Keywords of Search Engine Optimization based on Corpus

Runtao Li*

Chengdu Neuoft University, Sichuan 611844, China

*Corresponding author e-mail: liruntao@nsu.edu.cn

Abstract. With the development of network, keywords have become an important way of retrieval, which is the window for academic papers to enter circulation and citation. Therefore, the selection of keywords is very important to improve the retrieval rate. Based on this, this paper carries out a corpus based search engine optimization keyword research. Based on one million English corpora, this paper analyzes and studies 3933 key words from the definition and function of keywords, the characteristics of keywords, selection methods, arrangement order, composition, quantity and writing format. It also analyzes the algorithm and ranking principle of search engine, discusses the optimization of website structure, keywords and single page, the punishment of search engine and the rescue method after being punished by search engine. According to the statistical research on the frequency of key words, the number of keywords is mainly 1-2 words, accounting for 90% of the total number. It can be seen that the search engine optimization technology used in the research of keyword optimization in the website can improve the website ranking and realize the value of the website.

Keywords: English Corpus, Key Words, Search Engine, Ranking Principle

1. Introduction

Keywords [1-2] are words or phrases used by users when searching related pages, and also words to be used by search engines in building index tables. Search engine optimization [3-5] is a way to improve the ranking of websites in relevant search engines by using search rules of search engines. It optimizes and updates websites by special ways. Using keywords helps to obtain higher search engine query ranking. It is necessary to pay attention to keyword research to find out the most valuable keywords and make the website key Word in the search engine ranking relatively high, to strive for search customers click into, and so as to improve the site's customer traffic.

Keyword selection is very important; we should pay attention to the process of searching and selecting suitable keywords, which requires careful preliminary research. To discover the terms people use to search and their frequency, and to provide websites with these terms. Keyword selection is a strategic feature that helps identify the most relevant queries to the site. According to the research, find out the words users want to ask, and then customize and optimize the website. Search engine optimization is completed around product keywords, so the selection of keywords is related to the effective completion of the whole search engine optimization work. Keyword strategy mainly includes keyword selection, layout and density [6-7], aiming to improve page relevance. The basic principles of

1

keyword selection are: to be closely related to the website, to find the terms supported by the website content; to have a relatively high search volume, that is, to use the terms that people actually search for; to use the terms with relatively low competition and small amount of search results, the lower the competitiveness of keywords, the greater the chance of getting a good ranking.

Based on the research of keywords in search engine optimization of corpus [8], this paper analyzes search engine algorithm and ranking principle, optimizes website structure, keywords, single page search engine Punishment [9-10] and search engine optimization method. Finally, the optimization method of keywords is described in general. Research shows that the technology of determining search engine ranking for specific keywords can improve the ranking of web search engines and increase the number of customers visiting the website.

2. Keyword Optimization Technology for Search Engine

2.1. Keyword Extraction Algorithm
1. PageRank algorithm

PageRank is mainly designed for network ranking and is usually used to evaluate the effectiveness of website optimization in search engine optimization. Page ranking uses link links in the Internet to identify site rankings. The formula is based on the idea of voting

$$S(V_i) = (1-d) + d * \sum_{j \in In(V_i)} \frac{1}{|out(V_j)|} S(V_j) \tag{1}$$

In this formula, V_i is a web page, V_j is a web page linked to V_i, $S(V_i)$ is a web page's PR value, $out(V_i)$ is a web page, and d is a damping coefficient.

2. Textrank algorithm

Textrank is improved from PageRank, and its formula has many similarities. Here we give the formula of textrank:

$$WS(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V \in out(V_j)} w_{jk}} WS(V_j) \tag{2}$$

It can be seen that the formula only has one more weight term w than PageRank, which is used to indicate that the edge connection between two nodes has different importance.

3. Textrank generation summary

Treat each sentence in the text as a node. If two sentences are the same, the right edge between the nodes corresponding to the two sentences is the weighted edge. The methods to check sentence similarity are as follows:

$$Similarity(S_i, S_j) = \frac{\left|\left\{w_k \mid w_k \in S_i \& w_k \in S_j\right\}\right|}{\log(|S_i|) + \log(|S_j|)} \tag{3}$$

In a formula, S_i and S_j represent two sentences in turn, and then the numerator represents the number of words in both sentences at the same time, and the denominator is the sum of the number of words in the sentence. This specification can omit longer sentences in the same calculation.

