

Baidu SEO Ranking Factors Correlation Study 2023

BY
QING HE & MARCUS PENTZEK

Cover image by Midjourney, Prompt:

/imagine an image that combines the ancient Great Wall of China with the surreal aspect of flowing digital information. The Wall should morph into a ribbon of light and data that spirals into the sky. The background should be a starry night that subtly integrates circuit patterns, symbolizing the digital universe.

The rest of the cover design is inspired on the cover of the Baidu SEO Ranking Factors Correlation Study from 2020 with Searchmetrics designed by Jörg Fischer-Bildt (now working at flaschenpost SE).

Table of Contents

A Word from Marcus Pentzek.....	8
A Word from Qing He.....	10
Preface from Kun Tang.....	11
Introduction: Mastering the Art of SEO for Baidu.....	12
Figuring out the Benchmarks.....	13
Understanding the Nuance of Correlation in Ranking Factors.....	14
Distinguishing Correlation from Causation in SEO.....	14
Key Influencers in Search Rankings.....	15
Our Approach.....	16
Simplifying the Interpretation of Spearman Correlation Scores.....	17
Decoding the Data: Understanding Our Charts.....	17
Understanding Our Dataset.....	18
The Enduring Relevance of Web Search in China's Digital Ecosystem.....	19
Key Findings.....	21
Ranking Dominance by Baidu's Own Services.....	21
Domain and URL Configuration.....	21
Refined Techniques in On-Page Optimization.....	22
Advancements in Technical SEO.....	22
The Crucial Role of Backlinks.....	23
The Study Results in Detail.....	24
The Big Players in Baidu's SERPs.....	24
Baidu's own services ranking.....	24
Most prominent ranking Baidu Services.....	27
Top 500 Domains ranking.....	30
Top 10 most powerful players on Baidu SERPs.....	31
Domain Names and URLs.....	35
Top Level Domains.....	35
Top-Level Domains Distribution in the Baidu SERPs.....	35
.com URLs ranking.....	37
Chinese TLDs.....	38
.cn URLs ranking.....	38
.com.cn URLs ranking.....	40
Usage of www. vs. non-www.....	41
Domain Length.....	42
Chinese Characters in Domains.....	43
Numbers in Domains.....	45
URLs.....	47
Chinese Characters in URLs.....	47
Numbers in URLs.....	49
URLs with GET Parameters.....	50
Number of GET Parameters in URLs.....	51
URL Length.....	53

Path Length.....	55
Directory Count.....	57
Chinese Language indicator in URL.....	59
Content.....	61
Counting Characters, not Words.....	61
Keyword Character Coverage.....	63
Title, Meta-Description, Meta-Keywords.....	65
Title Length.....	65
Title equals H1.....	67
H1 is found in Title.....	68
Keyword used in Title.....	69
Position of the Keyword in Title tag.....	71
Keyword Character Coverage in Title.....	72
Keyword used in Title AND H1.....	74
Keyword used in Title AND H1 AND in further Content of the page.....	75
Meta Description Length.....	76
Keyword used in Meta-Description.....	77
Position of the Keyword in Meta Description.....	79
Keyword Character Coverage in Meta Description.....	80
Keyword used in Meta-Keywords.....	82
Keyword Character Coverage in Meta Keywords.....	84
Headlines.....	86
H1 headline.....	86
Number of H1 tags found.....	87
H1 and H2 and ... H6 found all in the document.....	88
Headlines in hierarchical structure.....	89
Keyword used in H1 headline.....	90
Position of Keyword in h1 headline.....	91
Keyword Character Coverage in H1 headline.....	92
H2 headline.....	93
Number of H2 tags found.....	94
H3 headline.....	95
Number of H3 tags found.....	96
H4-6 Headlines.....	97
Number of H4 to H6 tags found.....	98
Overall Average Headline Counts.....	99
Visible Text Analysis.....	100
Content Length.....	100
Chinese Content on the Page.....	103
Average (exact match) Keyword Coverage in Visible Content.....	105
Keyword Density.....	107
Keyword Character Coverage in Visible Content.....	109
First position of the Keyword in Visible content.....	111
Last position of the Keyword in Visible content.....	113
Bi-half Keyword Distribution.....	114

Tertial Keyword Distribution.....	115
Quartile Keyword Distribution.....	116
Keywords in emphasizing tags.....	118
Content Formats.....	119
Ordered Lists.....	119
Number of Ordered Lists.....	121
Number of Points per Ordered List.....	122
Keyword Usage in Ordered Lists.....	123
Unordered Lists.....	124
Number of Unordered Lists.....	126
Number of Points per Unordered List.....	127
Keyword Usage in Unordered Lists.....	128
Tables.....	129
Rows per Table.....	130
Columns per Table.....	131
Keyword usage in Tables.....	132
Images.....	133
Number of Images.....	134
Images with Alt-Attributes.....	135
Keywords in Alt-Attributes.....	136
Keywords in Anchortext (within the ranking page).....	137
Technical Considerations.....	139
https Usage.....	139
HTML file size in kb.....	142
Code Size versus Content Size in kb.....	143
Content-to-Code Ratio.....	144
Average Number of Resources loaded.....	145
Google Tag-Manager Integration.....	147
Adobe Tag-Manager Integration.....	148
Frameset Integration.....	149
Mobile Page referenced.....	151
Canonical Tag Usage.....	153
Self Referencing Rel=Canonical.....	155
Hreflang in Use.....	156
Number of Hreflang References.....	159
<html> Lang-Attribute = Chinese.....	161
<html> Lang-Attribute = English.....	162
<html> Lang-Attribute = any other language.....	163
HTML5 Adaption.....	164
Semantic HTML5 Completeness.....	165
Number of Obsolete HTML5 tags used.....	166
Schema.org.....	167
Publishing Date in JSON-LD.....	170
Publishing Date (or Update date) in JSON-LD is Up-to-Date.....	171
Website Legal Requirements.....	172

ICP License.....	172
ICP License externally links to GOV Website.....	175
Social Media Integration.....	176
Referencing Chinese Social Media.....	176
Referencing Chinese Social Media.....	178
Webmaster Tools.....	180
Chinese Search Engine Webmaster Tools.....	180
Western Search Engines' and SEO Tool Webmaster Tools.....	182
Links.....	184
Outbound links count.....	184
Internal Inbound links count.....	186
URLs with Jumpmark Links.....	188
Number of Internal Jumpmark Links.....	189
“Friendly Links”.....	190
Links Pointing to Non-Chinese Pages.....	192
Backlinks and Domain Strength.....	193
Number of Backlinks.....	194
Number of Referring Domains.....	196
Domain Authoritiveness.....	199
DataForSEO Rank.....	200
Majestic Citation Flow.....	202
Majestic Trust Flow.....	203
DataForSEO Backlinks Spam Score.....	205
About our Data Suppliers (alphabetically).....	207
DataForSEO.....	208
Dragon Metrics.....	209
Majestic SEO.....	210
About JademonD Digital.....	212
Glossary of Terms.....	216
Backlinks.....	216
Baidu-Property.....	216
Correlation Score.....	216
Friendly Links.....	216
Median Baidu Top 20.....	217
Organic Rankings.....	217
Ranking.....	217
Ranking Factor.....	217
SERP (Search Engine Results Page).....	218
TLD (Top Level Domain).....	218
Short Keyword.....	218
Still Short Keyword.....	218
Common Length Keyword.....	218
Long Keyword.....	219
Short Head Keywords.....	219
Mid Tail Keywords.....	219

Long Tail Keywords.....	219
UGC.....	219
URL.....	220

A Word from Marcus Pentzek

**Main Researcher and Author of this Study,
Director SEO at Jademond Digital**



My journey in SEO, especially with Baidu, has been a path of continuous discovery and myth-busting. Starting as the Chief SEO Consultant at Searchmetrics in 2020, I led the first major Baidu SEO ranking factors correlation study. This pioneering work emerged from a need to replace the myriad of speculative blog posts with data-driven insights. These posts, often recycled from outdated sources, propagated myths like the necessity of a Chinese TLD for ranking on Baidu – a misconception I debunked through rigorous analysis.

Now, back in China in 2023, I observe that these myths still echo in the SEO world, misleading clients and necessitating extensive clarification from our end. Recognizing the need for updated, in-depth research, I embarked on an even more comprehensive study. This new endeavor, set for release around Christmas 2023, scales down from 50K to 10K keywords but offers a more balanced and varied analysis, including additional factors and different keyword types. It promises to be a groundbreaking study for SEOs targeting the Chinese market.

What excites me about this study is the collaboration with three renowned SEO data providers: **Dragon Metrics** for Baidu SEO rankings, **DataForSEO** for Bing-China, Google-HK and backlink metrics, and **Majestic** for backlink metrics. This synergy enhances our study's depth and accuracy, making it a reliable resource in the field.

Looking ahead, we plan to conduct further studies on Bing-China and Google-HK in 2024. My goal is to illuminate the complexities of SEO in China, offering precise, nuanced insights, and to continue challenging the entrenched myths in this dynamic domain.

I also want to thank the experts from various companies for their valuable contributions to this study:

- **Adam Di Frisco**, Senior Digital Marketing Strategist, Atigro, www.atigro.com
- **Christina Xu**, Founder of China Marketing Corp, www.chinamarketingcorp.com
- **Dan Taylor**, Editor, ChineseSearchNews.com / Head of Technical SEO at SALT.agency
- **Hermes Ma (马骏)**, Head of Performance Marketing, Merkle, Dentsu, www.dentsu.com/cn
- **Nils Horstmann**, Managing Director, eviom GmbH, www.eviom.com
- **Owain Lloyd-Williams**, Independent SEO Consultant and Owner at Son of Jack Ltd, www.owainlloydwilliams.com
- **Qing He**, the co-author of this study, and Head of Search, JademonD Digital, www.jademonD.com
- **Sarah Presch**, Digital Marketing Director, Dragon Metrics, www.dragonmetrics.com
- **Simon Lesser**, Cofounder and CEO, Dragon Metrics, www.dragonmetrics.com
- **Stephanie Qian**, Head of Search, The Egg Company, www.theegg.com
- **Sylvain Sipp**, SEO | Ecommerce Consultant, linkedin.com/in/sylvain-sipp
- **Tait Lawton**, Founder of Nanjing Marketing Group, nanjingmarketinggroup.com
- **Veronique Duong**, International SEO & Marketing Digital Expert, veronique-duong.com

A Word from Qing He

**Co-Author of this Study,
Head of Search at Jademond Digital**



Since 2021, the Baidu search ecosystem has undergone significant changes;

Excluding Baidu's bidding advertising space, the remaining 10 positions on Baidu's search results top10 are almost all Baidu's own products, such as Baijiahao, Baidu B2B site, Baidu Zhidao, etc., There are fewer and fewer position left for enterprises, and some industries are difficult to squeeze on the top10.

Baidu algorithms continue to adjust, continue to crack down on a lot of seo violations, such as SEP click fast ranking, SEO package fast ranking, etc., SEO thresholds and costs gradually increased, "quick success and instant profit" many operations almost have no effect.

Although the difficulty of SEO optimization is more and more difficult, you can still find some keywords of the enterprise site can be ranked in the top 5, it is important to study the ranking factors of these sites.

In 2023, with the rise of AI, content output has undergone tremendous changes

I have observed that Baidu News Source has cancelled the inclusion of most self media news sources, including Sohu self media, Tencent account, Wangyi account, and even Baidu's own Baijiahao has been affected.

Although Baidu News Source has cancelled the inclusion and weight of many self media platforms, the ranking of official media has been improved. In the search for news sources column, it can be seen that a large amount of content in the information section comes from official media and large websites.

Due to the widespread use of intelligent AI writing such as ChatGPT, many self media accounts generate articles on a large scale, seriously affecting the quality of news sources. Although this approach has increased the visibility of Baijiahao,

the official does not want to gain attention in this way. Under these conditions, there are new challenges in completing the work of Baidu SERP.

Preface from Kun Tang

Founder and CEO of Jademond Digital

Baidu, once a quieter player in China's dynamic internet landscape, experienced a significant shift in September 2021. This change followed a directive from the Ministry of



Industry and Information Technology, which required Chinese internet companies to allow interlinking among each other. I anticipated this would diversify Baidu's search engine results page (SERP), bringing in a broader array of content, including from competitors within the Alibaba and Tencent ecosystems.

Two years later, the evidence has validated my predictions. Our SEO team observed a noticeable shift: B2B company websites began losing their prime search rankings for generic product terms, outpaced by the likes of JD.com and 1688.com. These sites benefited from their robust domain authority, a greater volume of indexed pages, and content that resonated more with users.

Concurrently, our reputation management team noted an emerging presence of Bilibili and Xiaohongshu (LRB) on Baidu's SERP, making it an increasingly significant channel for client brand management.

Our early engagement in this changing market landscape has been advantageous, allowing us to swiftly adapt our SEO strategies. For those who haven't had the opportunity to learn through direct experience, the Baidu ranking factor study is an invaluable resource. It offers detailed insights into the platform's adaptations, driven by regulatory changes, competitive dynamics, or evolving consumer behaviors, all substantiated with data

Introduction: Mastering the Art of SEO for Baidu

In the dynamic world of search engine optimization (SEO) and content marketing, especially for those targeting the Chinese audience via Baidu, understanding the nuances of achieving top rankings is crucial. Baidu has evolved significantly in recent years, enhancing its ability to discern website relevance with remarkable accuracy. This transformation marks an era dominated by Baidu's sophisticated machine learning algorithms, which adeptly match user queries with the most relevant URLs in real time.

The old SEO adage 'Anything Goes' no longer applies to the Chinese market. The days when SEOs could boost rankings on Baidu through sheer volume of link purchases are gone. Baidu now values the quality of links over quantity, aligning with a broader set of factors that determine website ranking.

For SEO professionals focusing on China, it's imperative to shift towards flawless technical SEO, high-quality content creation, and organic link-building strategies. This approach marks a departure from the outdated tactic of acquiring low-quality backlinks. Baidu, like Google, continually rolls out algorithm updates with names like Ice Bucket, Money Plant, Pomegranate, Breeze, Lightning, and Thunder. These updates primarily target backlink vendors, ad-heavy sites, and other specific areas.

Our insights draw upon the findings of the 2020 Searchmetrics' "Baidu Ranking Factors Correlation Study". This study reveals that a website's ranking on Baidu is determined by a plethora of factors or 'signals', each with its own importance. Together, these signals form a complex matrix that governs how websites are ranked on Baidu.

Figuring out the Benchmarks

The journey into Baidu's SEO landscape, initiated with Searchmetrics'/Marcus' 2020 study, has been enlightening. That first foray into Baidu Ranking Factors Correlation Studies revealed a multitude of signals that Baidu considers important, particularly in the context of keyword competitiveness. These insights prompted us to delve deeper into the finer details of these factors.

It's crucial to view this study as a navigational tool rather than a set of unchangeable rules. We're exploring the correlations within Baidu's algorithm, not establishing definitive ranking factors. As we're still in the early stages of this profound research, labeling our findings as 'ranking factors' might be premature.

However, the importance of this document as a benchmark cannot be overstated. It serves as a vital reference for webmasters, digital marketers, and SEO enthusiasts, providing a glimpse into Baidu's evolving patterns and trends. These benchmarks are foundational, offering a baseline for future analyses and comparisons.

We must also recognize Baidu's dynamic nature. As the search engine evolves, it may adjust the importance of certain factors based on topic and user intent.

With this in mind, we invite you to a thorough examination of our findings. We encourage practical applications of these insights and correlations, and look forward to receiving feedback on the resulting SEO outcomes.

Understanding the Nuance of Correlation in Ranking Factors

Ranking factors are the criteria used by search engines to categorize and rank websites in response to a user's query. With an ever-expanding online universe, these algorithms play a critical role in delivering pertinent results. Initially reliant on basic metrics like keyword density, search engines have evolved their ranking factors to be more complex and varied, countering manipulative practices like keyword stuffing.

The exact details and weights of these factors are closely guarded secrets of search engine providers. Revealing these specifications could lead to manipulation, undermining the integrity of search results. Ranking factors today are a mix of technical elements, like backlink counts, and user interaction signals.

For businesses, this evolution presents challenges and opportunities. As search engines refine their ranking criteria, prioritizing genuine quality improvements becomes key to achieving sustainable and organic ranking success.

Distinguishing Correlation from Causation in SEO

Understanding search algorithms and their ranking factors is central to SEO. By examining patterns across top-ranking pages, we can identify factors that seem to influence search rankings. For example, a high occurrence of a specific keyword in title tags among top pages suggests a strong correlation with ranking success.

However, in data analysis, it's crucial to differentiate correlation from causation. While our extensive data analysis might show correlations - like common on-page features among high-ranking pages - this doesn't prove causation. A feature common in top-ranking pages doesn't necessarily cause higher rankings.

Recognizing this distinction is key. While we can identify and analyze correlations, proving causation requires more focused studies. For instance, isolating and examining the impact of social signals on rankings would be necessary to establish causation.

Our data-driven approach lays a solid foundation for hypothesizing about potential ranking influences. We analyze various factors, both technical and qualitative, that may impact rankings. However, it's vital to interpret these findings critically, understanding that correlations provide direction, not definitive causative evidence.

In conclusion, while the term "Ranking Factor" suggests direct impact, our research indicates it's more accurately described as a rank correlation coefficient. This perspective emphasizes the relationship between factors and rankings, rather than direct causation.

Key Influencers in Search Rankings

The ranking process in search engines like Baidu is complex, influenced by various factors beyond simple search queries or keywords. This process is tailored by each user's unique online behavior, leading to personalized search results. Key factors that play a role include:

- **User Authentication:** Being logged into Baidu tailors results based on the user's search history and preferences.
- **Browser Choice:** The use of browsers like Edge, Chrome, or 360-browser affects results, as search engines personalize outcomes based on the technology used.
- **Geographical Context:** The user's IP location significantly alters results to reflect regional preferences and relevancies.

- **Personal Data Usage:** Cookies and other data on a user's device influence results, creating a personalized experience from past online activities.

Baidu, in its continuous evolution, regularly updates the way these factors are integrated to enhance user experience.

This white paper focuses specifically on the variations introduced by Baidu. This approach allows for a more streamlined and neutral analysis, maintaining relevance on a national level.

Our Approach

In advancing from our previous Searchmetrics study, this research expands its scope. We've broadened our analysis from the top 10 rankings, as used in the 2020 study, to the top 20. This expansion allows for a deeper and more nuanced understanding of the ranking dynamics.

Key to our methodology is the exclusion of Baidu.com sub-domains from our dataset. This decision is critical to maintaining objectivity, particularly given Baidu's tendency to favor its own products in organic search results. This exclusion ensures our analysis more accurately reflects the true landscape of search engine results pages (SERPs), including trends like predominant top-level domains (TLDs).

Additionally, our approach tailors its analysis to the unique characteristics of keyword usage. We recognize that the impact of exact match keywords varies across different competition levels and keyword lengths. By moving away from a generic one-size-fits-all approach, our study captures the subtleties in how keywords influence rankings.

To assess the relationship between different ranking factors, we employed the Spearman correlation method. This non-parametric measure is ideal for our

purposes as it evaluates rank correlations without assuming a linear relationship or interval scale measurement between variables. This method's flexibility and appropriateness for ranking factor correlations give us a more accurate and insightful view into how changes in one factor might influence another.

Simplifying the Interpretation of Spearman Correlation Scores

In our study, we've made a crucial adjustment to the Spearman correlation scores for enhanced clarity and ease of understanding. Under standard conditions, a positive correlation, such as between high backlink counts and better rankings, would typically yield a negative Spearman score. This counterintuitive result stems from the fact that superior rankings (like positions 1-3) are represented by smaller numerical values, while they often correspond to larger metrics, such as high backlink counts.

To avoid confusion and ensure that our findings are straightforward for readers to interpret, we've modified the Spearman scores by multiplying them by -1. This alteration reverses the sign of the score, making it more intuitive: now, a positive correlation directly translates into a positive Spearman score. This change simplifies the understanding of the relationship between ranking factors and their impact, ensuring that any significant correlations are immediately apparent and easily comprehensible.

Decoding the Data: Understanding Our Charts

Our charts primarily illustrate data for search result positions 1 through 20. They typically show the proportion of ranking pages that exhibit a particular attribute

under analysis. This approach helps in visualizing how prevalent certain attributes are across the top-ranking pages.

When dealing with actual counts rather than percentages, our default method is to use median values. This choice is deliberate; it helps us avoid the distortion that can occur due to outliers, a common issue when using mean values. Any deviation from using median values in our charts is clearly indicated on the graphs themselves.

This study's strength lies in its extensive dataset, meticulous analytical process, and the application of robust correlation methods. Our goal is to peel back the layers of Baidu's SEO mechanics, providing clear and precise insights into what influences rankings in this complex digital landscape.

Understanding Our Dataset

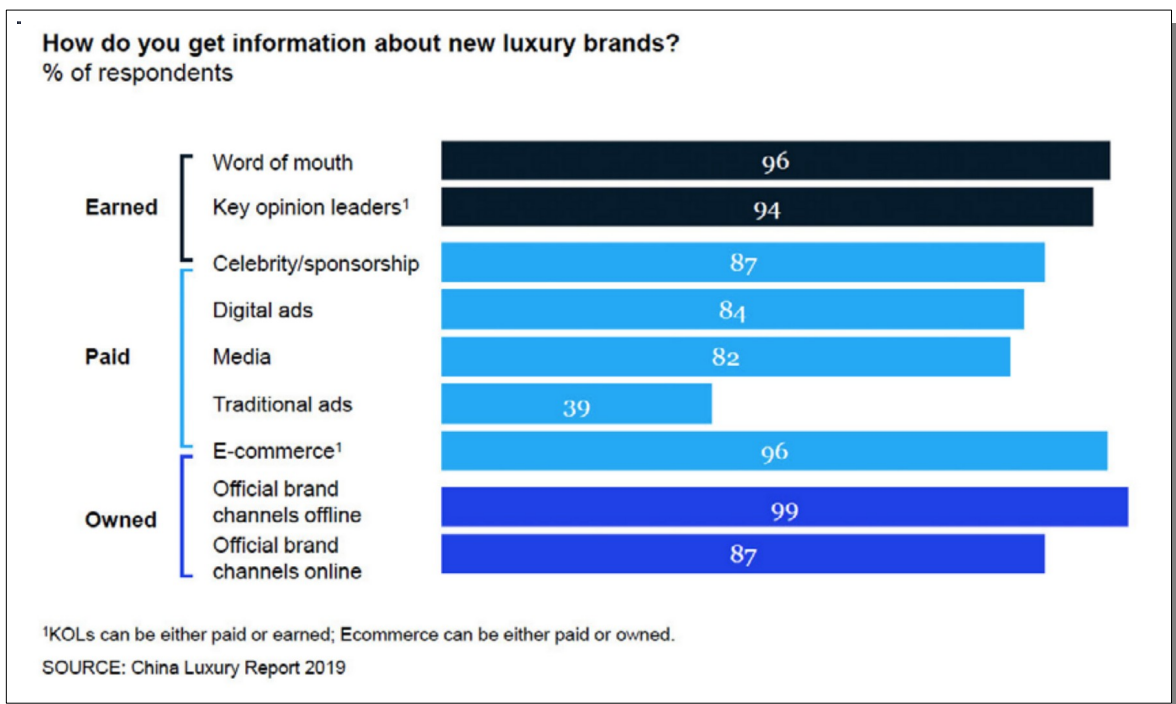
The foundation of our analysis is a diverse dataset of 10,000 keywords, carefully selected from various industries. This comprehensive collection ensures a well-rounded examination of Baidu's SEO ranking factors. These keywords were sourced from our client base, contributions from Chinese SEO experts, and industry classifications provided by Dragon Metrics.

Our dataset primarily consists of keywords in Chinese characters, supplemented by some incorporating Latin letters or numbers. To maintain the dataset's integrity and relevance, we deliberately excluded navigational and brand-specific terms, like "Weibo Login". Such terms often lead to unique, exact-match results, which could potentially distort the overall analysis.

By including a wide range of search volumes, keyword lengths, and intents, our dataset provides a well-balanced view. This approach allows us to capture and analyze the nuances of Baidu's ranking behaviors effectively, offering valuable insights into various search patterns.

The Enduring Relevance of Web Search in China's Digital Ecosystem

Despite the rapid evolution of China's digital landscape, where platforms like Taobao, Tmall, and JD dominate product searches, and WeChat redefines communication and commerce, traditional web search continues to hold a vital position. These modern platforms have indeed become integral, but web search still commands a significant market share. Its effectiveness in driving meaningful traffic and its proficiency in converting queries into potential customers are crucial in the complex customer journey within Chinese markets.



A 2019 McKinsey study sheds light on this dynamic, especially in the context of luxury goods consumers. These consumers expect a diverse information ecosystem. Apart from the popular paid channels and Key Opinion Leaders (KOLs) – a term in China often equated with social media influencers – there is a strong expectation for insight from official brand sources. This includes not only a brand's social media presence but also its dedicated website.

This scenario underscores the sustained importance of web search. While alternative digital platforms play a significant role, the ability of traditional web searches to provide comprehensive brand information and guide consumer decisions remains unmatched. In China's multifaceted digital world, web search thus remains a key player, indispensable for businesses aiming to connect with and convert their target audience effectively.

Key Findings

Our extensive 2023 Baidu SEO Ranking Factors Correlation Study dives deep into the myriad variables influencing search engine rankings. This rigorous analysis is crucial for SEO professionals looking to master Baidu's complex search environment. In this section, we distill our comprehensive research into key data-driven insights, highlighting the most critical findings that will shape SEO strategies for Baidu.

Ranking Dominance by Baidu's Own Services

The research uncovers a pronounced dominance of Baidu's proprietary services in the search engine results pages (SERPs). An extraordinary 34.91% of the top 10 search results come from Baidu's own offerings, demonstrating a clear proprietary advantage.

This trend is even more evident at the top of the SERPs, where Baidu services claim 60% of the first positions, a significant increase from previous years. Services such as Wenku, Baijiahao, Video Search, B2B, Baike, Zhidao, and Tieba are particularly prominent, highlighting the search engine's preference for its own suite of products.

Domain and URL Configuration

The study reveals a notable shift in domain usage trends, with Chinese Top-Level Domains (TLDs) experiencing a substantial increase in Search Engine Results Pages (SERPs) presence, rising from 9.5% to 20.61%. This indicates a growing preference for local domain extensions.

Further analysis shows that over half of the ranked pages fall under the 'www' sub-domain.

Additionally, the average domain length is found to be 15 characters, not counting Baidu's own domains.

Interestingly, the use of Chinese characters in domain names is relatively rare, and preliminary data suggests a potential negative correlation with higher rankings.

