines. The thicker line represents a stronger association. The strength of the association is calculated as follows:

$$strength = \log_e\left(\frac{1}{Prob_k}\right) \quad (1)$$

where,

$$Prob_k = \sum_{r=k}^{r=n} Prob(r) \quad (2)$$

whereas,

$$Prob(r) = \left[ \frac{n!}{[r!(n-r)!] p^r (1-p)^{(n-r)}} \right] \quad (3)$$

$A$  – the word in the middle of the graph

$B$  – the word on the edge of the graph

$n$  = number of documents (nu. of phrases), in which we can find term  $B$

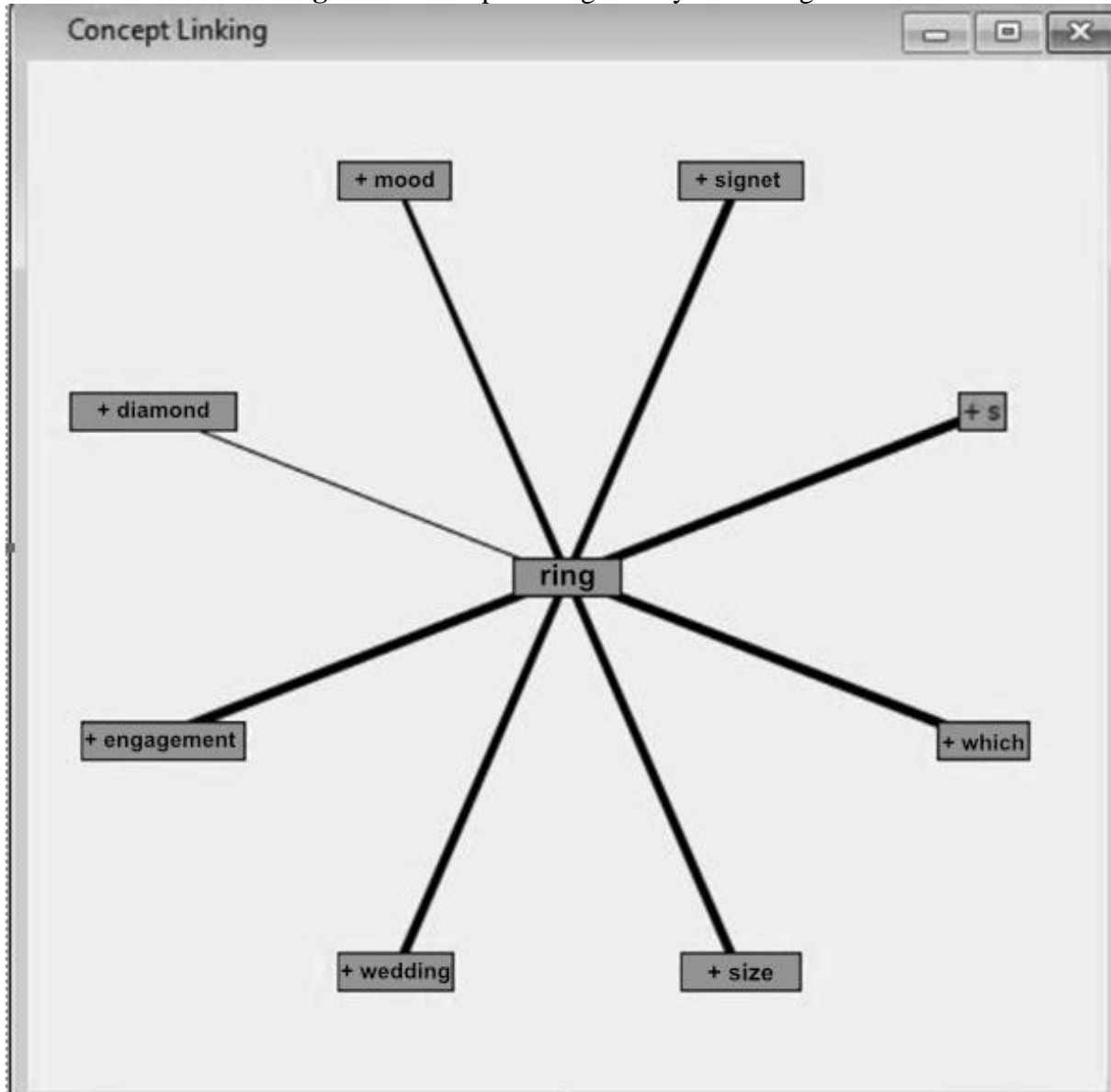
$k$  = number of documents, in which we can find both of terms  $A$  and  $B$

$p = k/n$  = the probability that the expression  $A$  occurs in conjunction with the expression  $B$ , provided that they are independent of each other.