2.2. Search Engine Optimization
1. Compatible with W3C. Use W3C standards to design websites, improve user experience and speed up Web downloads.

2. Network framework. Search engine indexing technology has not yet been able to list the frame structure of a web page. If you need to use the framework structure, it is recommended to add noframe in the framework code to improve the friendliness of the search engine.

3. Page size. In order to speed up the loading of the website, CSS and JS can be used to reduce the code from the outside.

4. Usage of flash memory. Instead of designing the entire site as a flash animation, insert it into HTML. The keywords / links in the inner page are stored outside the flash file and displayed as independent text links.

5. Use pictures. Generally speaking, search engines only read the text content, but not the image content. At the same time, image file will slow down the page loading time and increase the sliding load of search engine. Therefore, the core of image optimization has two points: one is to use the alt attribute function of image object to add text to search engine. The second is to compress the size of the image file as much as possible while maintaining the image quality.

6. Index structure. The directory structure is directly in the URL. Clear and concise index structure and standardized name are not only ideal indicators of user experience and website communication, but also ideal indicators for search engine optimization. As for the length, it is suitable to keep 65 English letters, and the page level of URL can be better controlled within 2 layers.

7. Static dynamic web pages. Many websites now have database driven URLs. This is not very good for search engines to skip pages and affect site ranking. Variable URLs are usually converted to static URLs through technical solutions.

3. Experimental Thinking and Design

3.1. Experimental Ideas
Based on one million English corpora, this paper analyzes the key words in the abstract from seven aspects: definition and function of keywords, characteristics of keywords, selection methods, arrangement order, composition, quantity and writing format. And through the analysis of website structure, keywords, single page optimization, search engine punishment, search engine optimization and other methods to elaborate the search engine algorithm and ranking principle.

3.2. Experimental Design
Besides email, search is the most widely used online behavior method. Internet users mainly search for information on search engines to find information and resources online. A survey conducted by a search engine marketing company found that before purchasing products and services, users would search online for products and related services, with 74% searching for products and 54% finding websites to buy.

The influence of website location on search engine ranking is mainly reflected in English websites. For the same English website, the English website abroad is higher than the index ranking in China. The speed of website access not only affects the user experience, but also affects the level of SEO. When the search engine browses the content of a web page, it will determine the ranking of the web page according to the index speed of the web page. The stability of the website affects the huge influence of the search engine. If the website cannot be accessed, it will greatly affect the user experience, and the search engine will ignore it.

4. Discussion

4.1. Research on Keywords of Search Engine Optimization based on Corpus
Table 1. Statistics of compound word frequency

Number of compound words	Frequency	Percent
Two word combination	3783	96%
Three word combination	130	3.3%
Combination of more than 3 words	20	0.7%

The key words are nouns and noun phrases. There are mainly two forms: compound noun and noun phrase. In this paper, the data obtained from 3933 keywords in English corpus are shown in Table 1. The statistical table of compound word frequency in keywords shows that most compound nouns in keywords are composed of two words.