Refined Techniques in On-Page Optimization

On-page optimization practices have evolved significantly. Title tags and meta descriptions are now meticulously crafted to the optimal length, ensuring they are fully displayed in Baidu's search results without being truncated. The careful use of exact match keywords in these elements, as well as throughout the content at a density below 1%, exemplifies a sophisticated approach to keyword optimization.

The study also highlights the importance of structured content organization. The effective use of headings, unordered lists, and images plays a crucial role. Particularly noteworthy is the trend of top-ranking pages incorporating target keywords in the alt attributes of images, underlining the significance of every on-page element in influencing search rankings.

Advancements in Technical SEO

The technical SEO terrain is witnessing a significant transformation. A standout trend is the substantial rise in HTTPS adoption, soaring from 55% in 2020 to 69.6% today, mirroring the industry's heightened emphasis on digital security.

Equally noteworthy is the shift away from mobile-specific sites, signaling a broader embrace of responsive design principles. This trend aligns with Baidu's preference for websites that deliver a seamless user experience across various devices.

Another area of interest is the use of canonical tags. While their application is widespread, the study highlights notable inconsistencies, revealing untapped potential for optimization.

Surprisingly, the adoption of Schema.org structured data has seen a significant uptick, despite Baidu's lack of formal recognition for this standard. This indicates a proactive stance among webmasters, keen on leveraging structured data to enhance their sites' clarity and relevance, preparing for future web standards.

The Crucial Role of Backlinks

Backlinks continue to be a pivotal factor in search engine rankings, as demonstrated by our latest analysis. We've found that backlinks from authoritative sources significantly correlate with a better site's position in Baidu's search engine results pages (SERPs). On the flip side, links with a high spam score are more often found with websites that do not have the best rankings.

These insights from our recent study provide a solid empirical basis for developing and refining SEO strategies. For SEO practitioners focused on Baidu, these findings are invaluable. They highlight the importance of not just acquiring backlinks, but ensuring their quality and relevance, guiding the way to improved search engine visibility through strategic and informed actions.

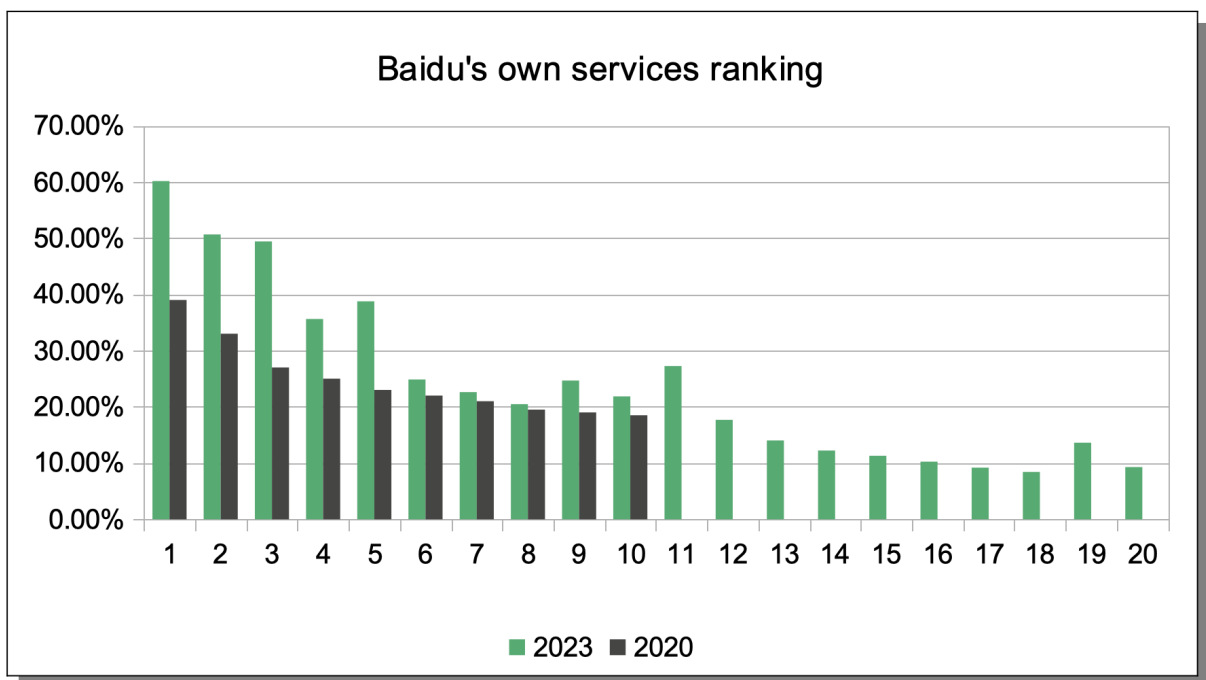
The Study Results in Detail

The Big Players in Baidu's SERPs

In this section, we delve into the dominance of Baidu's own services and the prominent presence of China's leading websites within the search engine results pages. This exploration sheds light on the landscape of Baidu's SERP real estate and its implications for search strategies.

Baidu's own services ranking

Baidu operates an extensive suite of proprietary services, many of which are anchored in user-generated content (UGC). Notable among these are Baike, Zhidao, and Wenku, each serving distinct informational niches within the digital ecosystem.



It's a widely recognized phenomenon within the industry that Baidu exhibits a discernible inclination towards favoring its own services in search rankings. While

these platforms frequently harbor invaluable content, reflecting the collective knowledge of their vast user base, there's an unmistakable strategy at play: Baidu's endeavors to retain users within its cohesive content universe, reinforcing its dominance and ensuring sustained user engagement.

	2023	2022
percentage of baidu results in top 10	34.91%	24.70%
percentage of baidu results in top 20	24.91%	-
percentage of baidu results on pos 1	60.13%	39.00%

The position distribution graph elucidates the percentage representation of Baidu's own services within the Top 20 rankings (encompassing the initial two search results pages) derived from our analysis of 10,000 keywords. This representation is juxtaposed with data from the 2020 study, wherein page 1 specifically delineated the Top 10 rankings.

Expert comments:



“Baidu’s own services do occupy a large proportion of Baidu’s natural search results. However, rather than considering this as a challenge for SEO, it’s better to use these services’ high-authority properties to gain more content exposure for clients. Especially recently, with the rollout of Baidu AI, Baidu’s Baijiahao is empowered and gets the significant visibility capability

on Baidu Search.”

Stephanie Qian, Head of Search, The Egg Company, www.theegg.com



“Baidu is still ranking its own services at the top of the results. This is logic in a way where Baidu wants users to stay within its system and drain all the traffic to their own subdomains. Marketplaces such as Alibaba, Taobao, and so on are not well ranked in Baidu because I had an information from Baidu that they want to keep the traffic on their own websites.”

Veronique Duong, International SEO & Marketing Digital Expert, veronique-duong.com



“Ten years ago already, our B2C clients had little to no chance ranking on Baidu's SERPs by themselves. Backlinks in China are not easy to be collected, as the Internet is so different. So we mostly consisted in Massive off-site content to develop brand awareness across the third-party UGC platforms (Zhihu, Xiaohongshu...) helping our clients get found on these platforms + sometimes in the Baidu SERPs with articles on these platforms. The data shows us, that this strategy is even more important today”

Sylvain Sipp, SEO | Ecommerce Consultant, linkedin.com/in/sylvain-sipp

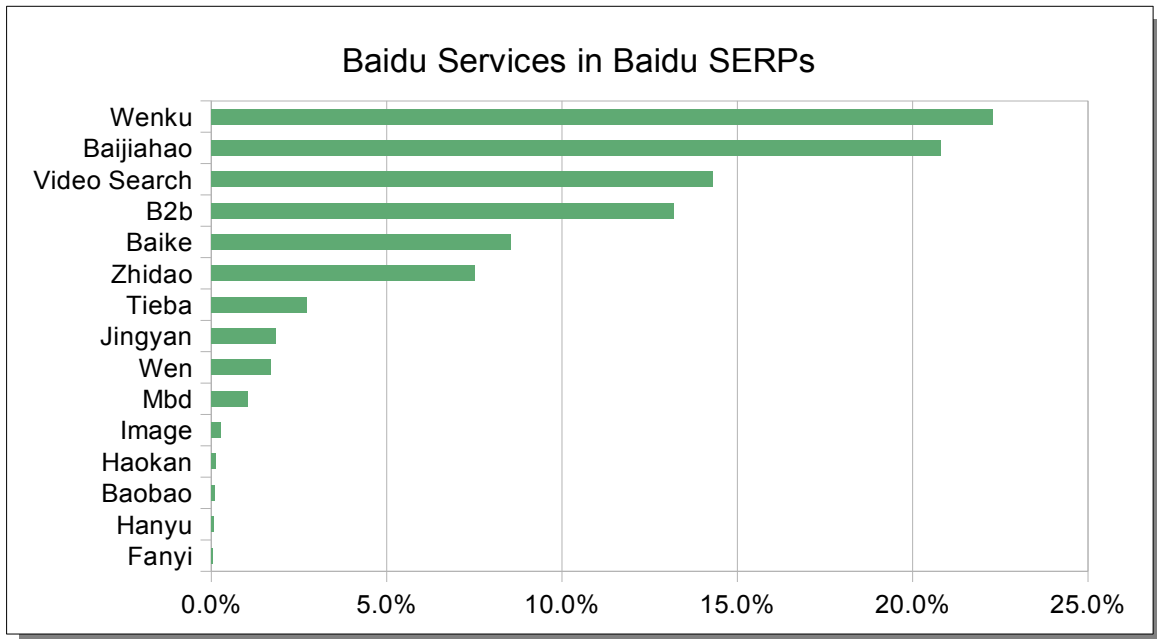


“Baidu Baike takes up less space than I thought it would. Also I wonder about including Zhihu as a 'Baidu service', as I heard Baidu owns some of it. That would result in even more Baidu-owned rankings and less room for regular websites.”

Tait Lawton, Founder of Nanjing Marketing Group, nanjingmarketinggroup.com

Most prominent ranking Baidu Services

Identifying top-ranking Baidu services is crucial for strategic 'parasite SEO'. By contributing high-quality content to these platforms, brands can elevate their Baidu visibility, even if their main site isn't in the top ranks.



In a previous 2020's analysis encompassing a broad spectrum of keywords, Baidu Zhidao, Baidu's UGC (User Generated Content) Q&A platform, manifested a commanding presence, being featured in 70% of keyword results.

However, recent data points towards a paradigm shift in this landscape. Baidu Wenku, the company's eminent file-sharing platform, now occupies 22.29% of the keyword results. In close pursuit is Baijiahao, another UGC-centric blogging platform, claiming a 20.80% stake.

Additionally, Video Search and B2B services have asserted their significance, capturing 14.29% and 13.19% respectively.

Baidu Baike, a platform bearing resemblance to Wikipedia but with stringent moderation mechanisms, is reflected in 8.53% of the keywords. Baidu Tieba, an

online community bearing hallmarks of conventional forums, garners a 2.71% share.

While the details spotlight ten of Baidu's principal services, it's important to acknowledge the extensive range of Baidu's digital portfolio.

Our study identifies 85 distinct services within Baidu's vast digital ambit. Of these, the 75 services beyond the primary ten account for an aggregate 6.09% of the dataset, with each individual service contributing marginally less than 1%.

Expert comments:



“Baidu Baike is quite unknown in Western companies, as Wikipedia's offerings are present here. However, Baike has a high visibility and extremely high trust, as the entries are manually checked by Baidu. Baike should therefore be an integral part of any China SEO strategy. Chinese references are usually required for the entries, e.g. PR publications on trusted third-party portals are suitable for this.”

Nils Horstmann, Managing Director, eviom GmbH, www.eviom.com

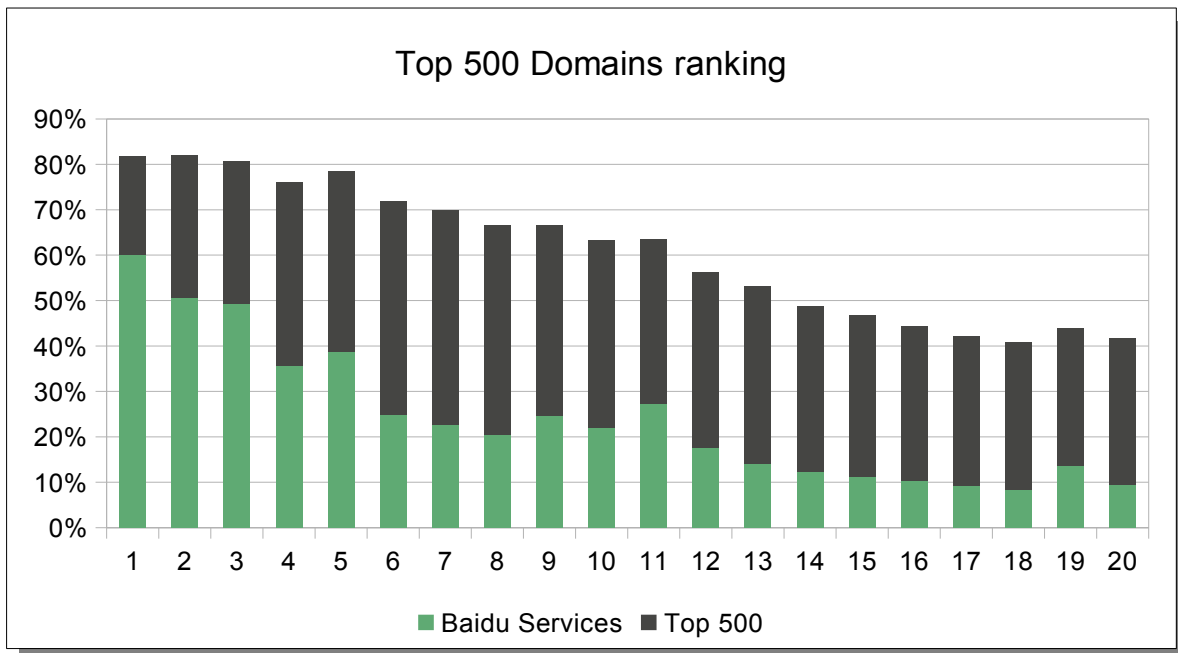


“It’s no surprise that internet giants like Sina, Weibo and Bilibili are dominating the SERPs - Whenever I help clients look through things like organic competitors in Dragon Metrics, websites like Xiaohongshu and jd.com come up all the time. I think this just hits home how much the Chinese search landscape is ruled by these players, and why it’s so vital that international brands have an understanding of what they’re going up against when they’re marketing in China. It’s not just a case of SEOing for Baidu like they’d do with Google, they need to be looking at other Chinese channels like WeChat, Weibo etc. too.”

Sarah Presch, Digital Marketing Director, Dragon Metrics,
www.dragonmetrics.com

Top 500 Domains ranking

The preeminence of influential domains within Baidu's search results prompts an exploration into their potential sway on ranking dynamics.



	Without Baidu	Incl. Baidu
percentage of top500 dom in top 10	38.70%	73.7%
percentage of top500 dom in top 20	36.70%	60.80%

Baidu's SERPs prominently feature its own services, yet a sizable footprint belongs to other major players. When analyzing the top 500 domains based on Searchmetrics SEO Visibility for China, we find that they, alongside Baidu, dominate the search landscape.

At Position 1, Baidu services occupy 60%, with the top 500 domains holding 22%. As we progress to Positions 2 and 3, Baidu's share slightly diminishes to 51% and 49%, while the top domains rise to 31%. From Position 4 onwards, the top

domains consistently surpass Baidu. For instance, by Position 6, they claim 47% of the share, overshadowing Baidu's 25%.

In summary, while Baidu's services are undoubtedly prevalent, the combined force of the top 500 domains remains a significant contender in the SERPs.

Top 10 most powerful players on Baidu SERPs

1. Baidu.com

Ranking themselves with their many (mostly UGC content) service offers like Baiken, Wenku, Baijiahao, ...

2. Sohu.com

Big Chinese internet portal with tons of contents distributed over several topical sub-domains.

3. QQ.com

QQ.com is the official website for Tencent QQ, a renowned instant messaging software service developed by Tencent, a major Chinese tech company. Beyond messaging, QQ.com offers a plethora of services including news, social networking, music, online shopping, movies, and more. Its integrated platforms make it one of the most visited websites in China.

4. Zhihu.com

Recognized as China's counterpart to Quora.com, Zhihu.com stands as the foremost Q&A platform in the nation. Users turn to Zhihu for expert answers, insightful discussions, and knowledge-sharing on a broad array of topics, solidifying its reputation in the Chinese digital landscape.

It is notable, that Baidu owns shares of Zhihu. We can not be certain, if that could have an influence on it's rankings on Baidu search.

5. 163.com

Originating as one of China's earliest email service providers, 163.com has since evolved into a multifaceted portal under the NetEase company.

Today, it offers a wide range of services including news, online gaming, music streaming, and e-commerce. Renowned for its diverse content offerings, 163.com remains a pivotal player in China's digital ecosystem.

6. Weibo.com

Functioning as China's parallel to X (formerly known as twitter), Weibo.com is a leading social media platform facilitating microblogging services. Users can post, share, and comment on short messages, multimedia, and articles. Due to its immense popularity and active user base, Weibo has become a critical channel for news dissemination, celebrity interactions, and public discourse in China.

7. Sina.com.cn

A pioneering digital platform in China, sina.com.cn is the parent portal of Weibo and is operated by SINA Corporation. Initially established as a news and information portal, it has diversified its offerings to include entertainment, finance, and other content categories. With its expansive reach and diverse services, sina.com.cn holds a significant position in China's online landscape.

8. Weather.com.cn

The official website of the China Meteorological Administration, weather.com.cn provides comprehensive weather forecasts, alerts, and related meteorological information for locations across China. Catering to a broad audience, from everyday citizens to professionals in sectors affected by weather conditions, the platform offers timely and accurate data crucial for planning and decision-making.

9. Iqiyi.com

Often likened to Netflix in the West, Iqiyi.com is one of China's premier online streaming platforms. Developed by Baidu, China's leading search engine, Iqiyi offers a vast library of movies, TV shows, animations, and other video content. With both free and premium subscription models.

10. bilibili.com

Bilibili is a distinctive Chinese video-sharing platform, characterized by its emphasis on user-generated content, particularly in areas like anime, comics, and gaming (ACG) culture. Known for its "bullet screen" feature, where user comments stream across videos in real-time, Bilibili fosters an engaged community atmosphere.

Expert comment:



“Obviously each industry is going to have a different landscape and a different set of these big players in the Chinese internet space. In the travel industry SERPs are dominated by Little Red Book and big, trusted online travel agencies like Ctrip. And on desktop, travel queries often trigger travel journal videos on Bilibili.

Understanding your industry and the players in that space, and establishing your organization on those platforms is crucial to visibility on Baidu SERPs.”

Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com



“The goal in SEO should always be the highest possible SERP penetration (Search-Engine-Result-Page), which - in addition to the website - also includes all other touchpoints and Baidu properties. Therefore, also use "gated" content, i.e. content that cannot be indexed by Baidu, e.g. from WeChat, to make it "crawlable" for Baidu and to seed it via Weibo, for example.”

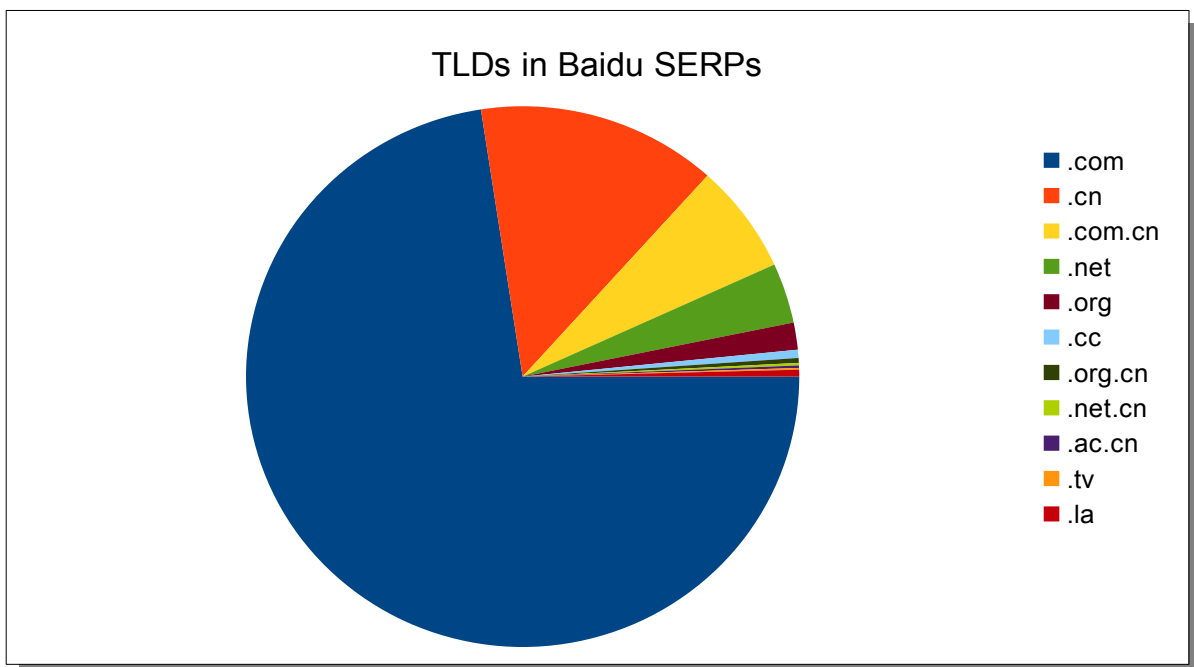
Nils Horstmann, Managing Director, eviom GmbH, www.eviom.com

Domain Names and URLs

Top Level Domains

We are analyzing the nuanced dynamics connecting domains and URLs to their potential rankings in search listings.

Top-Level Domains Distribution in the Baidu SERPs



The .com domain leads Baidu rankings at 72.59%, followed by .cn at 14.06% and .com.cn at 6.55%. Marcus's 2020 study saw .com at 75% and Chinese TLDs combined at 9.36%. The data challenges prior beliefs and highlights shifting domain preferences in Baidu's algorithm.

Searchmetrics' 2020 data positioned .com domains at 75% and combined Chinese TLDs at 9.36%. This latest data refutes the once-held notion of needing a Chinese TLD for Baidu ranking, emphasizing the evolving domain preferences within Baidu's algorithm.

	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
.com	72.59%	69.87%	74.10%	73.36%
.cn	14.06%	15.72%	13.25%	13.50%
.com.cn	6.55%	5.69%	6.05%	7.53%
.net	3.60%	4.03%	3.31%	3.51%
.org	1.61%	2.57%	1.80%	0.78%
.cc	0.49%	0.90%	0.42%	0.27%
.org.cn	0.29%	0.28%	0.28%	0.30%
.net.cn	0.15%	0.13%	0.17%	0.15%
.ac.cn	0.15%	0.05%	0.13%	0.22%
.tv	0.11%	0.07%	0.15%	0.09%
.la	0.08%	0.21%	0.04%	0.02%

Still, the higher percentage of .cn TLD domains of 15.72% on short-head keyword rankings compared to midtail and longtail, and at the same time lower percentage of .com domains ranking (69.87% compared to an average of 72.59%) might emphasize some importance of .cn TLDs for more competitive keywords.

Expert comment:

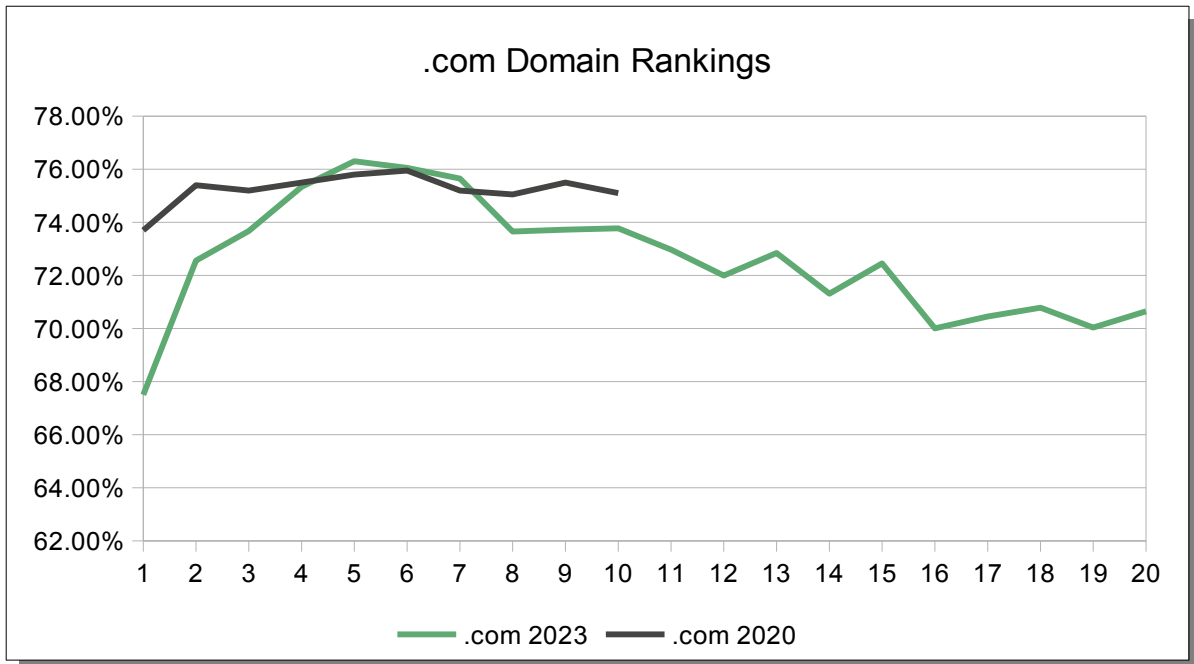


“It’s a common misconception that a .cn domain is required or is influential on Baidu rankings, so it’s useful to see data backing up the fact that there’s no issue in ranking a .com on Baidu. That said, the rise in .cn domains on Baidu causes me to speculate on whether this is due to increased domestic self-reliance of Chinese sites and pride of focusing only on China.”

Simon Lesser, Cofounder and CEO, Dragon Metrics, www.dragonmetrics.com

.com URLs ranking

In the 2020 study, a notable trend emerged: the majority of positions in Baidu's SERPs were dominated by .com domains, with Baidu.com sub-domains excluded from this analysis. This section delves deeper into how that looks 3 years later.