### FUTURE RESEARCH DIRECTIONS

The first step in keyword analysis is to gather keywords. The area of analysis depends on the focus of the sold assortment of the given customer. We decided to analyze the area of jewelry. When collecting keywords, we used various online marketing tools at our disposal, such as: the website of the customer for whom we processed the analysis, the websites of competitors, discussion forums, social networks and various other resources dedicated to the field of jewelry. In the keyword cleanup process, we removed duplicate keywords, deleted zero-search keywords, and irrelevant keywords. The goal of categorization is to categorize your keywords, with the most similar words within one category. We categorized keywords using the statistical software SAS Text Miner. By creating a concept linking graph or term map, which allows the display of word associations, we have reached the following results:

**Figure 1.** Concept linking for keyword "ring"



**Source:** own edit in the SAS Text Miner, 2020

*Figure 1* shows the strength of the association of the word ring with other words. As we can see in the graph, the strongest association of the word ring is with the words engagement, marriage, seal, etc., because there is the thickest line between these words. However, what is not visible from the graph is the number of times the word is in all connections and how many times it is in connection with the word in the middle, i.e. with the word ring. For example, a "diamond ring" is found a total of 23 times. In connection with the word ring only 12 times, so a narrow line is shown on the graph because the association is weak. The term "seal" has a ratio of 10/10 because it is always associated with the word ring. However, the problem is that in the graph, a maximum of about 8 words can be assigned to the word in the middle (default setting), so it always displays only the words that occur most often with the selected word. In this part, we decided to point out one of the possibilities of using statistical methods. This is a clear, visual means of creating categories and subcategories, which we can effectively use, especially in the case of smaller files.

## CONCLUSION

Using a statistical method to visualize the association of keywords through the concept linking graph, we created several categories. It was through categorization of keywords that we created a proposal for an information structure for the website owner. One of the main categories was the Rings category, which contained subcategories such as engagement rings, diamond rings, etc. In the same way as we mentioned above, we also created categories such as Chains, Earrings, etc.

After incorporating the findings from the keyword analysis, the site for which we developed the keyword analysis moved to a better position in the search results, which is the result of each keyword analysis. Better positions in search results will ensure higher conversions for the website, which resulted in an increase in the number of orders and a consequent increase in turnover, respectively profit. Because each website aims to increase conversions, we can say that we've met the goal of keyword analysis for the selected website.

This paper highlights the growing importance of using less traditional statistical methods that focus on textual or unstructured data. The issue of representation and categorization of keywords is currently the subject of much research and its use is very current in various fields.

## ACKNOWLEDGMENT

*This research was supported by the project VEGA 1/0193/20 Impact of spatial spillover effects on innovation activities and development of EU regions.*

## REFERENCES

- Angot, J. & Milano, P. (2001). Linking concepts and data. In Thiétart, R. Doing management research (pp. 132-146). London, GB: SAGE Publications Ltd. doi: 10.4135/9781849208970
- Berners-Lee, T., Cailliau, R., Luotonen, A., Frystyk Nielsen, H. & Secret, A. (1994). The world-wide web. Communications of the ACM, 37.8: 76-82. ISBN: 0002-0782/94/0800.
- Binka, M. (2019). Analýza klíčových slov: Návod krok za krokem. první vydání, Brno, Česká republika.
- Feldman, R. & Sanger, J. (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge, GB: Cambridge University Press. 410 p. ISBN 978-0-521-83657-9.
- Chakraborty, G., Pagolu, M. & Garla, S. (2013). Text mining and analysis: practical methods, examples, and case studies using SAS. North Carolina, USA: SAS Institute Inc., Cary. ISBN 978-1-61290-551-8
- Jurčík, M. (2017). Berners-Lee – Tvorca World Wide Webu, Goldman, Vydavatel'stvo MEDIAGE.