Keyword technology is the core content of network marketing technology, and the keyword based on website is the premise of network marketing activities. Keyword is one of the key elements of SEO optimization, and keyword quality is very important for website. Therefore, in order to establish the main keyword direction in the process of keyword selection, the business goal of search engine must be clear, and the selection of keyword is the first step. Analyze and summarize the key issues to identify a large number of keywords, select the suitable keywords for the site and make appropriate layout. Keyword selection is one of the most important SEO tasks, but it is often lack of discussion and research. Without the right keywords, SEO work will get twice the result. Hun shows a lot of terms that can help consumers find out what to use in search engines. If the web page does not contain keywords to be optimized, the search engine cannot return those keywords for the query. Even if some links from famous and relevant authoritative websites point to the website, if there is appropriate anchor text, then the link is meaningless. The website should be designed and developed in a search engine friendly way. Therefore, the key words are very important and worthy of careful consideration. It is necessary to study how to find the most accurate keywords.
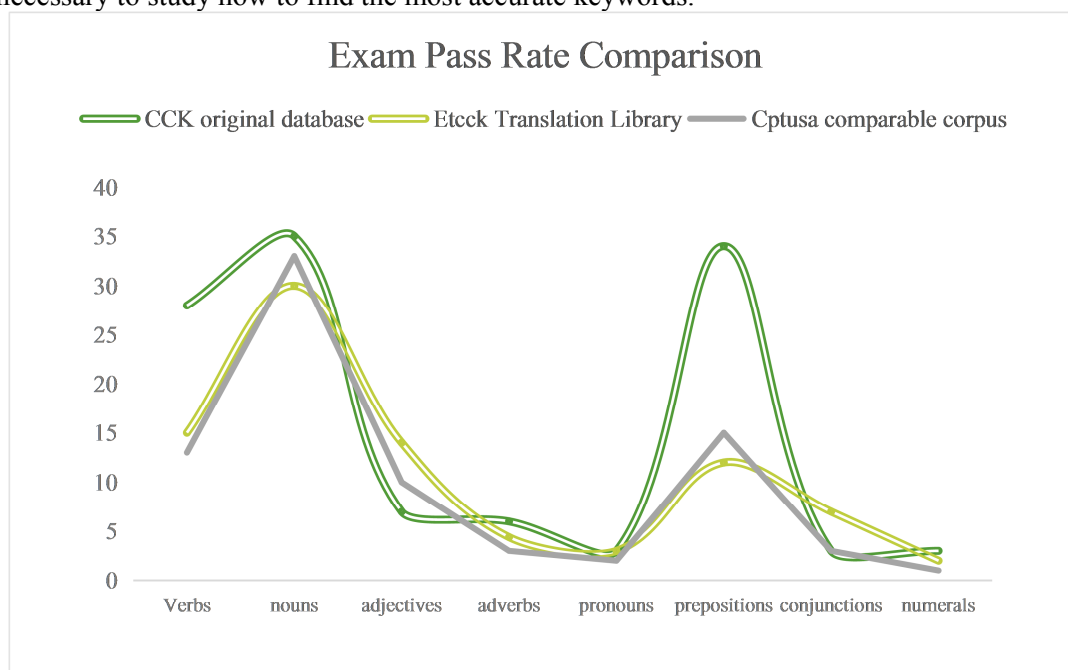

Figure 1. Distribution of parts of speech in CCK, etcck and cptusa corpora

It can be clearly seen from Figure 1 that the parts of speech of etcck and cptusa are similar; however, it is quite different from CCK. It can be seen that the number of verbs used in the original CCK database is much higher than that in the ECCC translation database and cptusa comparison corpus, which corresponds to the linguistic characteristics of English names. On the contrary, the frequency of prepositions used in etcck and cptusa comparable corpora is much higher than that in CCK original corpus.

Statistical table of keyword frequency
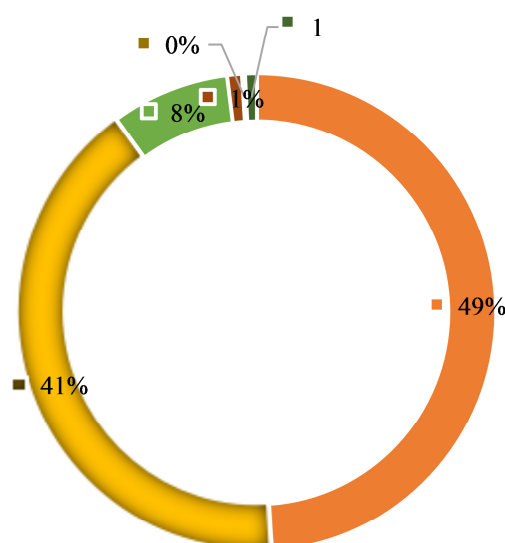
Figure 2. Statistics of keyword frequency

As shown in Figure 2, according to the statistics of keyword frequency, the number of keywords mainly contains 1-2 words, accounting for 90% of the total. More than three words rarely appear. In the end, the key words can be composed of 1-7 words

In a keyword group, each keyword group is separated by a semicolon. In a keyword group, each word is separated by a space.

4.2. Keyword Optimization

Keywords are the foundation of a website. The first thing about keywords is to choose keywords. Choose the right keywords, search engine optimization will become easier, and the website will naturally run faster. Principles for selecting Keywords:

1. Keywords should not be too broad

Common keywords bring a lot of traffic, but the pertinence of these users is not very high, because the purpose of searching common keywords is not very clear, and the user environment may not be in the range of market conversion rate. The order volume is very low, in fact, additional costs have been incurred, but there is no reasonable refund. Factors such as region and brand should be considered. There will be little competitive pressure on these keywords.