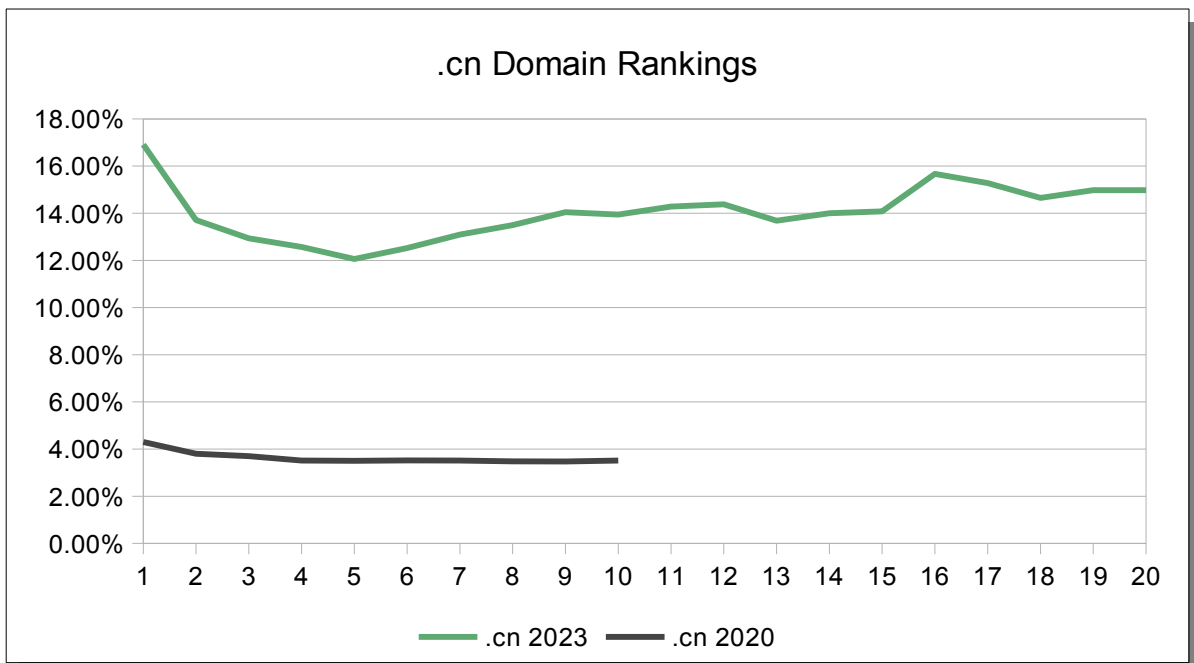
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.55	-0.19	0.46	0.86
percentage of ranking URLs with .com	72.59%	74.10%	74.10%	73.36%

The .com TLD accounts for 72.7% of Baidu rankings, down from 75.5% in 2020. It claims 68% at Position 1 and peaks at 76% between Positions 5 to 7. A positive correlation of 0.55 indicates the .com domain's accepted prominence in ranking practices.

Chinese TLDs

Despite the 2020 study underscoring the dominance of .com domains in Baidu's rankings, persistent myths suggest the necessity of a Chinese TLD for effective Baidu positioning. This section seeks to demystify such claims and further investigates the potential influence of a Chinese TLD on Baidu rankings.

.cn URLs ranking



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.59	0.06	-0.52	-0.77
percentage of ranking URLs with .cn	14.06%	15.72%	13.25%	13.50%

The .cn TLD has risen to a median of 14.02% in Baidu rankings, up from 3.5% in 2020. It holds 17% at Position 1, with positions 2 to 10 averaging around 13-14%.

The correlation of -0.59 highlights the evolving dynamics of .cn's influence in Baidu's algorithm.

Expert comments:



“.cn jump in prominence over the last few years is a major takeaway for websites branching out to mainland China as part of their international strategy. Although in the past we've never thought Baidu favored .cn domains (they seemingly still may not), I always advised organizations to strongly consider the .cn ccTLD. If you're already jumping through all of the various hoops to develop a web strategy for China, you may as well go all-in and signal as much trust with the users and algorithms as possible.”

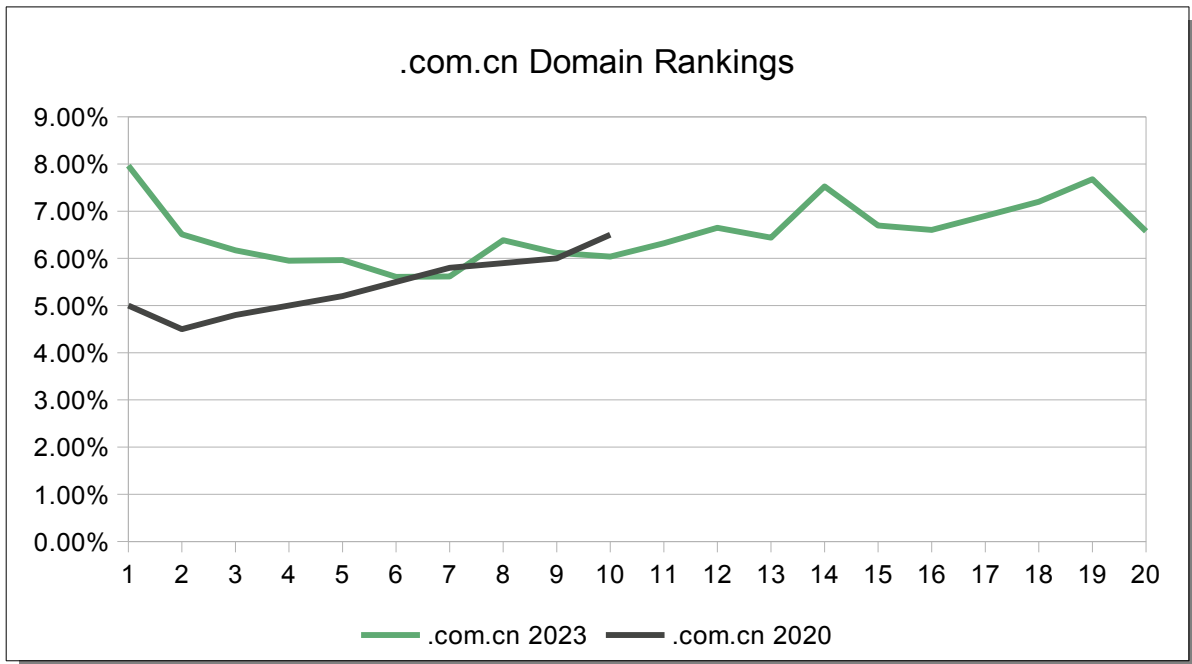
Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com



“I've never held a strong belief that a .cn domain as a ranking factor has a strong weighting (versus other factors), so I always wonder how much of the .cn usage is driven by, what is in my opinion, an antiquated ranking myth.”

Dan Taylor, Editor, ChineseSearchNews.com / Head of Technical SEO at SALT.agency

.com.cn URLs ranking

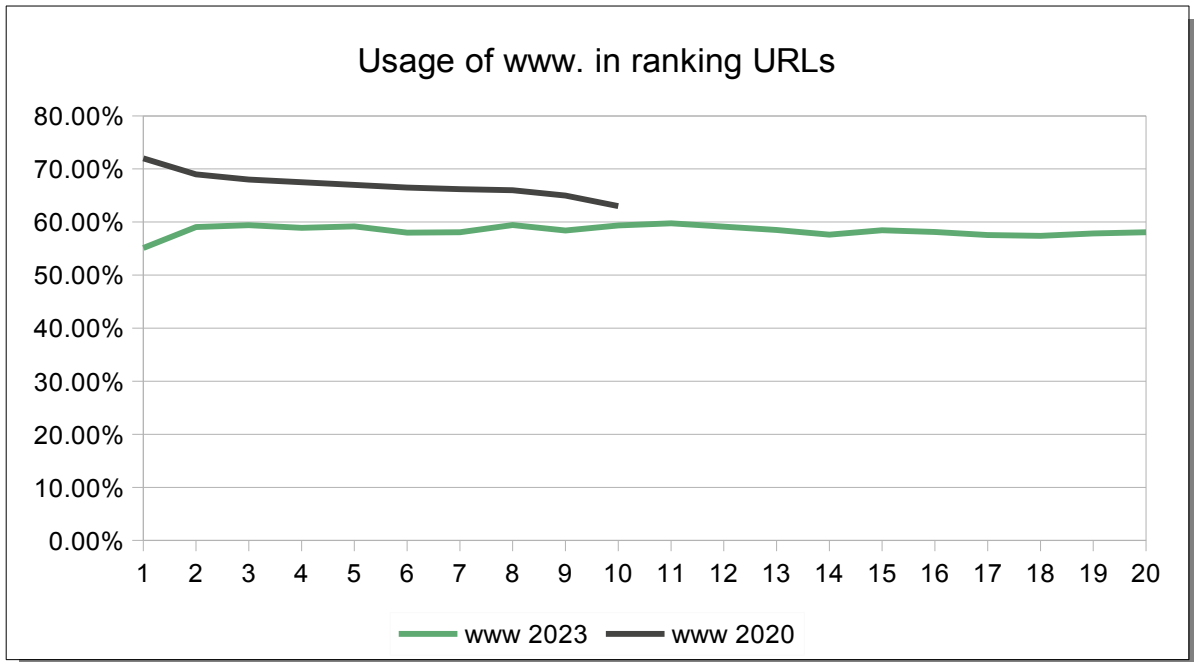


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.59	0.06	-0.52	-0.77
percentage of ranking URLs with .com.cn	6.55%	5.69%	6.05%	7.53%

The .com.cn TLD now appears in 6.48% of Baidu rankings, up from 5.3% in 2020. It captures 8% of Position 1, with positions 2 to 10 averaging between 6-7%. The correlation of -0.51 offers insight into its relationship with Baidu's ranking dynamics.

Usage of www. vs. non-www

Let's look at the potential strength of 'www' URL structures in influencing search rankings.

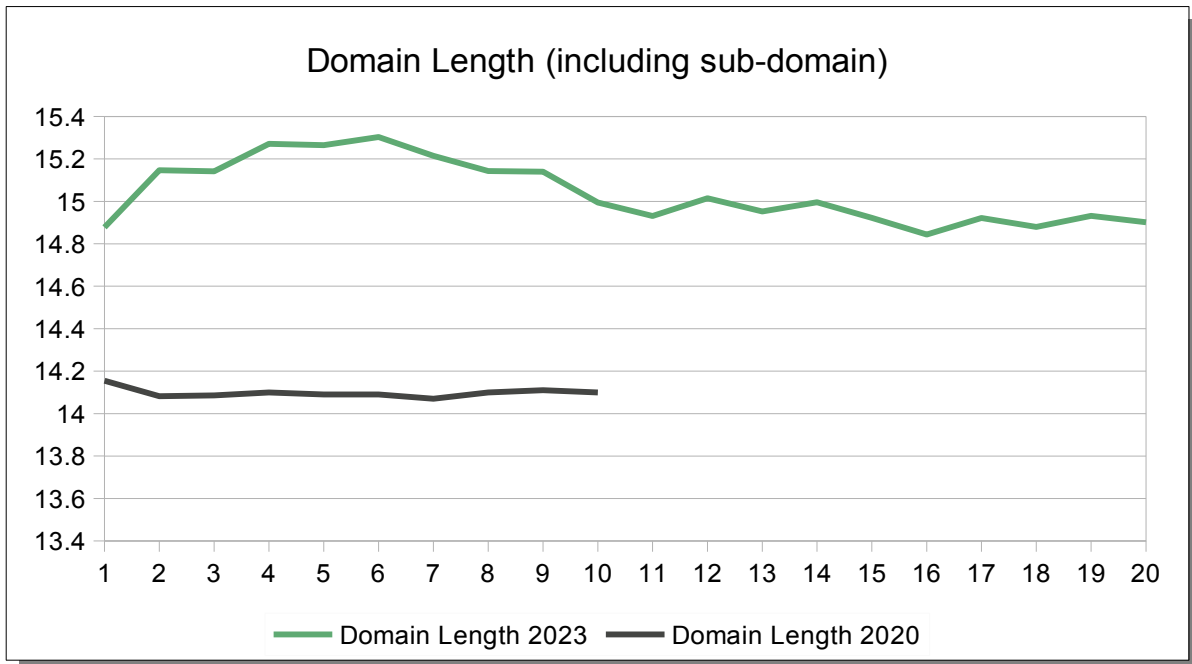


Rankings of URLs with www	
correlation	0.1
www-sub-domain ranking	58.42%

Approximately 55.11% of websites in Position 1 employ the 'www.' prefix, rising to around 59% by Position 10. This marks a decline from 2020's median of 67%. The correlation value of 0.1, down from 0.63 in 2020, suggests a diminished association between 'www.' subdomain usage and top Baidu rankings.

Domain Length

Given the presumption that shorter domains, inclusive of sub-domains, might be more memorable, there exists a conjecture that they could be favored by Baidu's ranking algorithm.

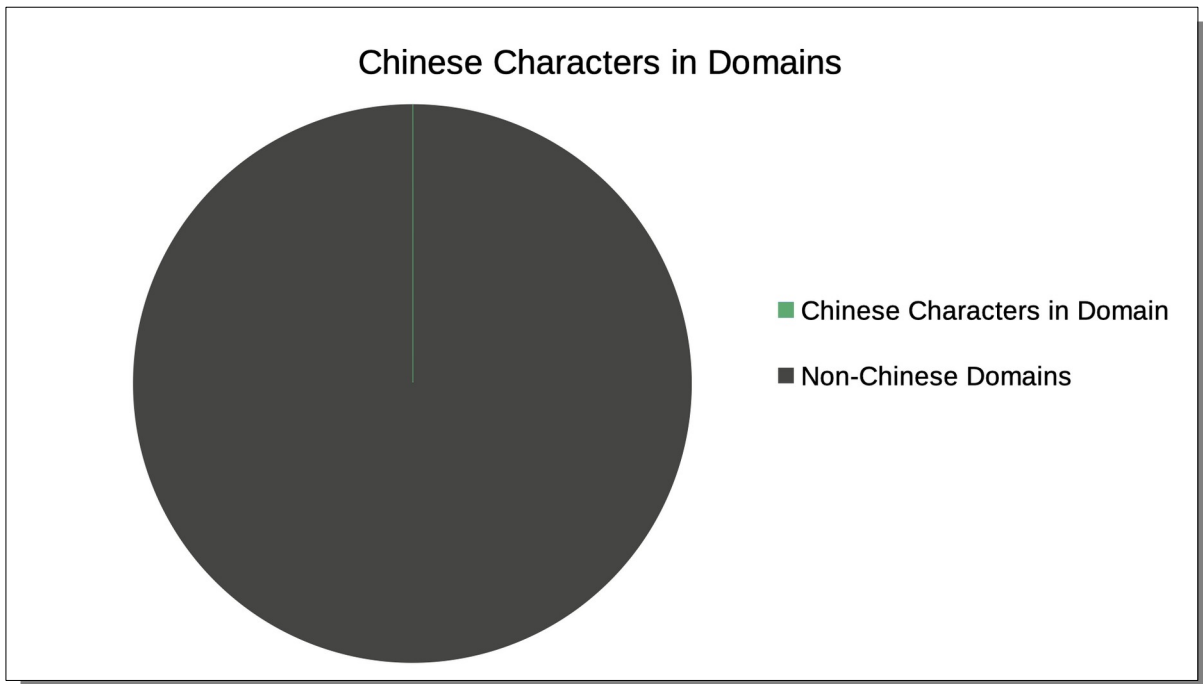


Domain Length (incl. Sub-Domain)	
correlation of shorter domains	-0.04
average (sub)domain length	15.04
median (sub)domain length	14.99

Domains occupying Position 1 average 14.88 letters in length, with a subtle progression peaking at 15.30 letters by Position 6. Throughout the Baidu Top 20, domain lengths oscillate between 14.84 and 15.30 letters. The correlation of -0.04 suggests a marginal association between domain length and ranking outcomes.

Chinese Characters in Domains

Foreign companies often inquire about the significance of Chinese characters in domain names for enhancing search rankings in China. Let's look if that is the case.



A detailed analysis of Baidu's SERPs reveals that only 0.0035% of ranking domains utilize Chinese characters. This minimal representation suggests that Latin-based domain names remain dominant, indicating that Chinese characters in domain names aren't essential for optimal Baidu rankings.

Expert comments:

“Although most websites prefer to create domain names in English, creating domain names in Chinese Pinyin can make your website easier to remember and search for Baidu users.”

Qing He, Head of Search, JademonD Digital,
www.jademonD.com



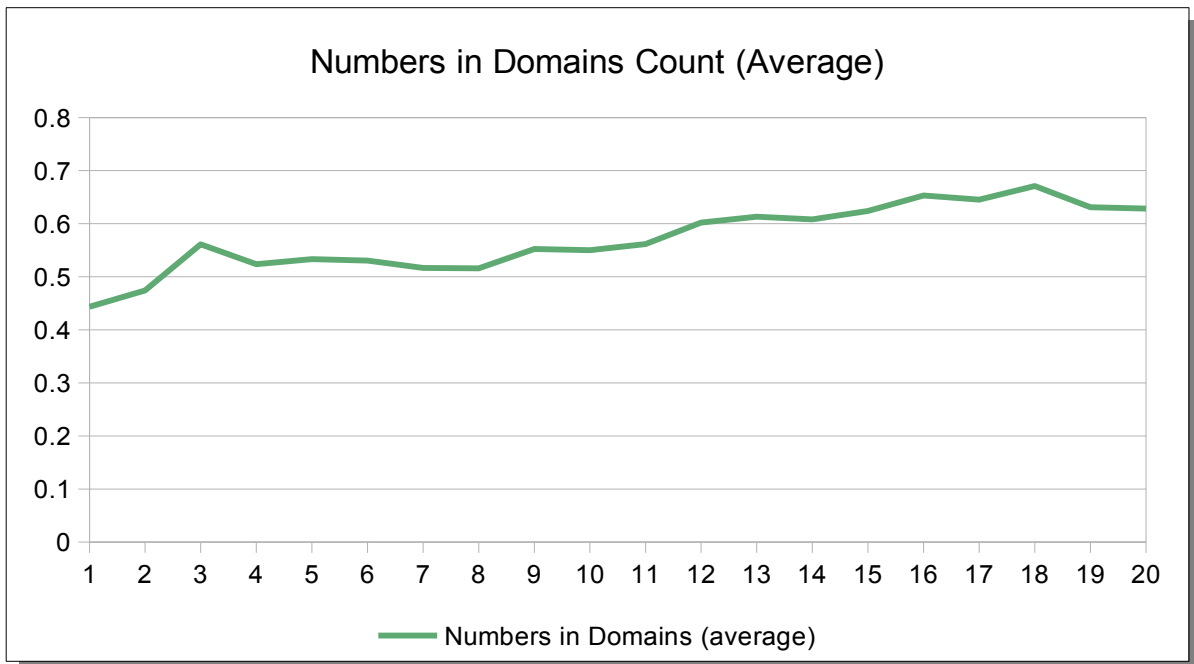


“It is not recommended to incorporate Chinese characters into URLs, as advised by Baidu's official source in 2019. Chinese characters may lead to recognition issues. Despite technological advancements in the last few years, when sharing links, Chinese characters in the URL are still encoded, such as in <https://baidu.com/%E6%B5%8B%E8%AF%95>, which is not user-friendly. Consequently, this practice is not favoured in China.”

Christina Xu, Founder of China Marketing Corp, www.chinamarketingcorp.com

Numbers in Domains

The presence of numbers in domain names has generated debate. This section examines the empirical impact of numerical domains on Baidu rankings. Taking the symbolism of numbers in the Chinese culture and the existence of strong domains like 1688.com leads us to correlate the existence of numbers against better rankings.



As median values would have been zero and we wanted to provide a more granular view, we used average values for this chart.

	Numbers in URL (Path)
correlation of fewer numbers Domains	-0.05
median count of numbers in ranking domain	0
average count of numbers in ranking domain	0.58

Chinese numerical symbolism, evident in domains like 1688.com, suggests a preference for numbers in domains. However, data reveals an average of only

0.58 numbers per domain in Baidu's Top 20. Furthermore, a negative correlation indicates fewer numbers in higher-ranked positions.

Expert comment:



“Although I wouldn’t automatically recommend numbers in domains for Chinese sites, it’s important to remember that numbers not only have lots of lucky/unlucky connotations, but also commonly have double meanings in Chinese. For example “520 - 五二零 - Wǔ’èr líng” is phonetically similar to “Wǒ ài nǐ - I love you”, and “56 - 五六 Wǔliù” sounding like “无聊 - Wúliáo - bored/boring”.

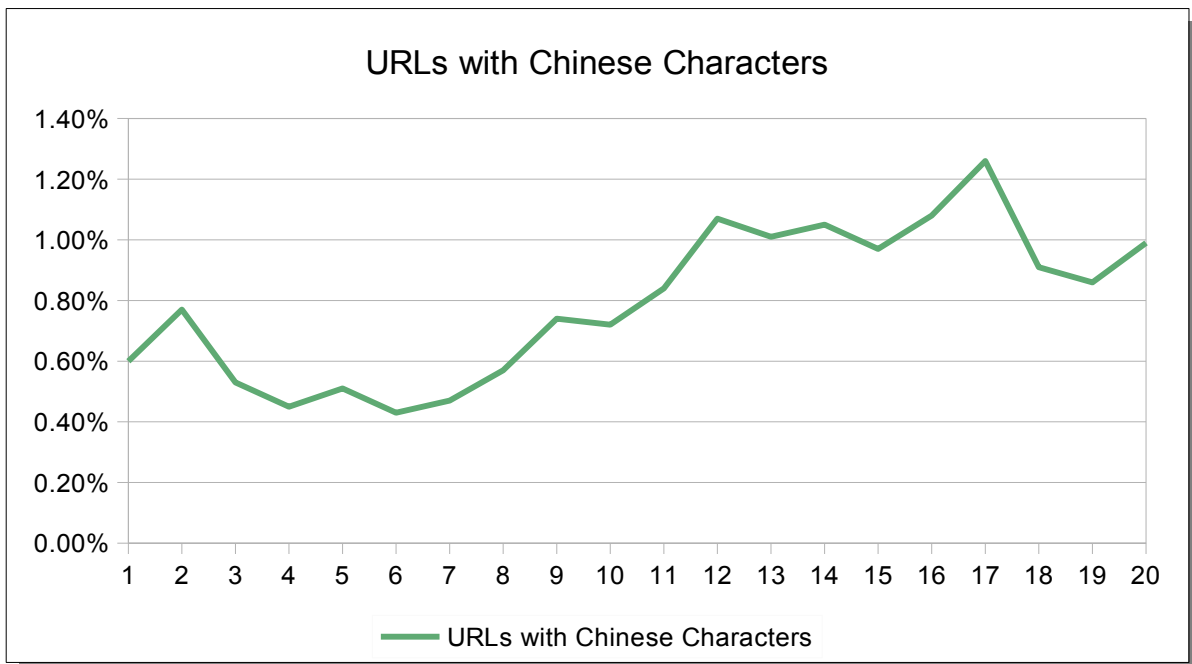
These are definitely worth considering when coming up with a Chinese brand, website, or subsite.”

Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com

URLs

Chinese Characters in URLs

Drawing upon the notion that Baidu may prioritize content cues within URLs, it becomes plausible to consider the potential weightage of Chinese characters in URLs within the Chinese digital landscape.



URLs with Chinese Characters	
correlation	-0.8
percentage ranking URLs	0.8%

Chinese characters appear in a modest 0.8% of ranking URLs on Baidu. A pronounced negative correlation of -0.8 suggests their limited prominence may affect rankings. For the top 10 positions, an average of 0.58% URLs employ Chinese characters, while positions 11-20 see a slight rise to 1%. This data

underscores that, despite Baidu serving a predominantly Chinese audience, the usage of Chinese characters in URLs remains relatively sparse in its top rankings.

Expert comment:

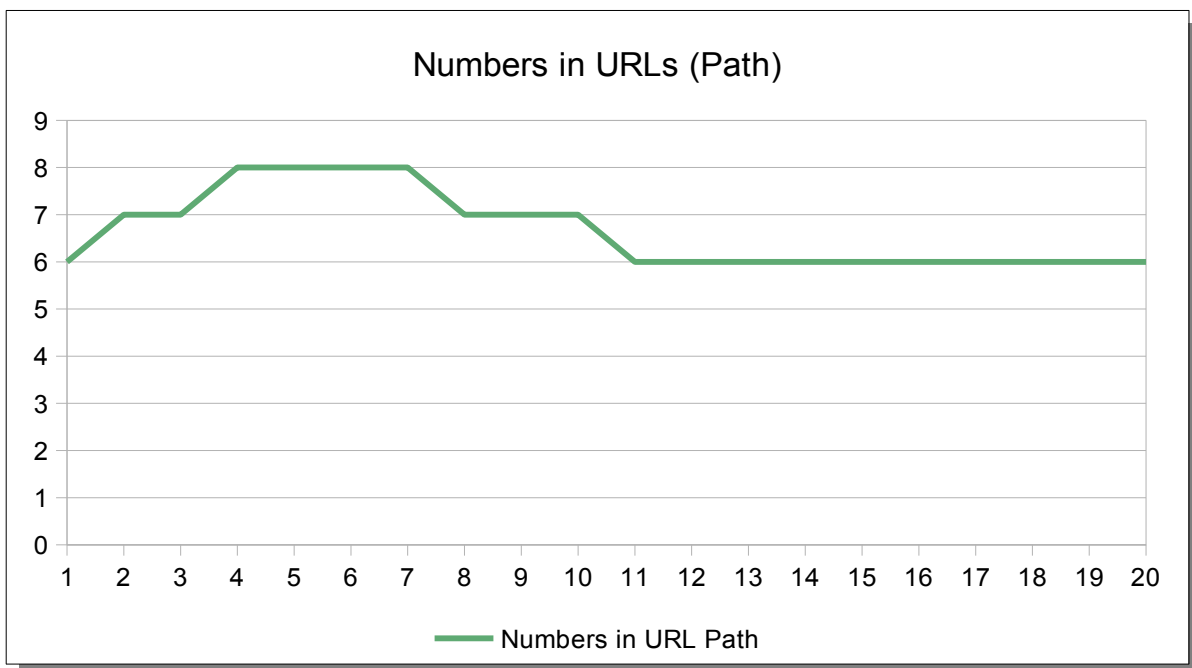


“We understand that global clients might wonder whether Chinese characters in URL could bring potential advantages to their website ranking. However, even Baidu itself has issued official documentation to state that they do not recommend the use of Chinese characters in URLs.”

Stephanie Qian, Head of Search, The Egg Company, www.theegg.com

Numbers in URLs

The presence of numbers within a URL path emerges as a potential ranking determinant. This factor examines whether numeric characters in the URL structure, exclusive of the domain name, influence a page's ranking position on Baidu's search results.

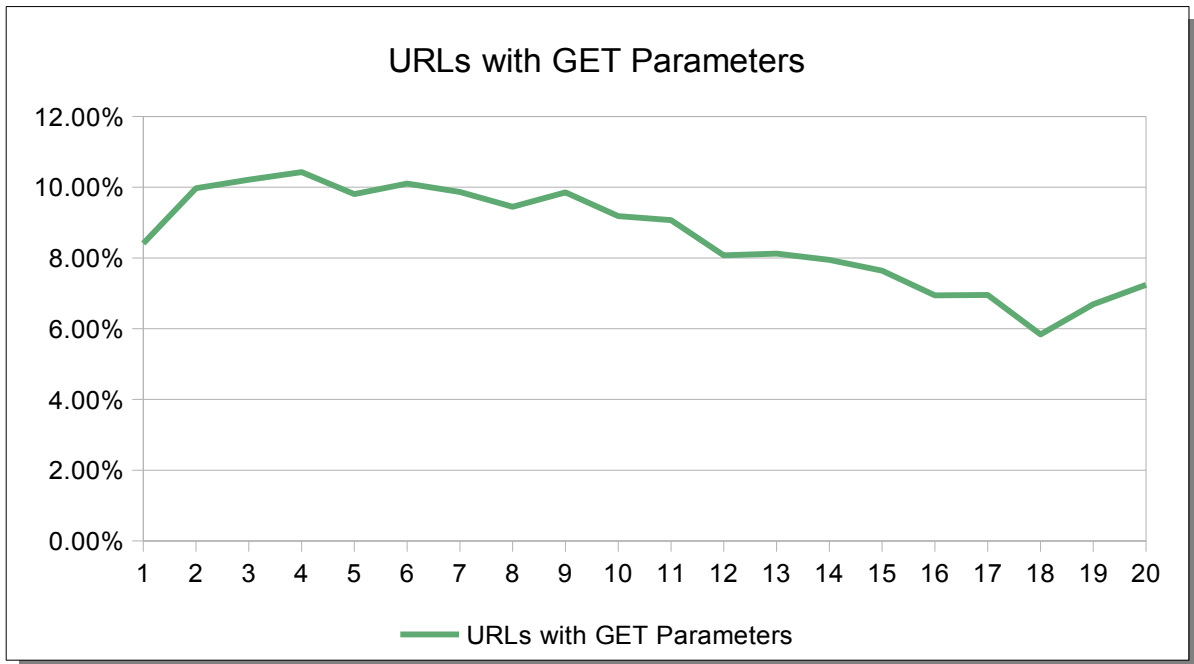


	Numbers in URL (Path)
correlation of fewer numbers URLs	-0.06
median count of numbers in URLs	6
average count of numbers in URLs	8.7

We find an average of 8.7 numbers present within the URL paths. The median stands at 6, with top 10 URLs typically containing 6 to 8 numbers, and positions 11 to 20 maintaining a consistent median of 6. This data effectively dispels the prevailing notion that numbers within URL paths adversely impact Baidu rankings.

URLs with GET Parameters

Can the presence of GET parameters in URLs can impact search engine rankings? This aspect looks at the potential influence of GET parameters in domains on Baidu's ranking algorithms.

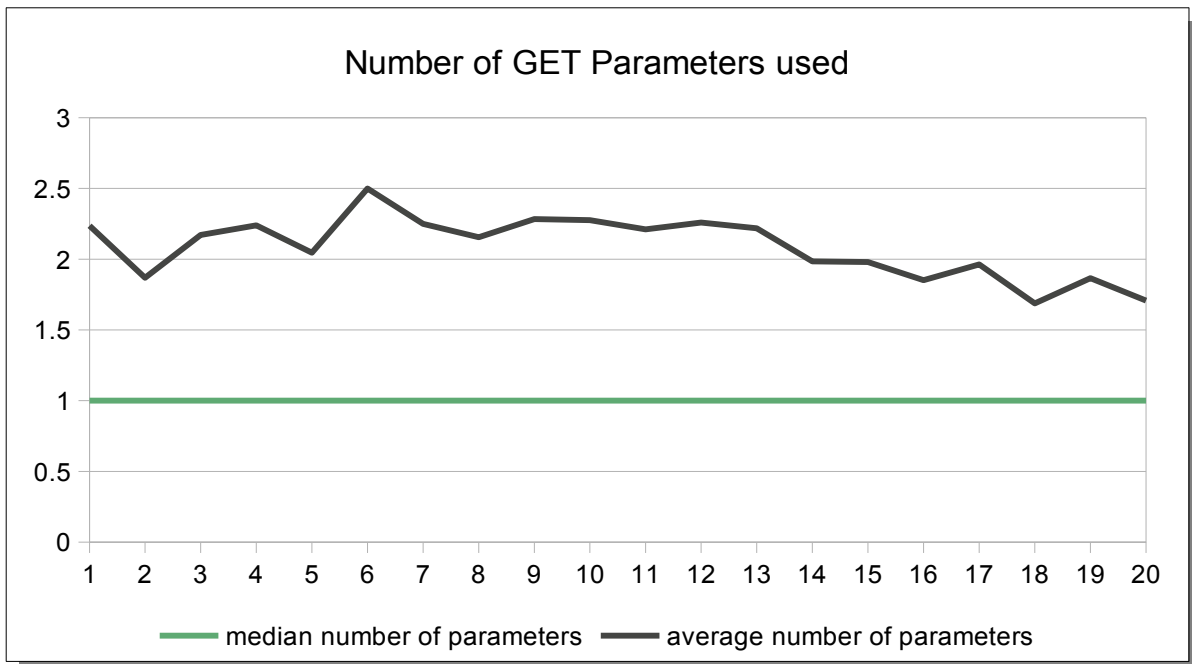


URLs with GET Parameters	
correlation of fewer parameters in URLs	-0.04
URLs with GET parameters	8.54%

Despite the common assumption that GET parameters might clutter URLs, our data indicates a minor negative correlation of -0.043 between the absence of GET parameters and Baidu rankings. On average, 8.54% of URLs in the top 20 positions contain GET parameters. While the top 10 positions see an average presence of around 9.5% for URLs with GET parameters, positions 11 to 20 show a slightly decreasing trend. Thus, the inclusion of GET parameters does not necessarily impede performance on Baidu.

Number of GET Parameters in URLs

Having seen an unexpected correlation between better rankings and the usage of URL GET Parameters, let's examine if the number of parameters, within those URLs containing them, correlates with ranking positions. This further exploration will help determine if the quantity of parameters holds any ranking significance.



As the median values are all 1, we wanted to provide a more granular view, so we also added average values for this chart.

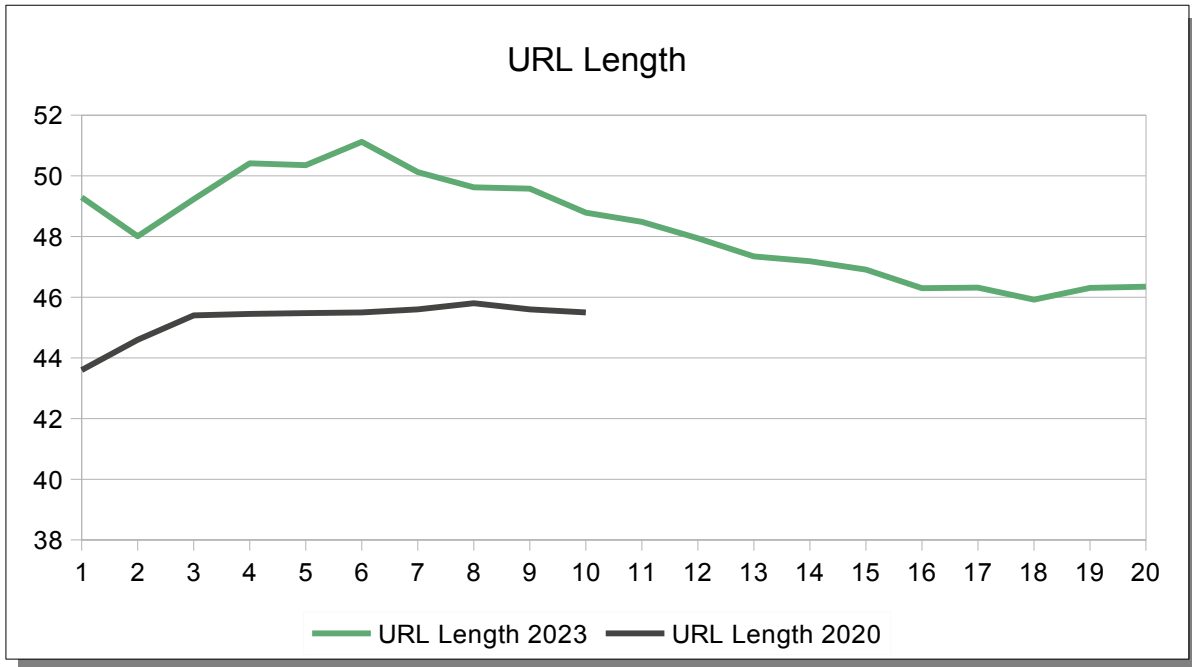
We see a chance that there might be correlations between number of GET Parameters and competitiveness of keywords, so in the following table, we introduce the values per search volume group (short head, midtail or long tail):

	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	0.06	-0.02	-0.09
average no	2.11	1.94	2.19	2.17
median no	1	1	1	1

A nuanced observation emerges from the data: overall, higher-ranked pages tend to feature more GET parameters (IF they have any at all, see previous chart “URLs with GET Parameters”). However, when focusing solely on Short Head keywords, the trend inverts, suggesting fewer parameters correlate with better rankings. This indicates that the Best Practice to not use GET Parameters for important URLs, might be a ranking factor when it comes to more competitive keywords.

URL Length

Given the prevailing hypothesis that succinctness in URLs may be favored, it's worth investigating whether shorter URL lengths exhibit a discernible advantage in Baidu's ranking algorithm.



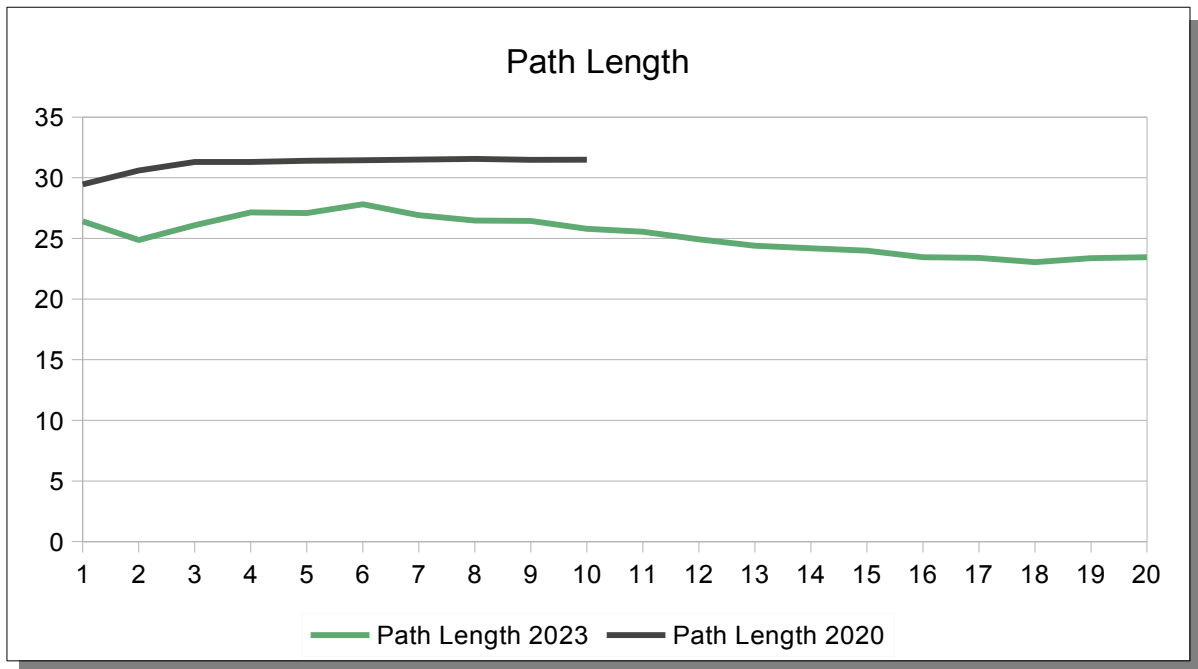
	URL Length
Correlation of shorter URLs	-0.8
median length	48.25

Baidu's ranking algorithm exhibits a clear inclination towards longer URLs, underscored by a pronounced correlation of -0.8. In the current landscape, Position 1 URLs average a length of 49.29 characters, with some positions, notably the fourth, stretching to 50.42 characters, while positions around 15 to 20 average between 46 to 47 characters. Notably, the median URL length has

evolved from the previously recorded 45.46 characters in 2020 to a more substantial 48.25 characters.

Path Length

Considering the prevailing notion that more concise paths may offer a ranking advantage, the correlation between path length and its potential prominence in Baidu's SERPs warrants exploration.



	Path Length
Correlation of shorter paths	-0.8
median length	25.24

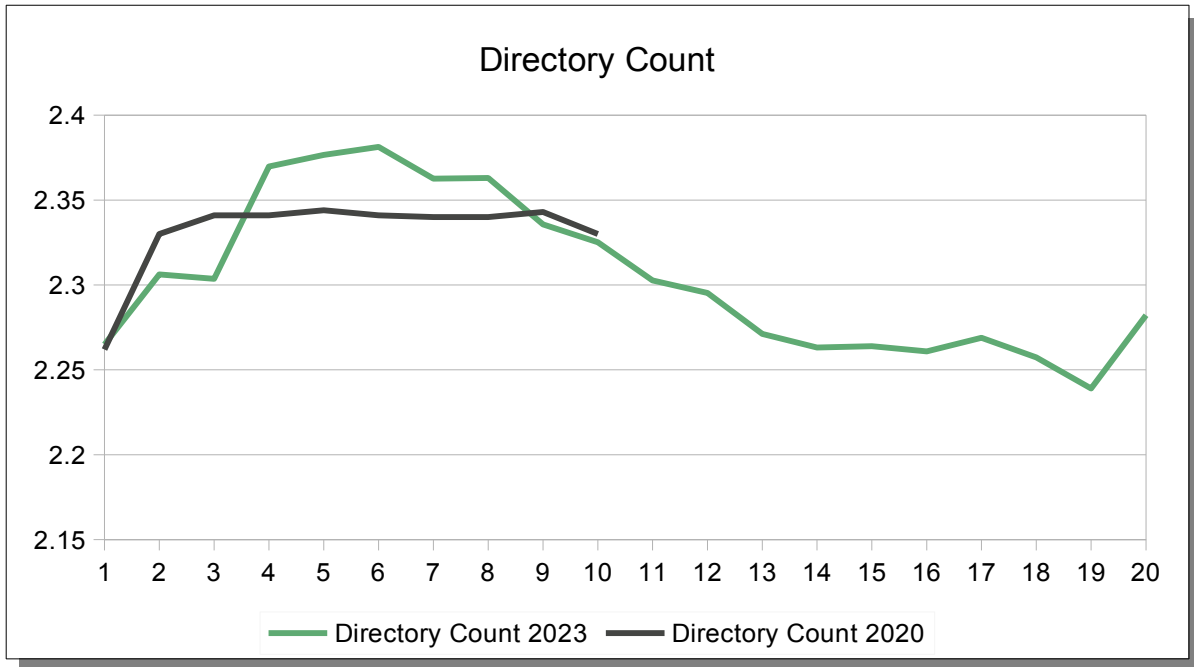
The path length (URL excluding protocol, domain, and sub-domain) presents a strong negative correlation of -0.8, indicating longer URL paths tending to acquire better Baidu rankings.

Position 1 averages a path length of 26.4 characters, with variations across the top 20, mostly hovering around the 25-character mark for Top10 rankings and around 24 characters for rankings between positions 11 and 20.

The median path length has shifted to 25.24 characters, suggesting a trend towards slightly shorter paths than the 29.4 to 31.5 character range observed in 2020.

Directory Count

Given the widespread recommendation by many Baidu SEO experts to maintain content at the root level, the impact of directory count within URLs on search rankings becomes an intriguing area of inquiry.



	Directory Count
Correlation of fewer directories	-0.02
median count	2.3

The directory count within URLs exhibits subtle variances across Baidu's premier positions. With averages ranging from 2.306 to 2.376 directories per URL for the foremost positions, the correlation of -0.023 insinuates a limited influence of directory count on rankings.

Despite prevalent advice from some Chinese SEO experts to anchor pages at the root level, data suggests Baidu's leading pages often encompass 2 to 3 directories.

Consistently, the data reinforces the prevailing tendency for Baidu's Top 10 rankings to favor concise URLs at both the domain and path levels.

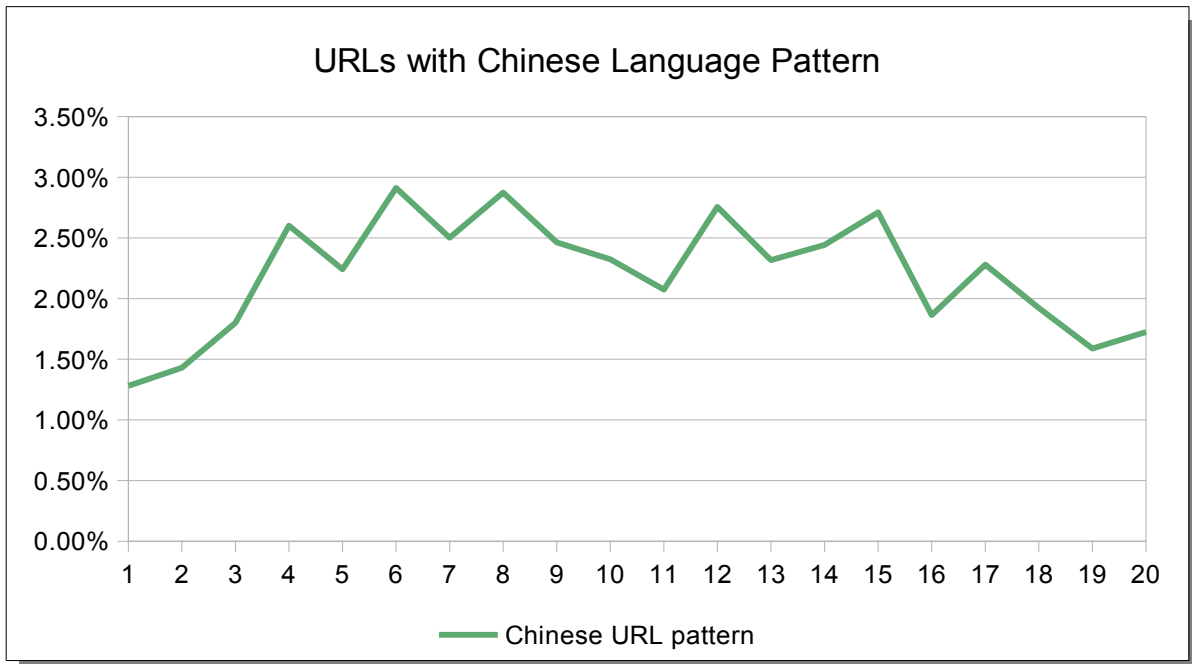
Chinese Language indicator in URL

In light of the influx of Western businesses into China and the apparent Baidu preference for .com domains, this examination evaluates the efficacy of incorporating Chinese sub-folders or sub-domains into global .com websites for better Baidu rankings.

Examples of patterns looked for:

- cn.domain.tld
- zh.domain.tld
- domain.tld/cn/
- domain.tld/zh/
- domain.tld/zh-CN/
- domain.tld/zn-Hans/
- ...

The list is not exhaustive – we looked for many more such patterns, we have regularly (or seldom) witnessed ourselves.



URLs with Chinese Language pattern	
Correlation of fewer directories	0
median percentage	2.3%

In 1% - 3% of the Top 20 ranking URLs can be found URL patterns that indicate that eventually a multilingual website might be ranking with its Chinese path.

There is nearly no correlation (not negative, not positive) of such URLs with better rankings.

Content

Counting Characters, not Words

Analyzing word count for SEO purposes necessitates a distinct approach when dealing with Chinese content. Unlike Indo-European languages, where word delineation is clear, Chinese operates on a character-based system, with each character imbued with inherent meaning, independent of phonetic guidance.

To elucidate, Chinese script is composed of characters aligned seamlessly, reminiscent of pieces in a linear jigsaw puzzle.

Example:

CHINESE Original: 写中文句子不需要空格。

ENGLISH translation: This is an example of a Chinese sentence without spaces.

The conspicuous absence of spaces in most Chinese constructs is evident.

While single words are acknowledged, particularly during the foundational learning stages for children and adult novices, they are rare in formal Chinese texts. On the rare occasions spaces are employed, they demarcate what's termed a 'split sentence'. Given the considerable length disparities between standard and split sentences, utilizing them for SEO-centric word count analyses proves challenging.

With no clear wordings distinguishable in Chinese sentences, what does "Keyword coverage" mean in such a context?

We simply check, if the analyzed keyword is found as a string within the analyzed elements (such as Title tag, Meta-Description, H1 tag, ...).

Expert comment:

“Content has been becoming increasingly important in SEO for years. There is usually a severe lack of content and resources at the Chinese branches of foreign companies, and the localization of corporate content from HQ is often not suitable for China. Artificial intelligence offers a lot of opportunities here, e.g. for Chinese subtitles of YouTube videos that are blocked in China and which are instead played on a Chinese video platform that can be indexed by Baidu, such as bilibili.”

Nils Horstmann, Managing Director, eviom GmbH, www.eviom.com

Keyword Character Coverage

In Chinese linguistic constructs, individual characters often convey distinct meanings while not constituting words in isolation. For instance, the character 林 (pronounced “lín”) signifies “woods”, and 森 (pronounced “sēn”) denotes “forest”. However, the confluence of these characters, 森林 (pronounced “sēnlín”), represents the comprehensive term “forest”. Such amalgamations enhance the semantic depth of single characters.

While “森林” is the definitive representation for “forest”, myriad combinations may allude to similar concepts, such as “the woods”.

Imagine the target keyword to be “森林”, but in the content we rather find words like the following:

- 林地 (lín dì) – woodland
- 林间 (lín jiān) - among the woods; in the forest
- 林下 (lín xià) - under the trees
- 林业 (lín yè) - forestry
- 林冠 (lín guān) - canopy (of a forest)
- 林带 (lín dài) - forest belt or zone
- 林果 (lín guǒ) - fruits of the forest; nuts and berries
- 林栖 (lín qī) - forest dwelling
- 林木 (lín mù) - trees and woods
- 竹林 (zhú lín) - bamboo forest
- 橡林 (xiàng lín) - oak forest

Even though the concrete keyword “森林” might not be included, we still find many words helping to support the meaning of “forest” or woods.

So the pure existence of one character of our target keyword, already helps making the content relevant for the target keyword.

Consequently, we have incorporated “Keyword Character Coverage” into our analytical parameters.

“Keyword Character Coverage” delineates the methodology wherein we assess the presence of each constituent character of a keyword within specific content segments, such as titles or H1 headlines. This metric ranges from 0%, indicating the absence of all keyword characters, to 100%, signifying the inclusion of every character, though not necessarily in exact sequence or match.

Expert comment:

“Most of the time Chinese keywords are much more complicated. The meaning of a 2-characters keyword can change dramatically when add a character. Therefore, it cannot be simply considered that a 3-characters keyword is covered by just 2 characters of the keyword.”



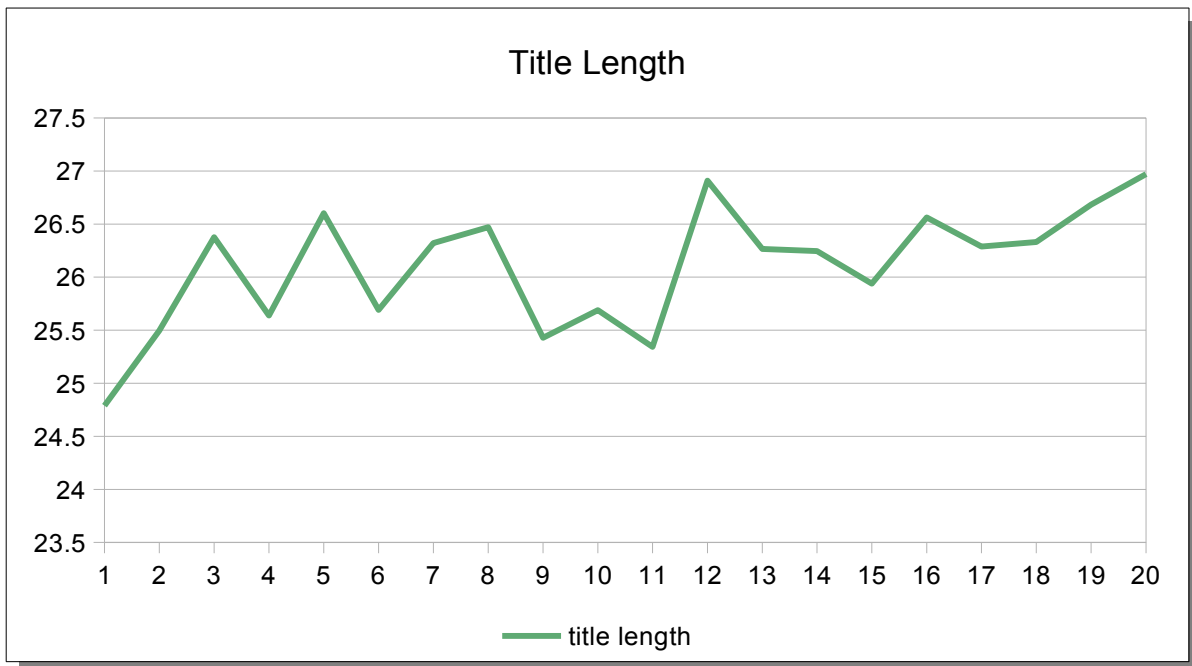
Qing He, Head of Search, Jademond Digital,
www.jademond.com

Title, Meta-Description, Meta-Keywords

In this section, we delve into the examination of Onpage head elements: the title, meta description, and meta keywords. These elements, often fundamental to SEO practices, are assessed for their potential influence on Baidu rankings.

Title Length

The factor "length of title" examines the number of characters present in a page's title tag, considering its potential influence on Baidu's ranking algorithm. This metric assesses whether shorter or longer titles impact a page's position in search results.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.01	-0.01	-0.05
characters	26	24.7	25.4	27.8

Titles on Baidu, on average, span between 25 to 27 characters, aligning closely with Baidu's display limit of 28 characters for desktop search results. Notably, shorter titles correlate with superior rankings, suggesting an advantage in crafting title tags that fit within Baidu's display constraints.