2. Don't use too many "cold" keywords

It's easy to get a good ranking by using the "cold" keyword, but the number of known users is very small, so don't use the company name as the main keyword. Even with a certain brand awareness, no one wants to know the company name, and the CEO wants to turn people who don't know the company and its products into customers.

3. Investigate users' search habits

Website designers and operators are well aware of their products and industry characteristics. When choosing keywords, some keywords are very important. Designers will think that users will search, but this is not true. The way of thinking of users and designers is not necessarily the same. Therefore, when choosing keywords, we need to do some research. Choose the most needed keywords, and their competition will be reduced. The most effective keywords are less competitive on the Internet, but they are sought after by users. Therefore, after combining the two, a detailed review should be conducted to list the most effective keywords.

4. Match keywords to the target page

A good keyword target page can meet the needs of visitors, answer their questions, provide them with the services and products they need, and make changes when appropriate. The landing page must contain information related to the search keywords.

5. Conclusions

In the research on the keywords of search engine optimization in corpus, based on the millions of English corpora, this paper further analyzes the definition and function of keywords, the characteristics, selection methods, arrangement order, composition, quantity and writing format of keywords. It also analyzes the algorithm and ranking principle of search engine, discusses the optimization of website structure, keywords and single page, the punishment of search engine and the rescue method after being punished by search engine. The research shows that keyword technology is the core content of network marketing technology, and the keyword based on website is the premise of network marketing activities. Keyword is one of the key points of SEO optimization, the quality of keyword selection is very important for a website, so in the work of keyword selection, we should select the keyword suitable for the website and make appropriate layout arrangement.

References

[1]   Jun, L. I. , Huang, G. , Fan, C. , Sun, Z. , & Zhu, H. . (2019). Key word extraction for short text via word2vec, doc2vec, and textrank. Turkish Journal of Electrical Engineering and Computer ences, 27(3), 1794-1805.

[2]   Purba, E. Y. , Efendi, S. , Sirait, P. , & Sihombing, P. . (2019). Collaboration of rsa algorithm using em2b key with word auto key encryption cryptography method in encryption of sql plaintext database. Journal of Physics: Conference Series, 1230, 012009-.

[3]   Loureiro, S. M. C. , Stylos, N. , & Bellou, V. . (2020). Destination atmospheric cues as key influencers of tourists' word-of-mouth communication: tourist visitation at two mediterranean capital cities. Tourism Recreation Research(4), 1-24.

[4]   Husain, T. , Sani, A. , Ardhiansyah, M. , & Wiliani, N. . (2020). Online shop as an interactive media information society based on search engine optimization (seo). International Journal of Computer Trends and Technology, 68(3), 53-57.

[5]   Bhandari, R. S. , & Bansal, S. . (2019). An analysis between search engine optimization versus social media marketing affecting individual marketer's decision-making behavior. Jindal Journal of Business Research, 8(1), 78-91.

[6]   Gao, R. , & Shah, C. . (2020). Toward creating a fairer ranking in search engine results. Information Processing & Management, 57(1), 102138.

[7]   Setiawan, A. , Harahap, Z. , Syamsuar, D. , & Kunang, Y. N. . (2020). The optimization of website visibility and traffic by implementing search engine optimization (seo) in palembang polytechnic of tourism. CommIT (Communication and Information Technology) Journal, 14(1), 31.

[8]   Karisma, I. A. , Putra, I. N. D. , & Wiranatha, A. S. . (2019). The effects of "search engine optimization" on marketing of diving companies in bali. E-Journal of Tourism, 130.

[9]   Mashayekhi, M. , Analoui, M. , & Bidgoli, B. M. . (2019). The introduction of criteria for assessing an aligned parallel persian-english corpus at the sentence level. International Journal of Engineering and Technology, 11(4), 686-701.

[10]   Othman, A. , & Jemni, M. . (2019). Designing high accuracy statistical machine translation for sign language using parallel corpus: case study english and american sign language. Journal of Information Technology Research, 12(2), 134-158.