Expert comments:



“The concise script structure of the Chinese language provides a lot of flexibility when it comes to localization and content creation for Chinese websites. The Chinese language is one of the world's best at using a few characters to express extensive meaning. This can be seen with the official Starbucks website, who's title tag in the English website is "Starbucks Coffee Company", and

on the Chinese domain is "星巴克 | 用每一杯咖啡传递星巴克独特的咖啡体验 - Starbucks | Get Starbucks' unique coffee experience with every cup of coffee", which in english takes up 80 characters, well above the character max on Google.”

Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com

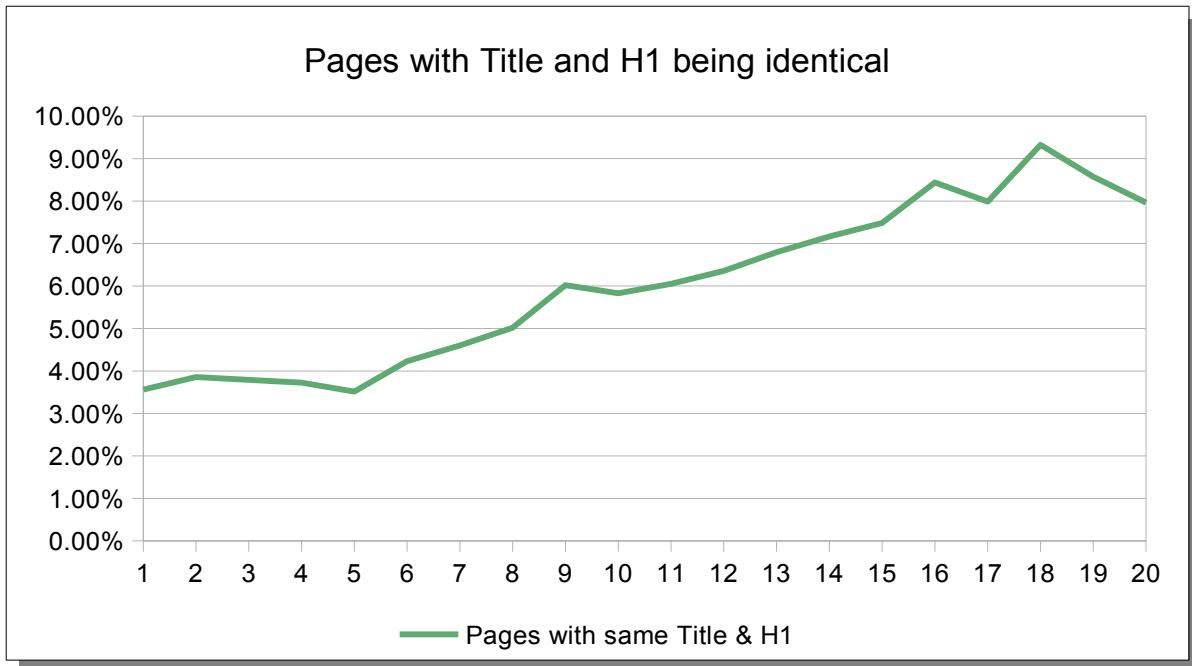
“Baidu tends to provide good rankings to pages including keywords in the first few words of the Title. But in the current situation of increasingly fierce competition, the often seen practice of piling up a large number of popular keywords in the title is completely ineffective.”

Qing He, Head of Search, JademonD Digital, www.jademonD.com



Title equals H1

The alignment of the title tag with the H1 header can suggest missed opportunities in keyword diversity. Let's check how the current top ranking pages are doing.



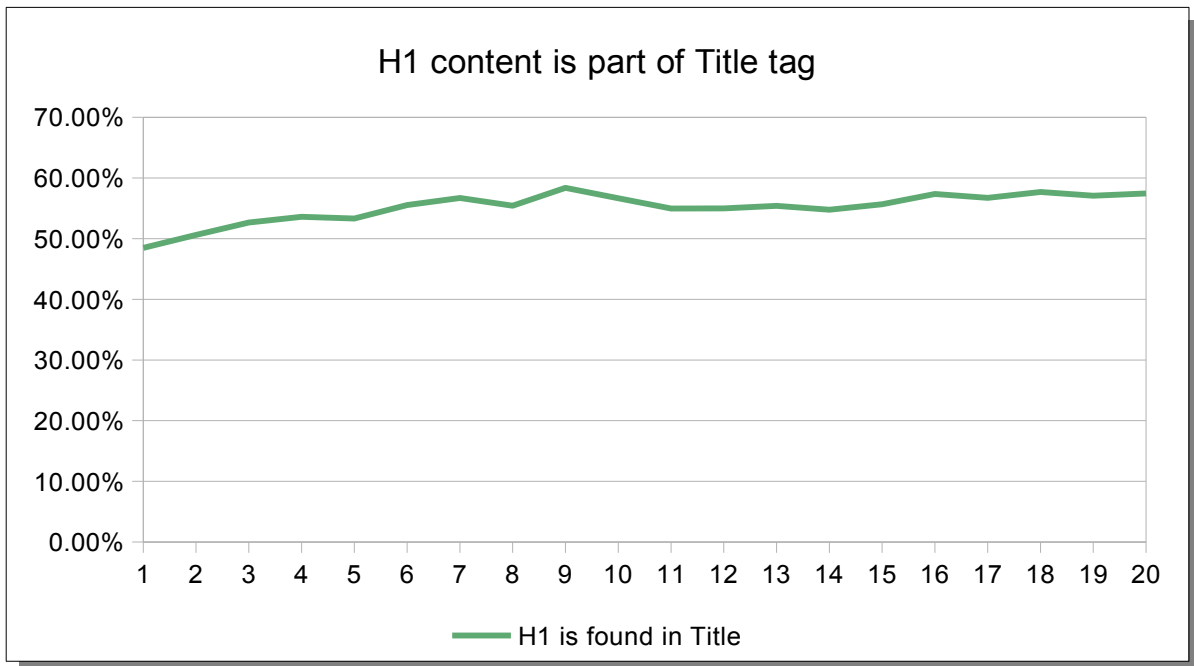
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.07	-0.09	-0.07	-0.06
percentage	6.2%	5.1%	6.5%	6.7%

A slightly negative correlation exists between identical title and H1 tags and superior Baidu rankings. While this may not directly influence rankings, it suggests that higher-ranking pages often employ greater keyword diversity between the title and H1.

Alternatively, it could indicate that as pages achieve better rankings, their titles deviate from the H1, possibly due to the inclusion of brand slugs in the title.

H1 is found in Title

The practice of incorporating the H1 within the title tag, often with an added brand name, may hinder keyword diversification opportunities. While not identical, this overlap could limit optimal keyword targeting. Our analysis will examine the extent to which the H1 is present within the title tag.

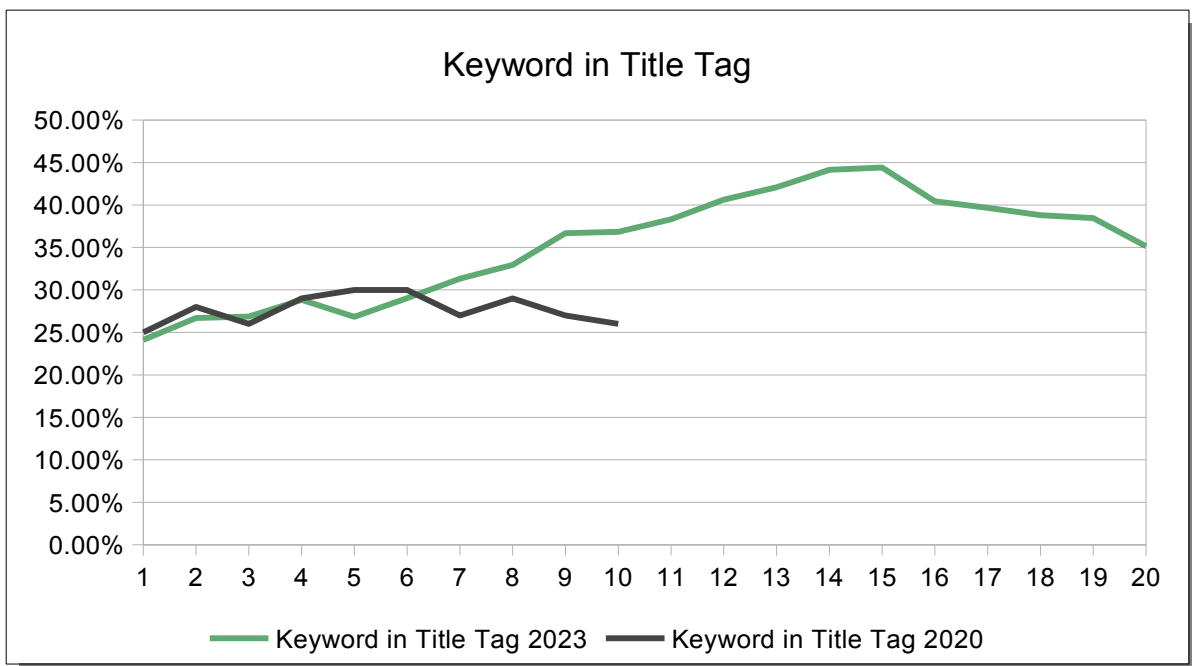


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	-0.09	-0.03	0.03
percentage	55.5%	51.9%	57.4%	56.5%

A notable 55.5% of top-ranking pages appear to employ the strategy of appending their brand as a slug to the H1 to form the title tag. However, a slight negative correlation is observed, suggesting that pages with superior rankings might be leveraging greater keyword diversity between their title and H1 tags.

Keyword used in Title

The presence of the target keyword in the title tag is a fundamental consideration in SEO, given its prominence in search results and potential influence on click-through rates and relevance to search queries. This factor examines the correlation between the inclusion of the exact match keyword in the title tag and its impact on Baidu rankings.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.1	-0.17	-0.14	-0.02
percentage	36%	54.4%	41.7%	18.6%

The data suggests a negative correlation between exact match keyword usage in the title and rankings. However, shorthead keywords are more often used in exact match than long-tail keywords.

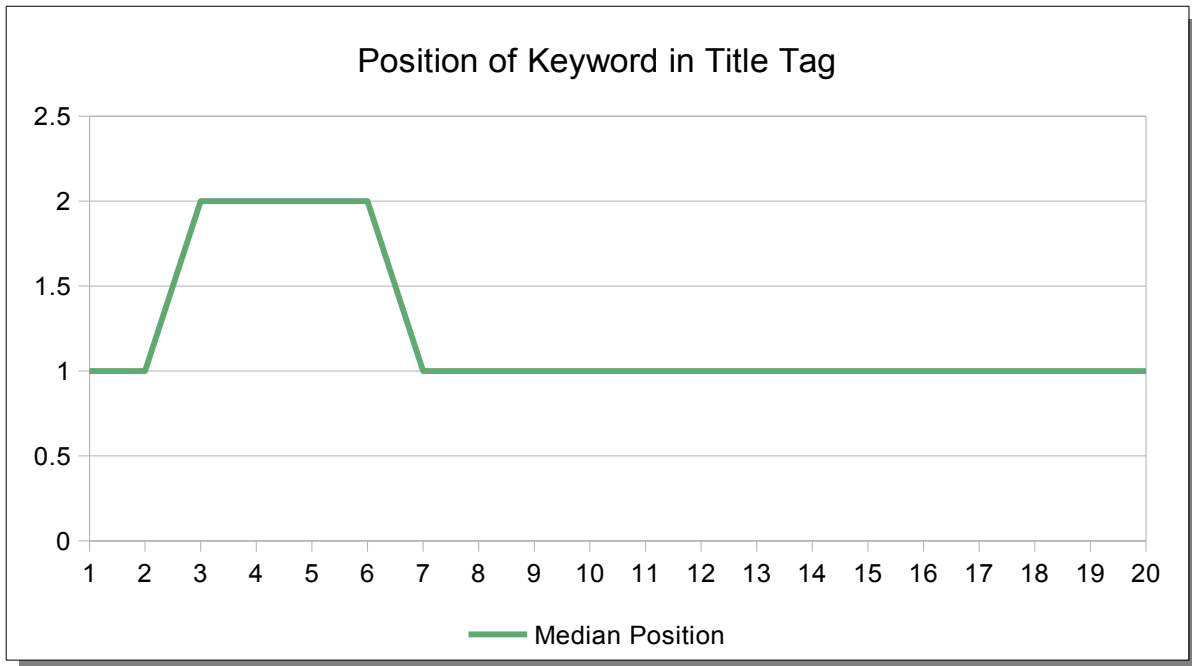
This prompts the question: are shorter keywords more frequently used in exact match in title tags? Further investigation is needed.

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.06	-0.05	-0.11	-0.14
percentage	85.3%	56.2%	34.1%	27%

Our analysis confirms that shorter keywords are more frequently used in an exact match within the title tags of top-ranking pages.

Position of the Keyword in Title tag

The position of the keyword within the title tag is explored as a potential ranking factor, assessing whether placing the keyword closer to the beginning might influence Baidu rankings.

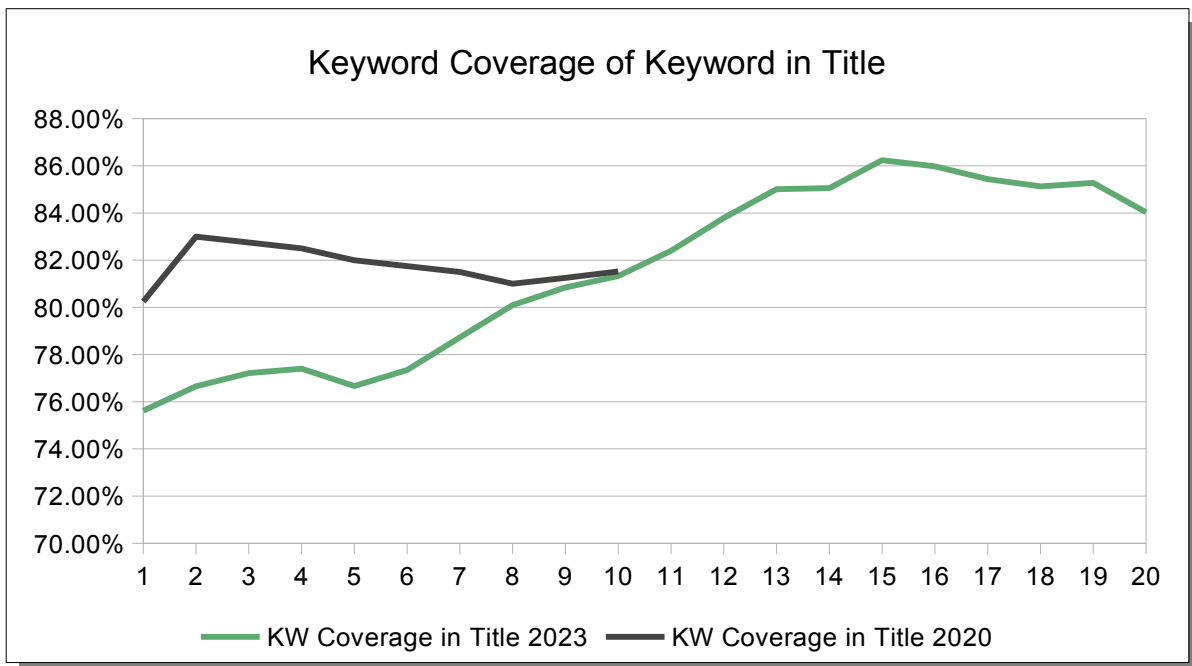


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	-0.09	-0.03	0.06
med. position	1	1	1	2

The analysis reveals that keywords are typically positioned directly in the beginning of the title tags.

Keyword Character Coverage in Title

Only 36% of top-ranking pages feature the exact match keyword in the title tag. This may be attributed to the nuances of the Chinese language, where adding a character can disrupt exact match while retaining the keyword's essence. Consequently, we will evaluate the character coverage of the keyword within the title tag.



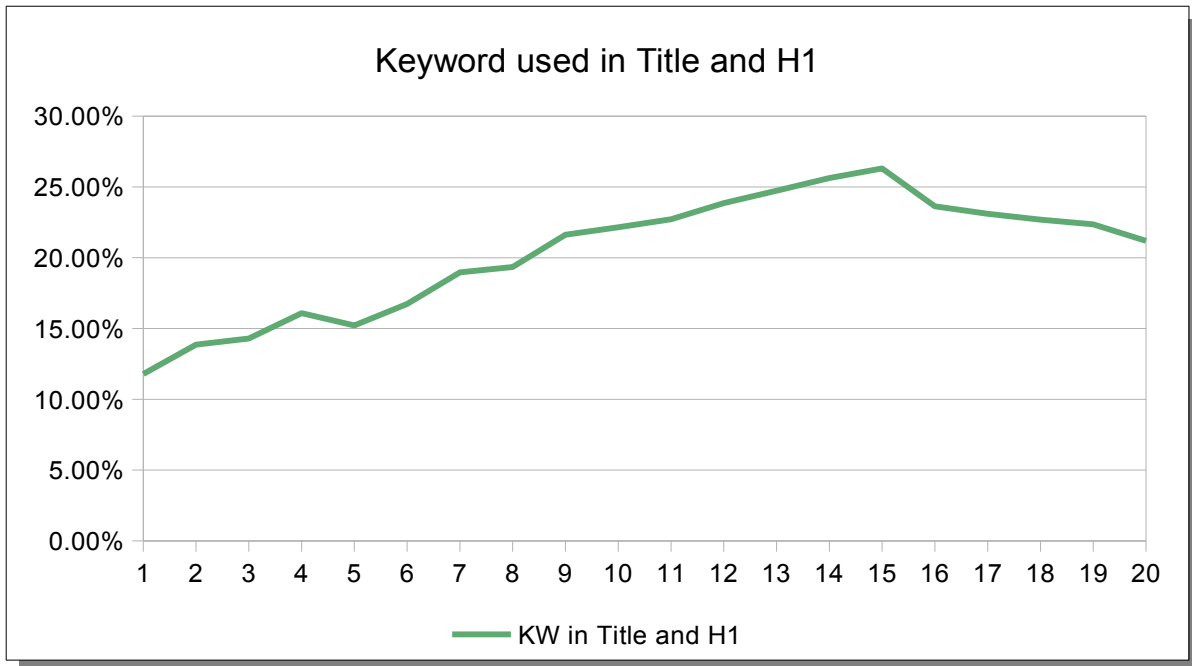
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.14	-0.18	-0.18	-0.11
char. coverage	81.0%	88.3%	85.3%	74.8%

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.086	-0.08	-0.14	-0.18
char. coverage	89.9%	87%	82.3%	79.3%

Observations indicate that short-head keywords typically exhibit more complete coverage in the title tag. Furthermore, the brevity of a keyword directly correlates with its likelihood of complete representation.

Keyword used in Title AND H1

The factor under examination is the concurrent utilization of the target keyword in both the title tag and the h1 tag, assessing whether this dual placement influences Baidu rankings.

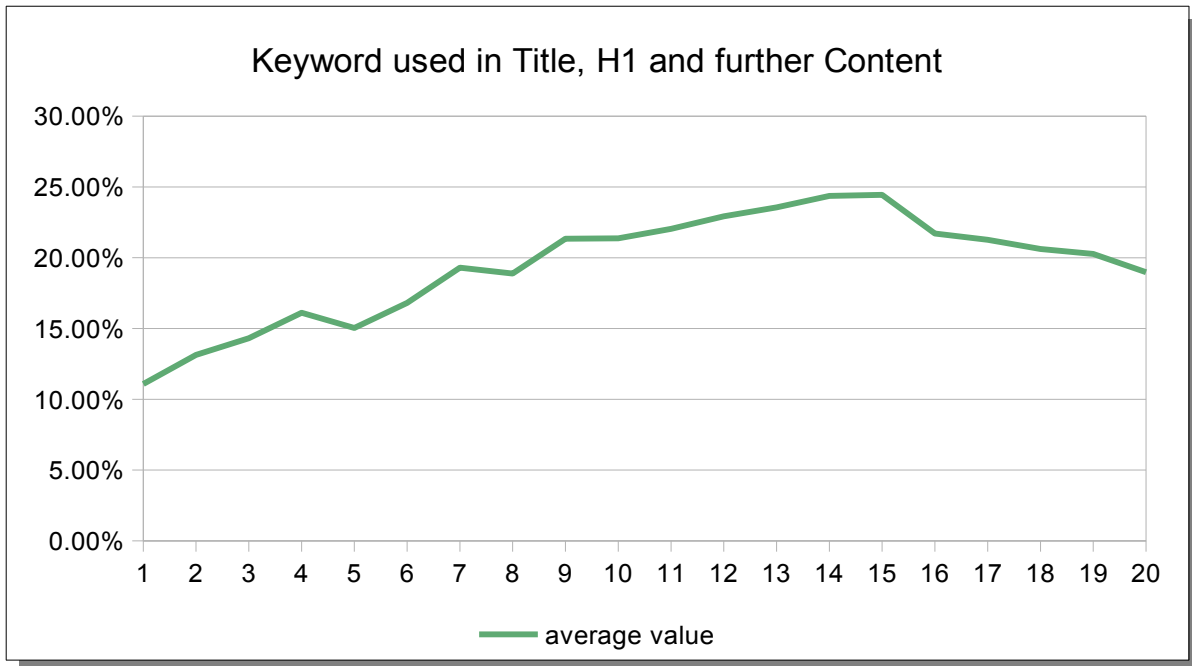


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.08	-0.14	-0.09	-0.01
kw in title & h1	21%	31.4%	24.8%	10.6%

The data reveals that competitive shorthead keywords have a 31.4% likelihood of appearing in both tags, compared to 24.8% for mid-tail and 10.6% for long-tail. This suggests a greater emphasis on keyword placement for high-competition terms.

Keyword used in Title AND H1 AND in further Content of the page

The factor explored is the simultaneous presence of the target keyword in the title tag, h1 tag, and the main page content, aiming to discern its impact on Baidu rankings.

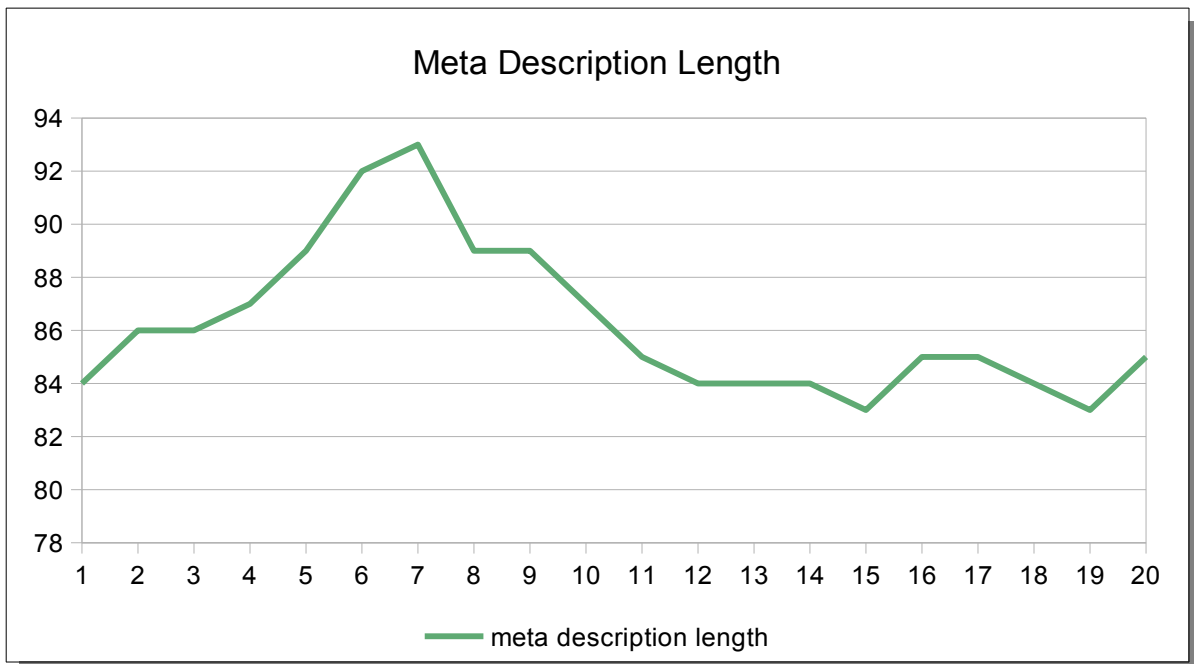


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.06	-0.11	-0.07	0.01
content before kw 1 st found	20%	29.9%	23.8%	10.1%

In our analysis, incorporating an additional factor (keyword presence in the content) exhibited minimal variation in the data. Consequently, it's evident that SEO practitioners tend to consistently incorporate the keyword in the title tag, h1 headline, and within the content itself.

Meta Description Length

The "length of meta description" factor assesses the character count in a page's meta description and its potential influence on Baidu rankings. This metric gauges whether description length affects search positions.

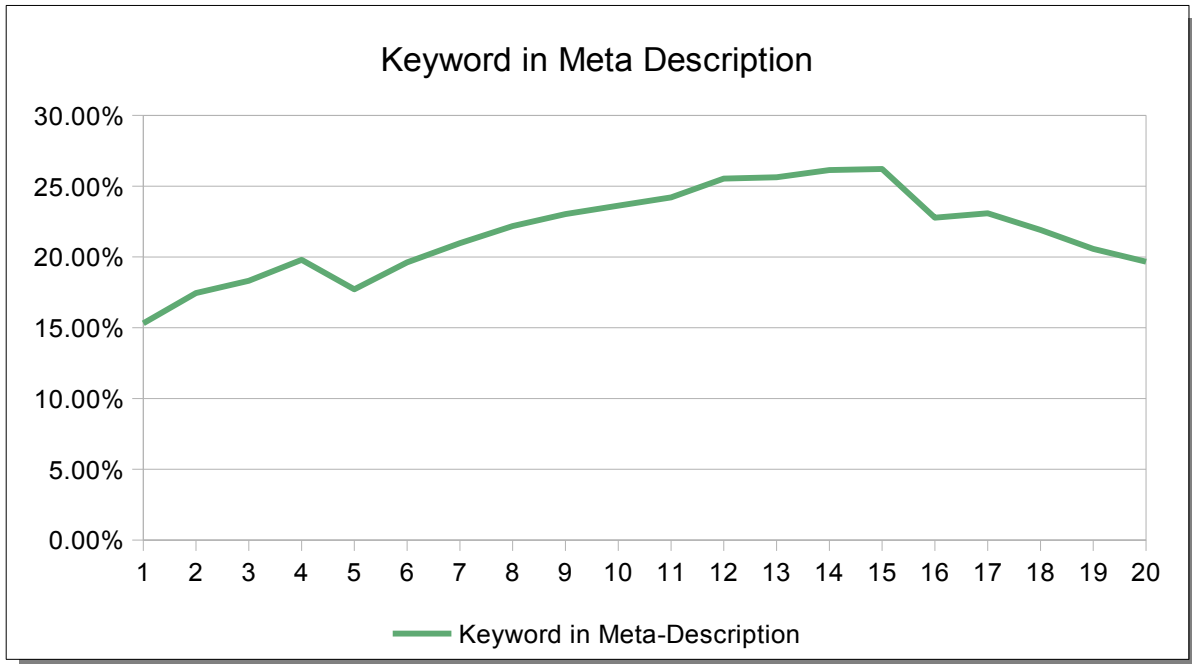


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.01	-0.001	0.003	0.02
characters	86	82	87	88

Meta descriptions on Baidu average between 82 to 88 characters, closely aligning with Baidu's display limit of around 86 characters for desktop search results. A correlation exists between shorter descriptions and better rankings, indicating the benefit of crafting meta descriptions within Baidu's display limits.

Keyword used in Meta-Description

The meta description, a concise summary of a webpage's content, is evaluated for its potential influence on Baidu's ranking algorithm. This factor examines whether the content and length of the meta description impact a page's position in search results.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	-0.07	-0.05	0.003
kw in descr.	22.2%	34.4%	25.7%	10.9%

The data suggests a negative correlation between exact match keyword usage in the meta description and rankings. However, shorthead keywords are more often used in exact match than long-tail keywords.

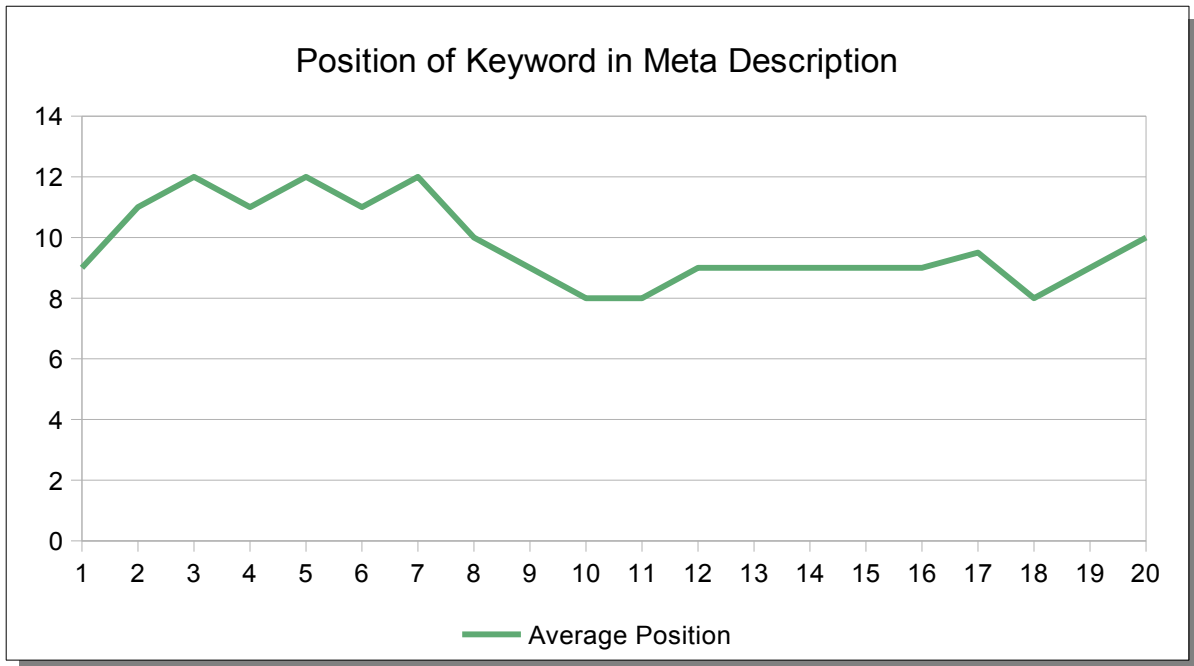
This prompts the question: are shorter keywords more frequently used in exact match in meta description tags? Further investigation is needed.

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.005	-0.01	-0.04	-0.07
kw in descr.	63.9%	40%	21%	14%

Our analysis confirms that shorter keywords are more frequently used in an exact match within the meta description tags of top-ranking pages.

Position of the Keyword in Meta Description

The position of the keyword within the Meta Description tag is explored as a potential ranking factor, assessing whether placing the keyword closer to the beginning might influence Baidu rankings.

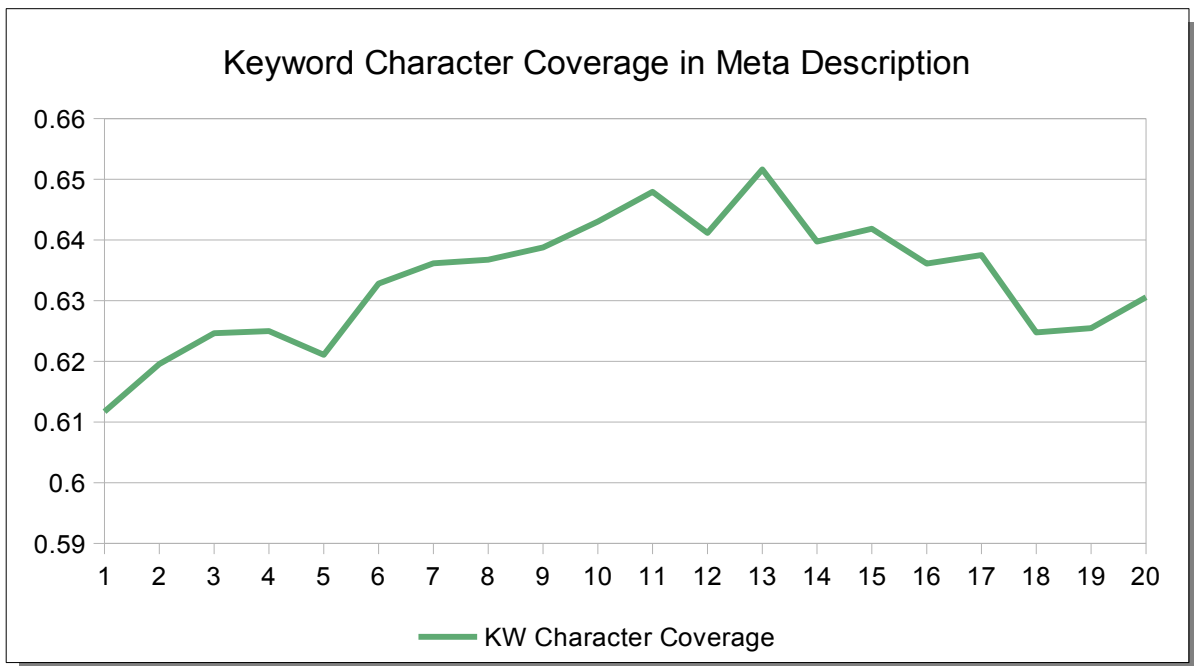


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.02	0.02	0.04	-0.02
med. position	10	7	11	13

The analysis reveals that keywords are typically positioned at an average spot of 10 in the Meta-Description. Notably, more competitive shorthead keywords are positioned closer to the beginning, averaging at 7, while longtail keywords are slightly farther at an average position of 13.

Keyword Character Coverage in Meta Description

Only 36% of top-ranking pages feature the exact match keyword in the title tag. This may be attributed to the nuances of the Chinese language, where adding a character can disrupt exact match while retaining the keyword's essence. Consequently, we will evaluate the character coverage of the keyword within the title tag.



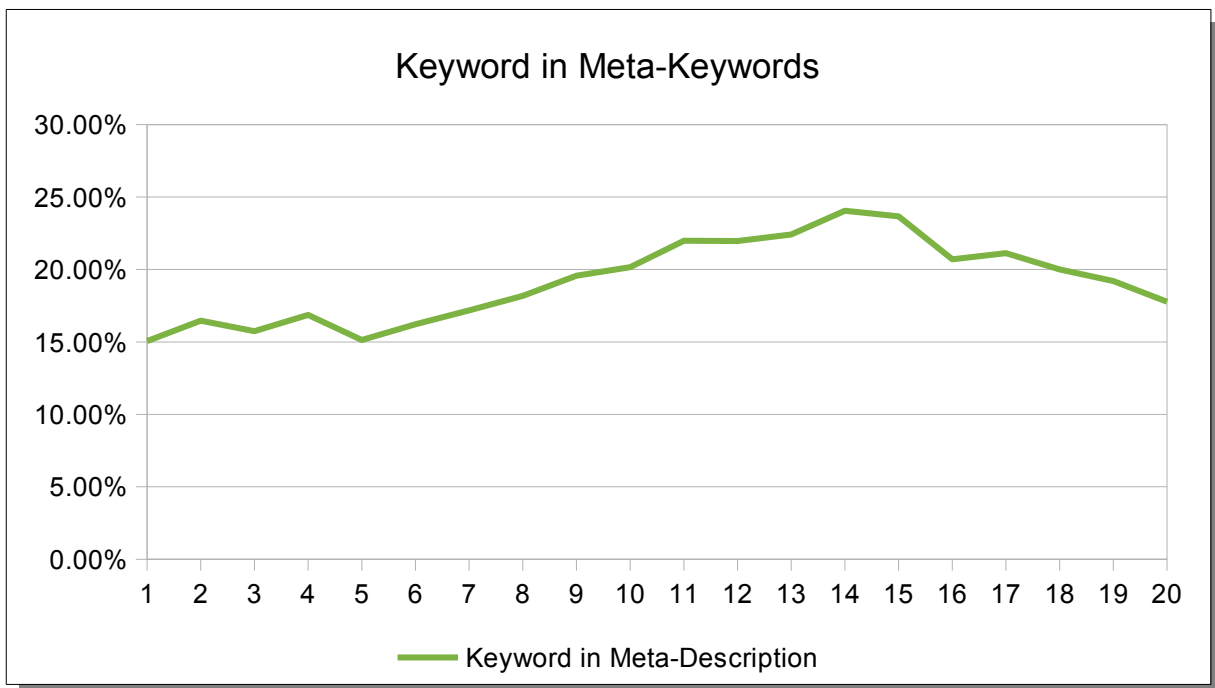
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.05	-0.04	0.01
char. coverage	63.5%	68.5%	66%	57.9%

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.01	-0.001	-0.14	-0.04
char. coverage	68.9%	67.5%	63.6%	61.6%

The keyword character coverage in the meta description is generally lower than in the title tag, suggesting a preference for synonyms or varied phrasing in the description. However, for more competitive keywords, there's a trend towards a more complete inclusion of the keyword in the meta description.

Keyword used in Meta-Keywords

The meta keywords tag, despite being publicly dismissed by a Baidu representative named Lee, persists in website implementations. This analysis probes potential correlations between the inclusion of target keywords in the meta keywords tag and Baidu rankings.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.04	-0.08	-0.07	0.01
kw in meta kw	19.5%	31.6%	21.9%	9.3%

The data suggests a negative correlation between exact match keyword usage in the meta keywords and rankings. However, shorthead keywords are more often used in exact match than long-tail keywords.

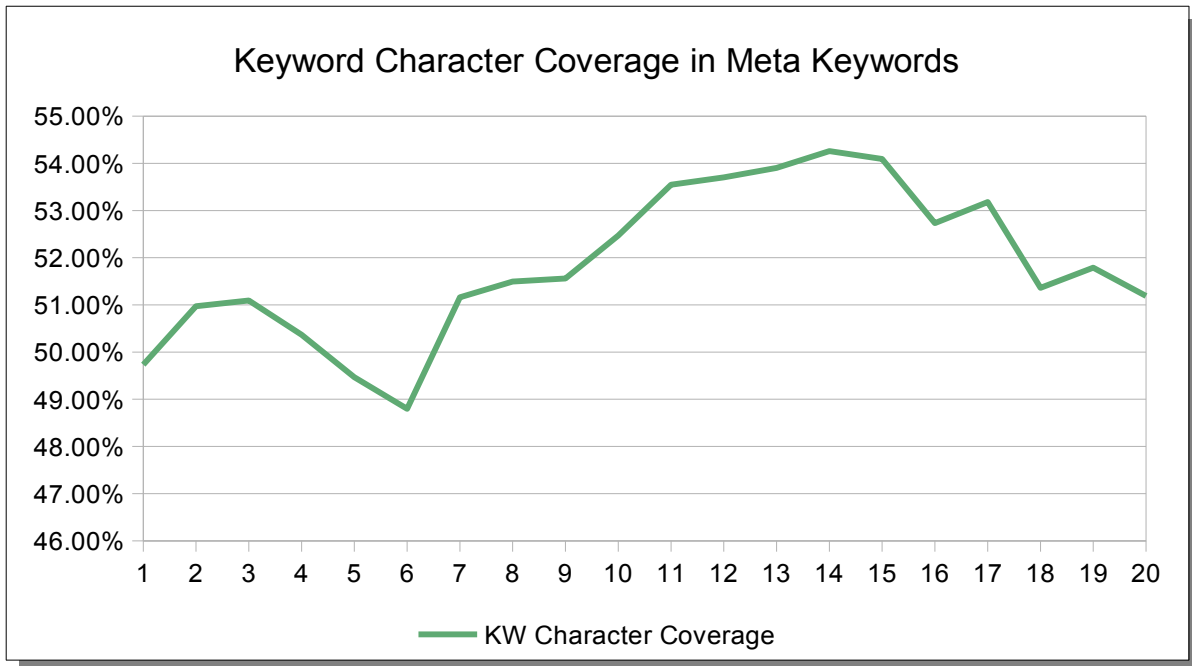
This prompts the question: are shorter keywords more frequently used in exact match in meta description tags? Further investigation is needed.

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.01	-0.06	-0.02	-0.07
kw in meta kw	52.6%	18.3%	33.3%	13.6%

Interestingly the Meta-Description more often finds shorter keyword of 1-2 characters in length in exact match than the Meta-Keywords.

Keyword Character Coverage in Meta Keywords

This metric determines how comprehensively a keyword, in terms of its Chinese characters, is represented in the Meta Keywords tag, assessing its potential influence on Baidu rankings.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.05	0.004	-0.05
char. coverage	57.7%	57.2%	46.96%	53.9%

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.01	-0.01	-0.03	-0.3
char. coverage	57.18%	56.4%	52.8%	49.5%

Despite Baidu's official stance disregarding the meta keywords tag as a ranking factor, many websites continue to utilize it. With an observed 57% keyword character coverage, it's evident that these tags are often meticulously curated.

Expert comment:



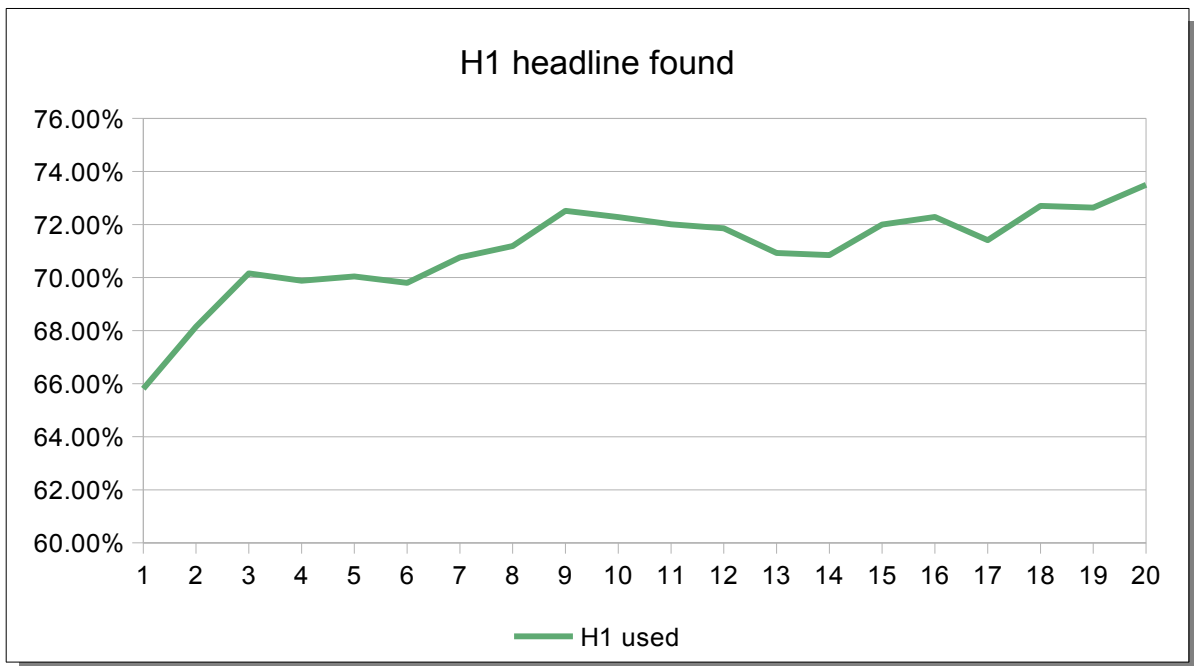
“The number of websites (including most of the major players on SERPs) still using Meta-Keyword tags on Chinese websites shocks me. Baidu often gets a reputation of lagging behind in practices like this, especially in the west where we mainly focus on Google. Not only do I think Baidu is way beyond this, they have also stated they haven't used meta keywords in quite some time. Between my personal testing and the data presented here I still believe it to be a complete non-factor.”

Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com

Headlines

H1 headline

The H1 headline, a primary HTML element, is evaluated for its potential influence on Baidu rankings. This factor assesses the significance of the H1 tag in the hierarchy of on-page elements.

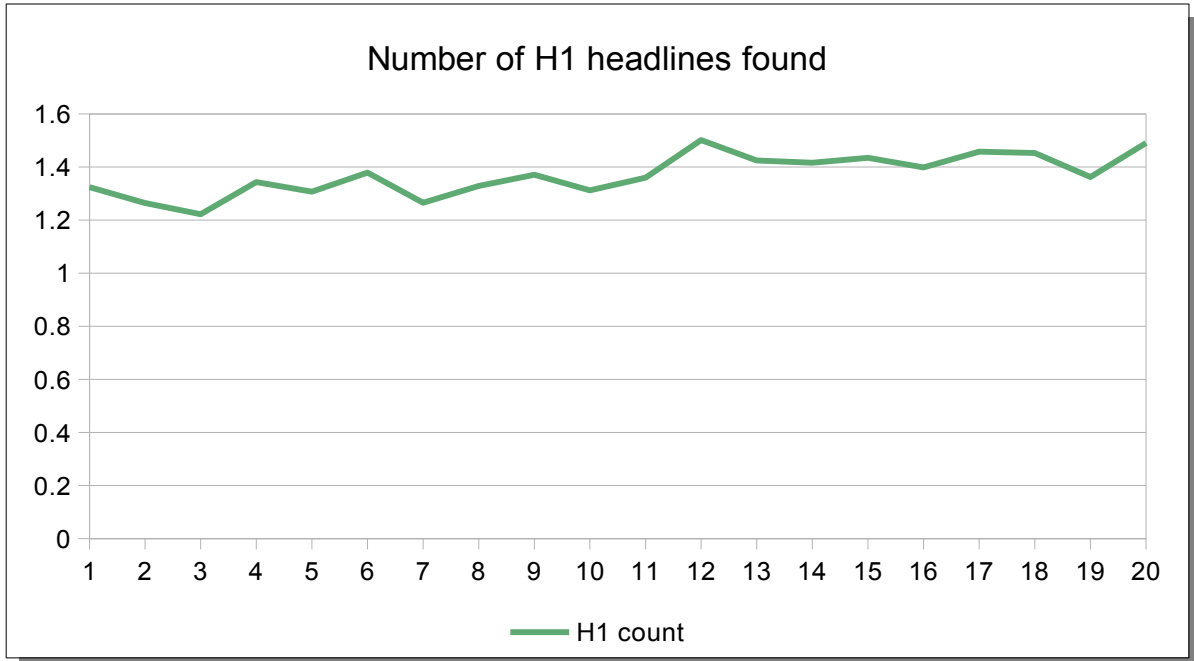


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.09	-0.03	0.03
percentage	71.2%	69.5%	72.7%	71.3%

On average, 71.2% of ranking pages utilize an H1 headline. A discernible negative correlation exists between the presence of an H1 headline and superior rankings on Baidu.

Number of H1 tags found

The potential ranking factor under scrutiny is the exclusive use of a single H1 headline on a page, assessing its influence on Baidu's ranking mechanism.



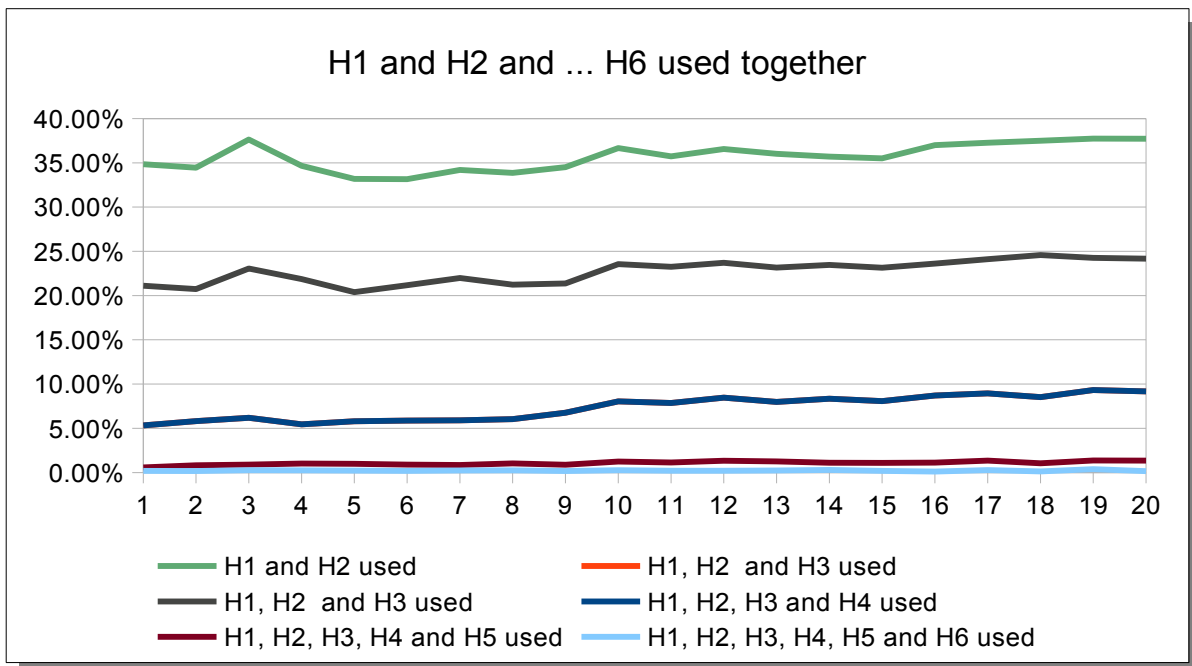
The following correlation score looks at the use of only one H1 used as this was our expected result:

	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.06	0.04	0.06	0.07
number	1.4	1.38	1.34	1.42

Observations indicate a modest positive correlation between superior rankings and the singular use of an H1 headline. Notably, of the 71% top-ranking pages employing an H1, the majority utilize just one H1 headline.

H1 and H2 and ... H6 found all in the document

The factor under scrutiny is the simultaneous implementation of both H1 and H2 headlines on a page, the simultaneous implementation of H1, H2 and H3 on a page, ... up to simultaneous H1 to H6 implementation, assessing its potential influence on Baidu rankings. This metric evaluates a first step whether a combined hierarchical structure of headlines impacts search positioning.

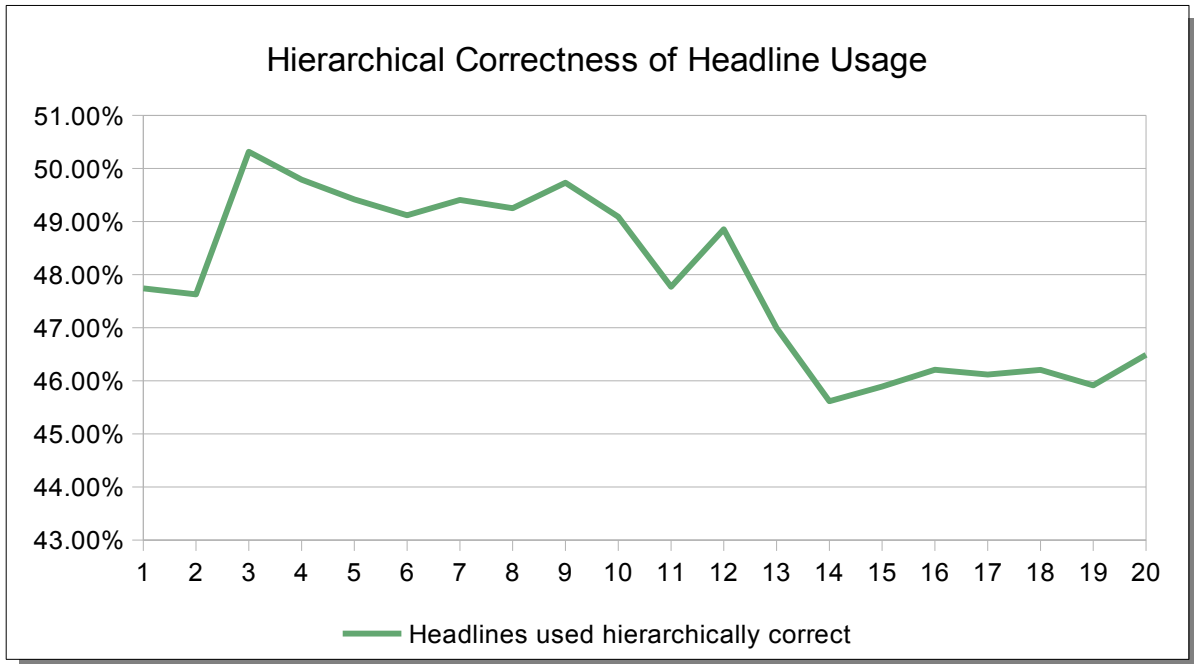


	H1 and H2	H1 to H3	H1 to H4	H1 to H5	H1 to H6
correlation	-0.02	-0.02	-0.05	-0.01	-0.0003
used all together	35.7%	22.8%	7.4%	1.1%	0.2%

In our findings, concurrent use of H1 and H2 headlines is limited, with only 35.7% of top-ranking pages employing both. As headline depth increases, their combined occurrence predictably decreases.

Headlines in hierarchical structure

The examination now shifts to the hierarchical correctness of headline usage. While the simultaneous employment of multiple headline levels is infrequent, the study probes the prevalence and potential impact of maintaining a structured and sequential order in headline tags.

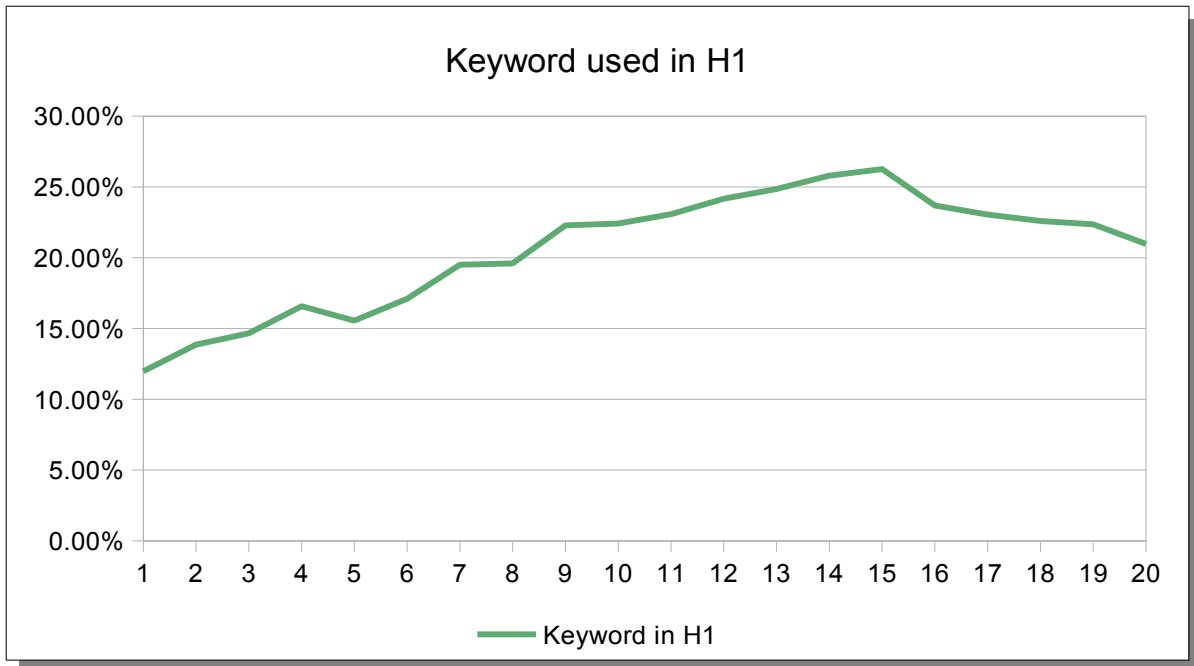


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.03	0.02	-0.05	-0.03
hierarch. H1-x	47.8%	48.8%	46.9%	48.1%

The findings indicate a subtle positive correlation between hierarchically correct headline usage and better rankings. Among the websites employing headlines, nearly half adhere to a structured and sequential order.

Keyword used in H1 headline

This analysis delves into the potential impact of keyword inclusion within the H1 headline on Baidu rankings, assessing if direct keyword representation in this primary header influences search result positioning.

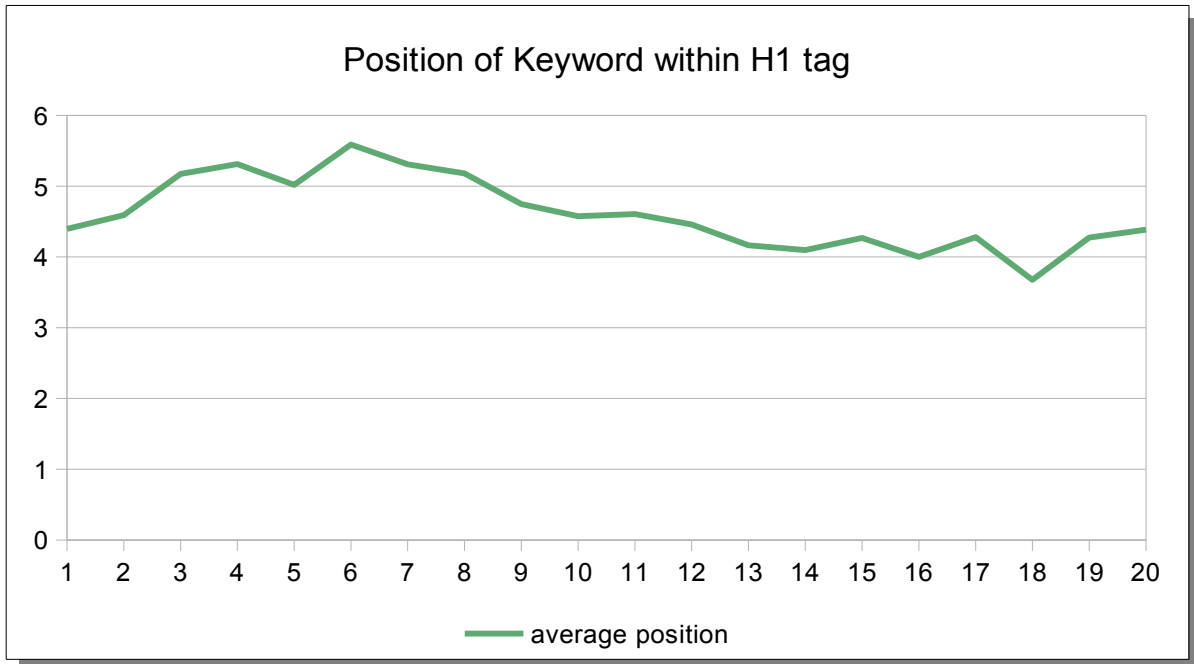


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.07	-0.13	-0.09	-0.002
kw in h1	21.1%	31.7%	25%	10.7%

Despite a mild negative correlation and a mere 21% of top-ranked pages incorporating the keyword in the H1, there's a marked difference based on keyword competition. Specifically, 31.7% of short-head keywords feature the keyword in the H1, contrasted with a mere 10% for long-tail keywords.

Position of Keyword in h1 headline

The position of the keyword within a headline is scrutinized as a potential ranking determinant, evaluating whether its placement influences Baidu rankings. This metric probes if leading, central, or trailing keyword placements affect search outcomes.

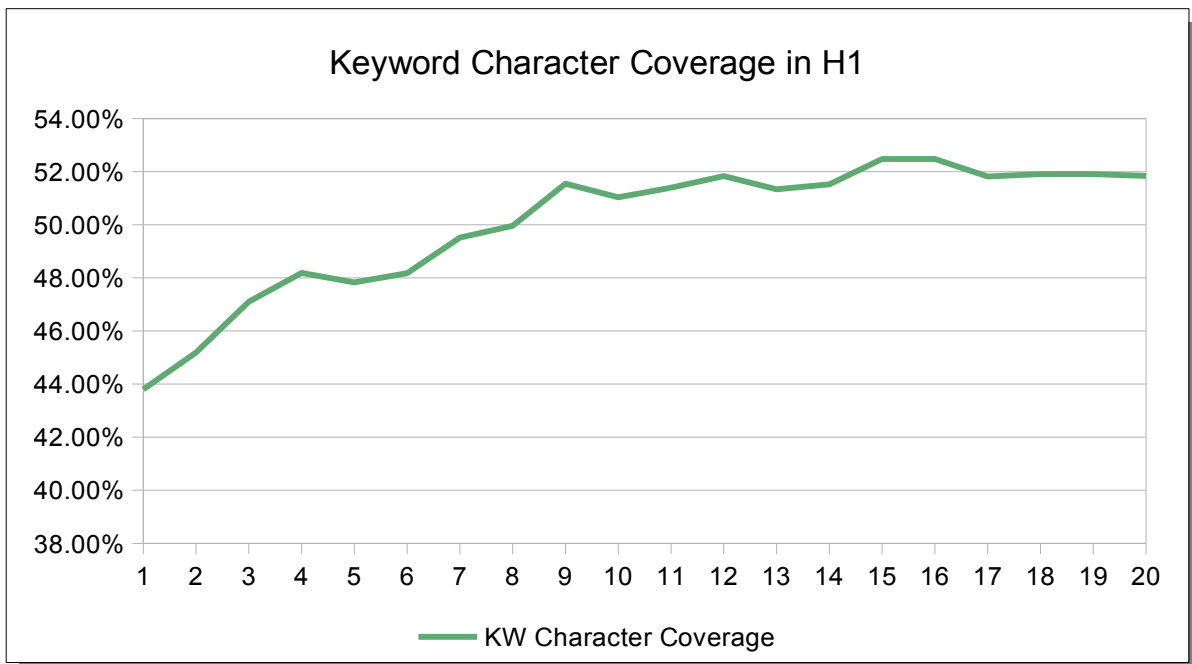


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.05	-0.09	-0.05	0.04
position	4.5	4	4.7	5.3

Keywords within H1 headlines predominantly occupy positions between 4 and 5, suggesting a central or slightly leading placement in Baidu's top-ranking pages.

Keyword Character Coverage in H1 headline

The "keyword character coverage" in the H1 tag evaluates the extent to which the characters of a keyword appear in H1, even if not in exact match. Given the modest 21% exact match presence in H1 headlines, this metric aims to discern the frequency of synonym or variation usage, offering deeper insights into keyword representation.

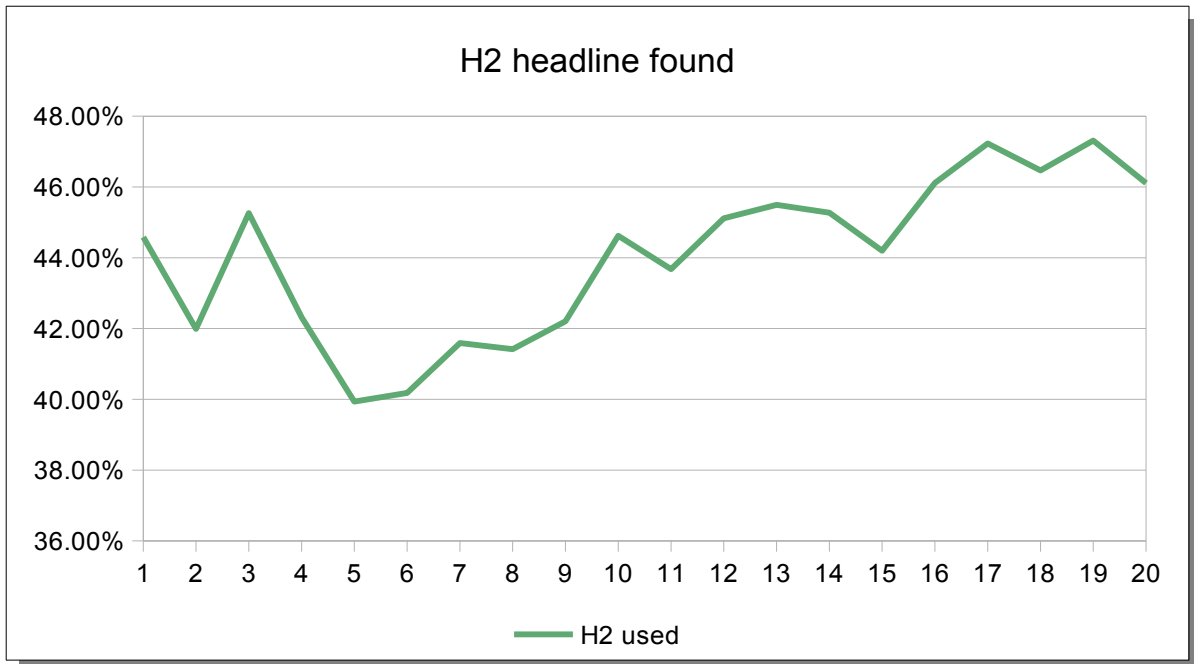


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.05	-0.11	-0.07	0.02
kw char. cov.	50.4%	52.5%	53.8%	46.3%

A slight negative correlation exists between comprehensive keyword coverage in H1 and superior rankings. However, while only 20% of H1 headlines feature the exact match keyword, over 50% of them cover the keyword's characters. This suggests a prevalent use of keyword variations in top-ranking sites.

H2 headline

The H2 headline, a secondary header tag, is examined for its potential influence on Baidu's ranking algorithm, assessing its role in content structuring and keyword emphasis.

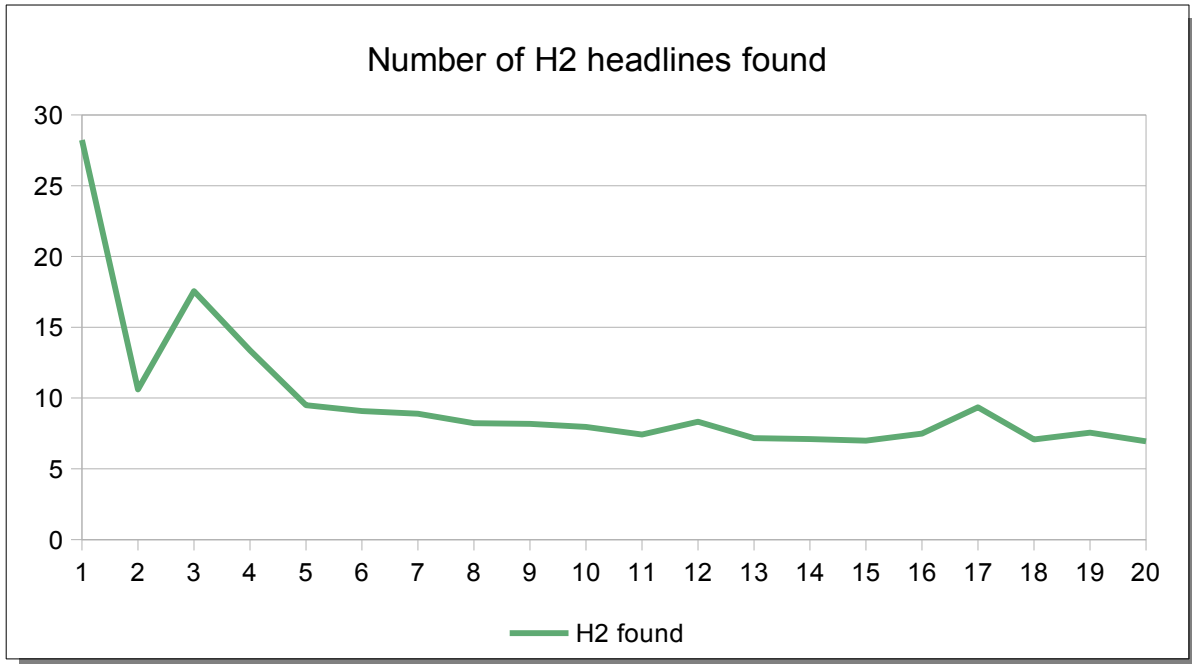


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	-0.04	-0.03	-0.03
H2 found	44%	45.4%	44.3%	42.8%

On average, 44% of ranking pages utilize an H2 headline. A discernible negative correlation exists between the presence of an H2 headline and superior rankings on Baidu.

Number of H2 tags found

The potential ranking factor under scrutiny is the use of H2 headlines on a page, assessing its influence on Baidu's ranking mechanism.

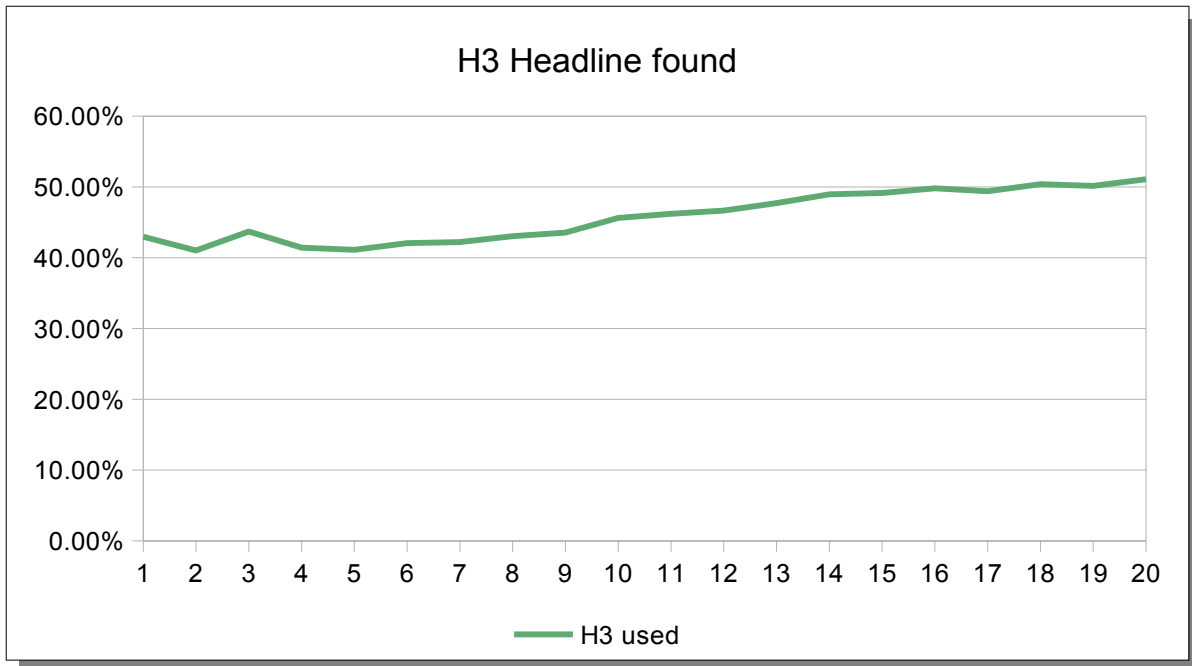


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.04	0.03	0.04	0.05
Number of H2	9.12	11.5	8.1	8.2

Observations indicate a modest positive correlation between superior rankings and the use of more H2 headlines.

H3 headline

The H3 headline, a tertiary header tag, is analyzed for its potential impact on Baidu's ranking outcomes, gauging its significance in content hierarchy and keyword highlighting.

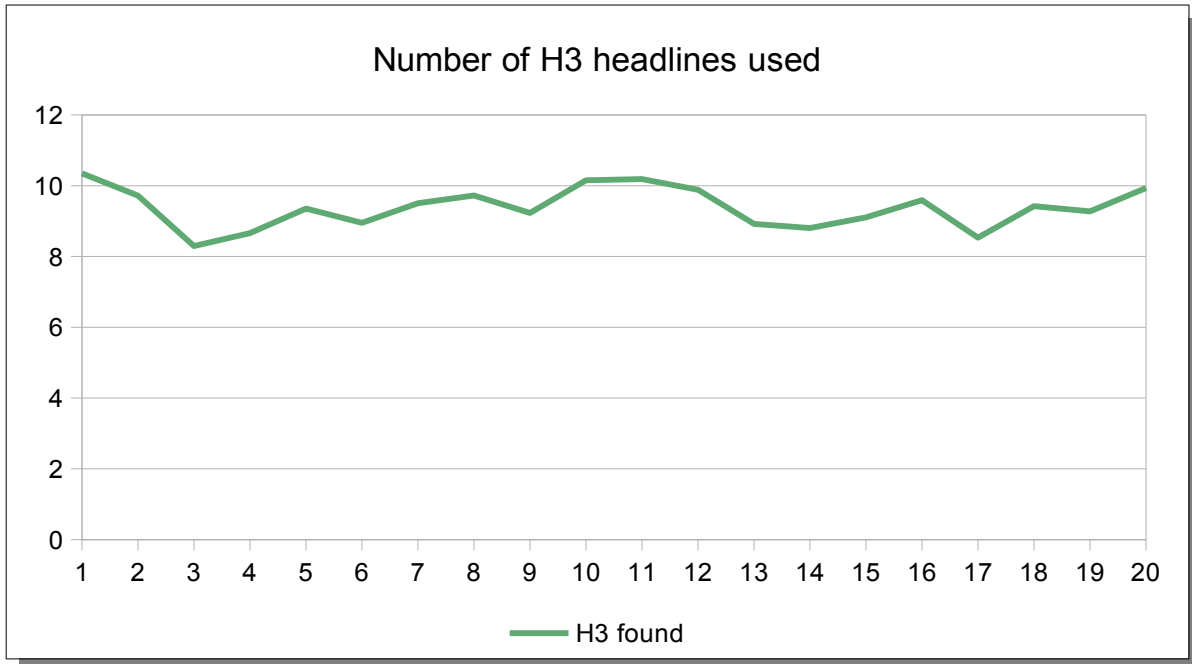


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.07	-0.09	-0.05	-0.07
H3 found	46%	45%	45.9%	47.2%

On average, 46% of ranking pages utilize an H3 headline. A discernible negative correlation exists between the presence of an H3 headline and superior rankings on Baidu.

Number of H3 tags found

The potential ranking factor under scrutiny is the use of H3 headlines on a page, assessing its influence on Baidu's ranking mechanism.

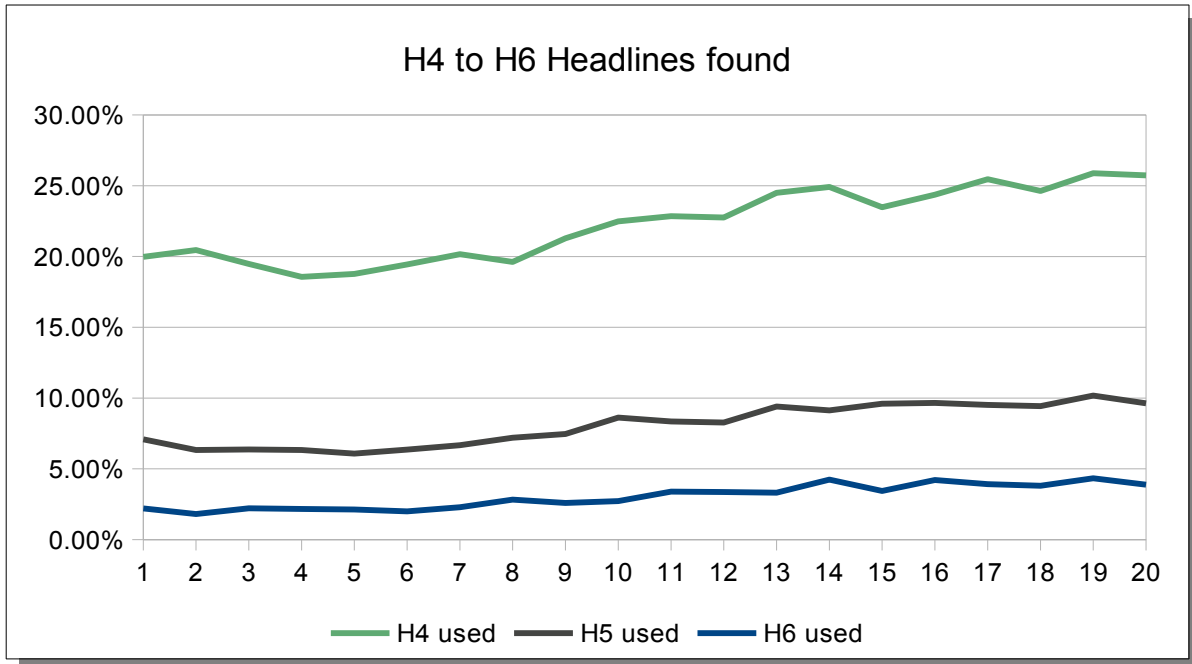


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.05	-0.05	-0.05	-0.05
Number of H3	9.4	9.9	8.5	9.7

Observations indicate a modest positive correlation between superior rankings and the use of more H3 headlines.

H4-6 Headlines

The H4 to H6 tags, analyzed for their potential impact on Baidu's ranking outcomes, gauging its significance in content hierarchy and keyword highlighting.

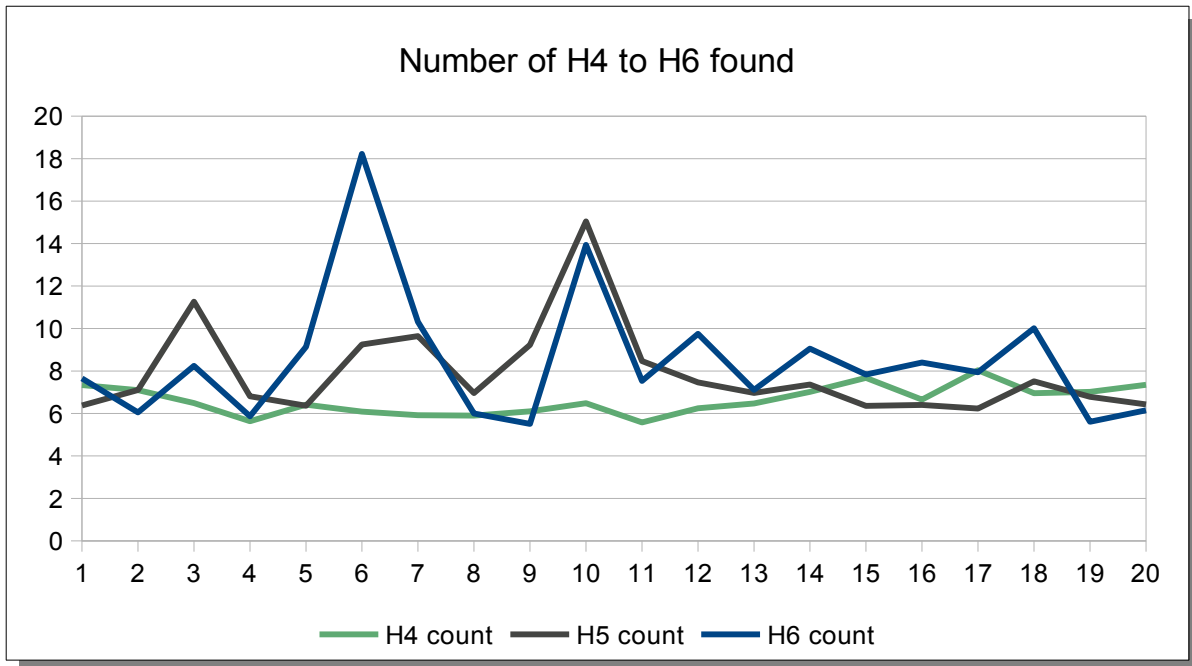


	H4	H5	H6
correlation	-0.05	-0.05	-0.04
percentage Hx found	22.4%	8.2%	3.1%

As headline depth increases, its presence diminishes among top-ranking pages on Baidu. Specifically, H4 appears in 22%, H5 in 8%, and H6 in a mere 3% of these pages.

Number of H4 to H6 tags found

The potential ranking factor under scrutiny is the use of H4 to H6 headlines on a page (only counted on pages using H4 to H6), assessing their influence on Baidu's ranking mechanism.

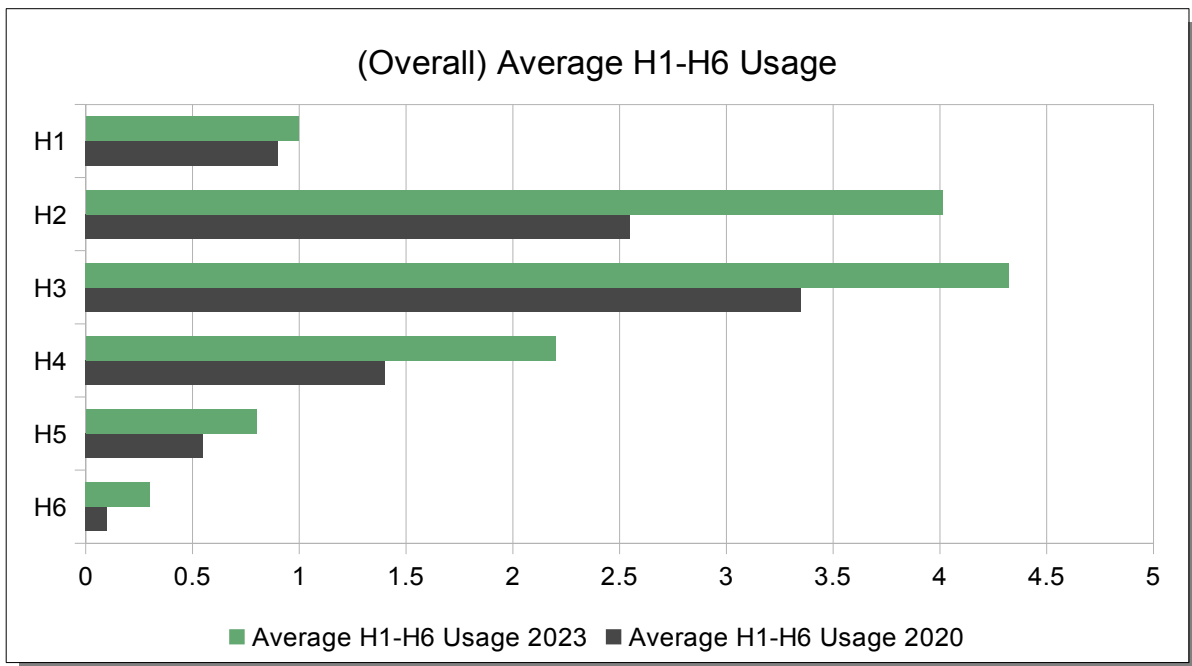


	H4	H5	H6
correlation	-0.1	-0.02	0.05
number	6.6	7.8	8.4

If used at all (H4 appears in 22%, H5 in 8%, and H6 in a mere 3% of top ranking pages), we find around 6-11 headlines of each used. That's just an observation, no SEO recommendation.

Overall Average Headline Counts

In the following analysis, we examine the average usage of heading tags—from H1 to H6—across web pages, including those that abstain from using these tags altogether. This comprehensive approach provides a holistic view of headline utilization, offering insights into both prevalence and distribution of heading elements within the current web landscape.



	H1	H2	H3	H4	H5	H6
avg. number	0.99	4.01	4.3	2.2	0.8	0.3

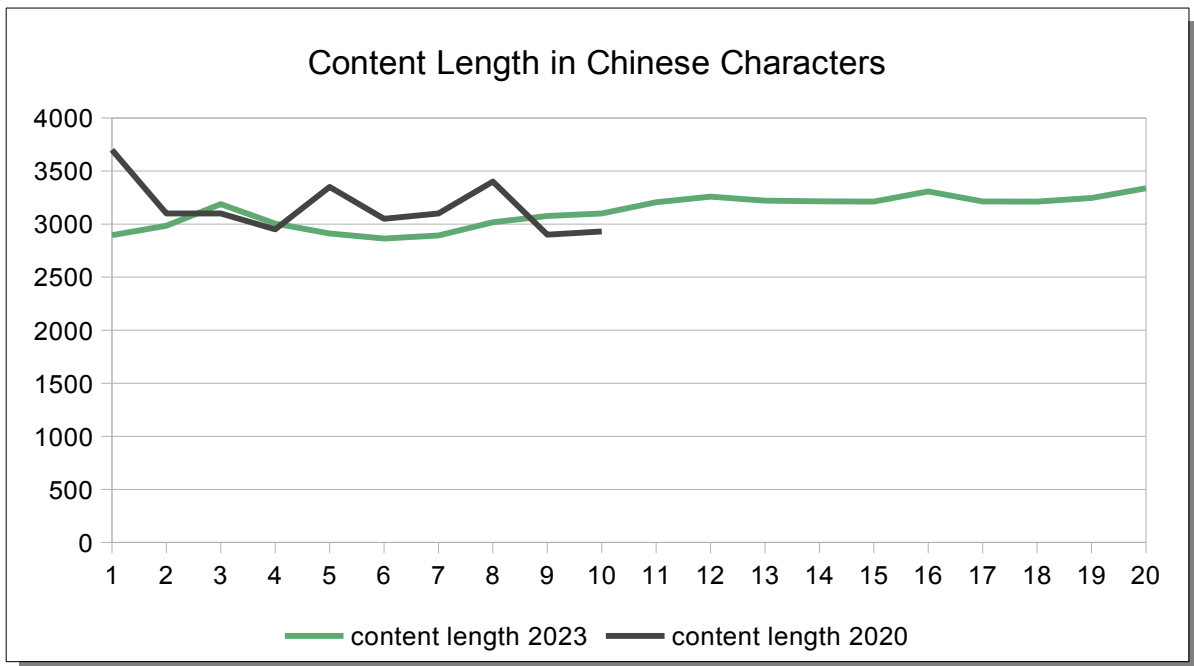
The above chart and table summarizes our findings of heading markup from H1 to H6 showing the averages across the Baidu Top 20 (2020 only across Top 10) ranking pages across our data-set.

Visible Text Analysis

Moving beyond the scope of headers, titles, and meta-tags, the analysis underscores aspects such as content length and keyword utilization, offering a profound understanding of how visible text influences SEO dynamics.

Content Length

The length of content emerges as a potential ranking factor, with the hypothesis that more comprehensive, often lengthier, materials may secure superior rankings. As it is hard to distinguish between words in Chinese, the content length is measured in number of Characters.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.05	-0.07	-0.06	-0.03
average	4,929	4,492	4,702	5,413
median	3,147	2,826	3,134	3,378

The factor of content length, measured in Chinese characters, presents a subtle shift in its correlation with Baidu rankings when juxtaposed against the 2020 data. While the previous year marked a positive correlation of 0.53, the current dataset reveals a minor negative correlation of -0.05 for the entire keyword set.

This trend deepens slightly for short head keywords at -0.07, stabilizes for mid tail at -0.06, and becomes more attenuated for long tail keywords at -0.03.

Intriguingly, the median content length for the top 20 positions in the current dataset (3100 characters) aligns closely with the 2020 study's corresponding median for the top 10 positions of 3194 characters.

Segmenting further: short head keywords average at 2826 characters, mid tail keywords hover around 3134 characters, and long tail keywords extend to 3378 characters.

Looking at the slightly negative correlation through all keyword demand groups, and also seeing that the median content length for short-head keywords is actually the lowest, followed by mid-tail keywords, and with longtail keywords showing the longest content by median calculation (the same when looking at average values), we can surely not say, longer content ranks better.

But we see a median content length of 2,826 Chinese Characters, that might be seen as an unofficial benchmark, that seems to work well.

Expert comments:

“For Q&A related article content, Baidu will tend to give richer answers, a variety of solutions of the content of the good ranking; If it is a short answer, even if it is a direct and effective solution to the problem, it may not be adopted by Baidu.”



Qing He, Head of Search, Jademond Digital,
www.jademond.com

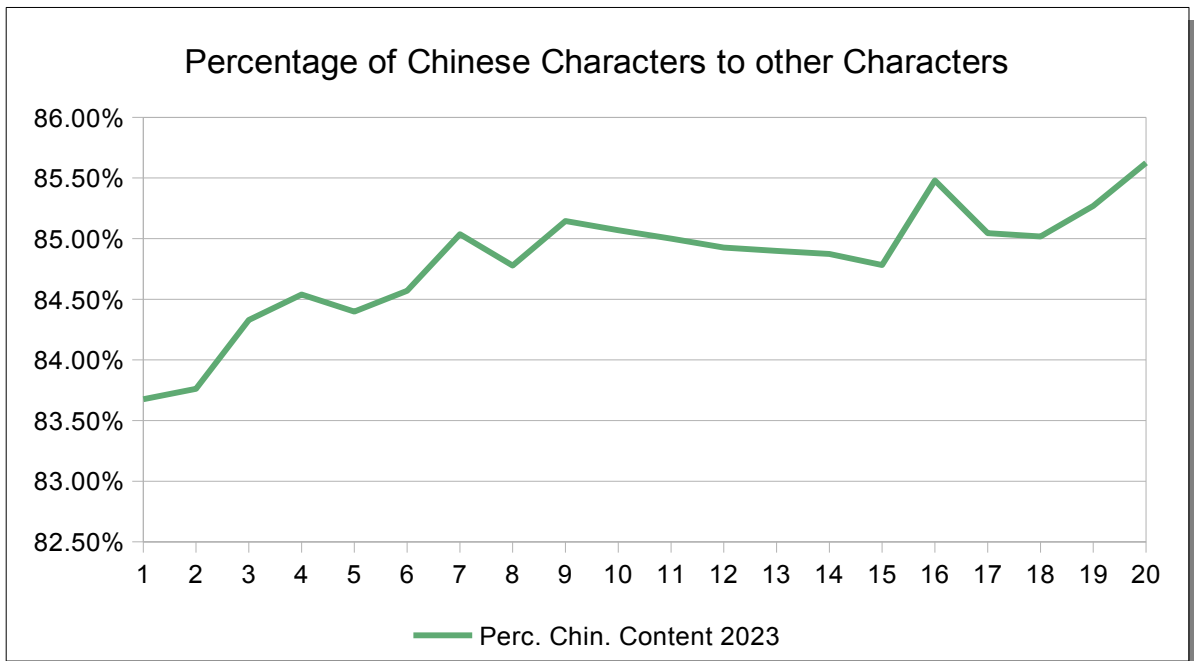


“Here is my opinion regarding the content: It's proven that a webpage with more than 1000 Chinese characters / words ranks better than a webpage that has poorer volume in terms of texts. Baidu needs to find information on the webpages in order to crawl and index them frequently.”

Veronique Duong, International SEO & Marketing Digital Expert, veronique-duong.com

Chinese Content on the Page

Chinese content on a page is crucial for Baidu rankings due to its primary audience. While international companies may blend English teasers with Chinese, we wanted to analyze if the proportion of Chinese characters to other characters do have an impact on Baidu SEO.



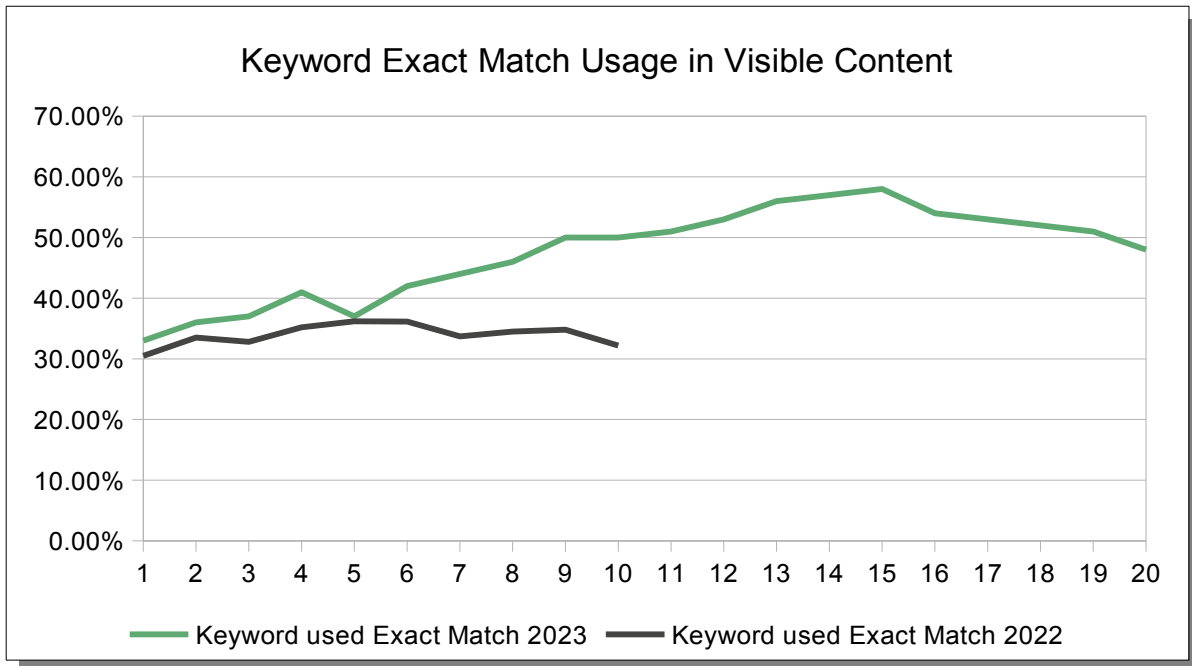
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.05	-0.02	0.01
average	81%	82%	81%	80%
median	85%	86%	85%	84%

There's a slight negative correlation between a higher percentage of Chinese characters and improved Baidu rankings. However, the percentage has increased compared to the 2020 study data, and also short-head keywords show a median

of 86% Chinese characters in top 20 rankings, compared to 85% for mid-tail and 84% for long-tail keywords underlining the importance of Chinese content.

Average (exact match) Keyword Coverage in Visible Content

Exact Match Keyword Coverage in Visible Content is a potential ranking factor, which assesses the extent to which target keywords are precisely represented within the main content displayed to users.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.11	-0.19	-0.16	-0.04
percentage	49%	66%	57%	30%

Despite the observable negative correlation between superior rankings and exact keyword usage in visible content, pages targeting competitive short-head keywords exhibit a higher keyword presence at 66%, compared to long-tail keywords at 30%. However, is this trend merely a function of short-head keywords inherently being shorter than long-tail ones? We'll delve deeper by analyzing keyword length.

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.06	-0.09	-0.12	-0.14
percentage	93%	74%	49%	36%

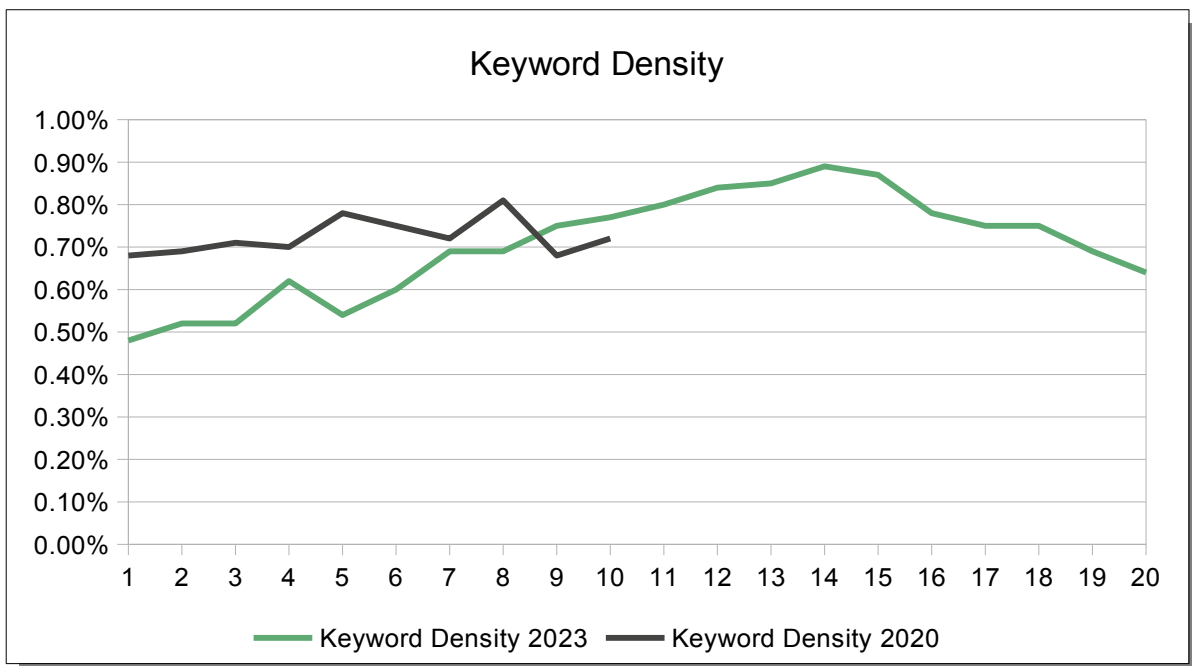
Upon examining exact match keyword usage across varying keyword lengths, it's evident that despite individual negative correlations, shorter keywords are more frequently used in exact match within top-ranking content.

Keyword Density

Keyword density, a commonly analyzed factor in Indo-European language SEO, presents challenges in the Chinese context due to the absence of spaces between words in written Chinese. To address this, we employ the formula:

$$\frac{(\text{keyword length}) \times (\text{count of keyword in content})}{(\text{number of characters in content})}$$

to determine keyword density.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.1	-0.15	-0.12	-0.03
kw-density	0.72%	0.48%	0.79%	0.34%

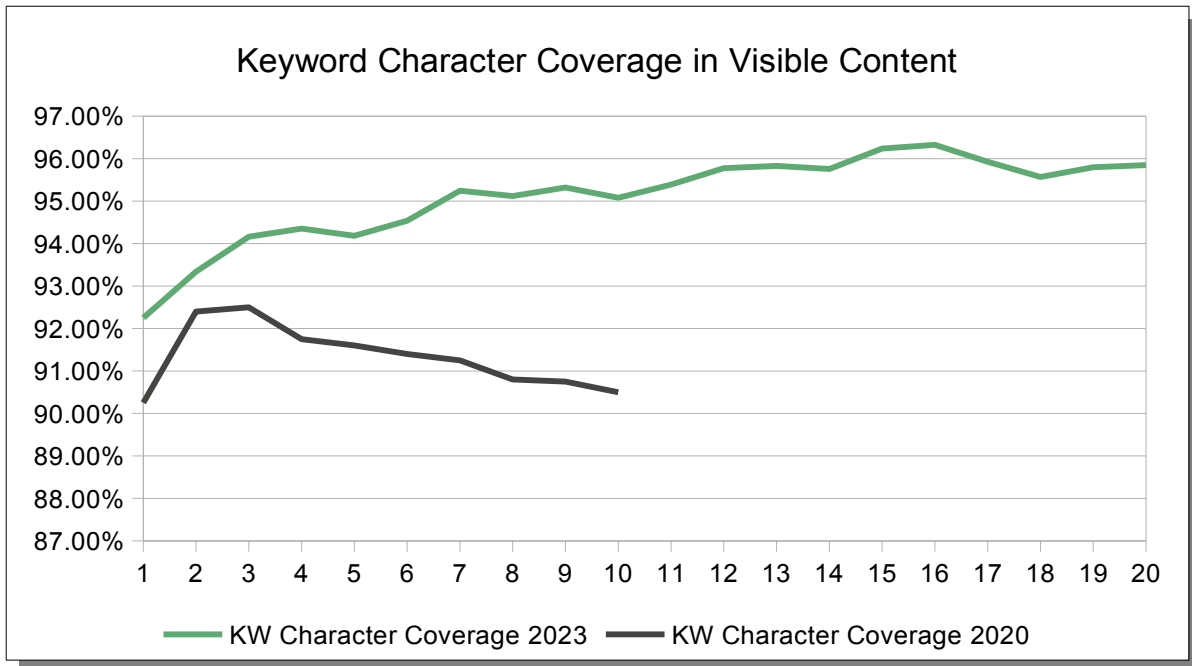
And Just as before, let's also look at the keyword length:

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.06	-0.05	-0.11	-0.14
kw-density	1.35%	0.48%	0.65%	0.5%

Keyword density averages 0.72%, well below often-cited benchmarks of 5% to 7% for Baidu SEO. There's a negative correlation between density and ranking, with mid-tail keywords at 0.79% density, and short-head and long-tail at 0.48% and 0.34% respectively. Analyzing by keyword length, shorter keywords have higher densities (1.35%), while longer ones are below 0.66%. Keywords of 3-4 and 7+ characters hover around 0.5%, but 5-6 character keywords stand at 0.65%.

Keyword Character Coverage in Visible Content

The factor being evaluated is the Keyword Character Coverage in Visible Content, gauging how comprehensively the characters of a target keyword are present within the primary content visible to users.



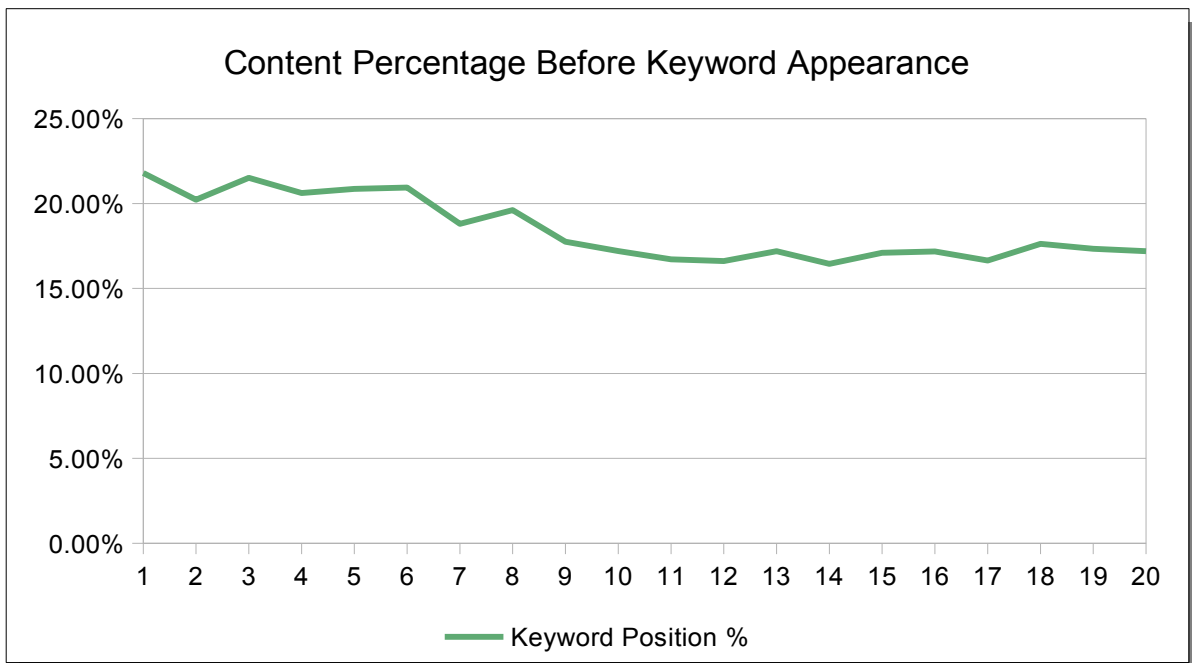
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.09	-0.09	-0.12	-0.09
kw-coverage	95.3%	96%	95.8%	94.3%

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	-0.07	-0.05	-0.08	-0.13
kw-coverage	95.2%	96.2%	95.5%	94.7%

The data underscores the importance of incorporating every character of the target keyword within the content. On average, 95.3% of keyword characters are present in ranking content. This holds true for long-tail keywords at 94.3% and for lengthier keywords at 92.7%, emphasizing its relevance across various keyword complexities.

First position of the Keyword in Visible content

Examining the initial placement of the target keyword in visible content offers insights into its potential impact on Baidu rankings. The subsequent analysis evaluates the content percentage preceding the keyword's inaugural appearance, shedding light on optimal keyword positioning within the content. The assumption is: the sooner, the better.

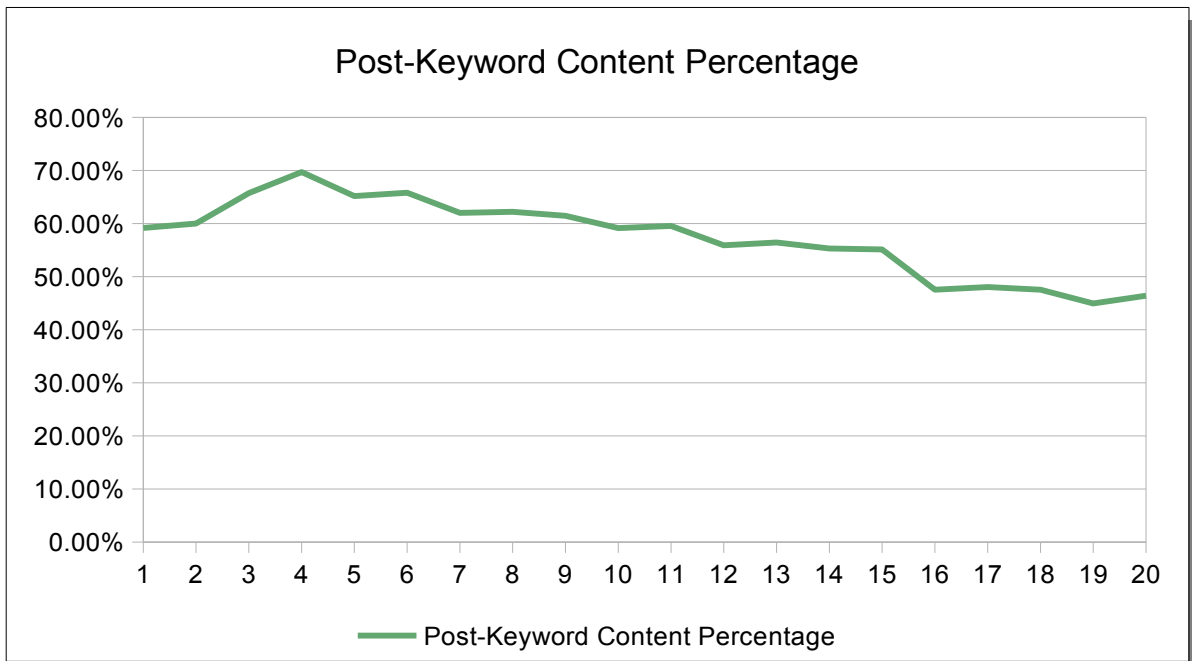


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.04	-0.01	0.001
content before 1 st kw found	18%	15.4%	22%	18%

In this analysis, the target keyword typically emerges after the initial 18% of content. Notably, for higher competition shorthead terms, this occurrence is expedited to just 15.4% into the content.

Last position of the Keyword in Visible content

This KPI quantifies the percentage of content subsequent to the final appearance of the target keyword within the text. Essentially, it gauges how late into the content the keyword is referenced for the last time. The assumption is: the later, the better.

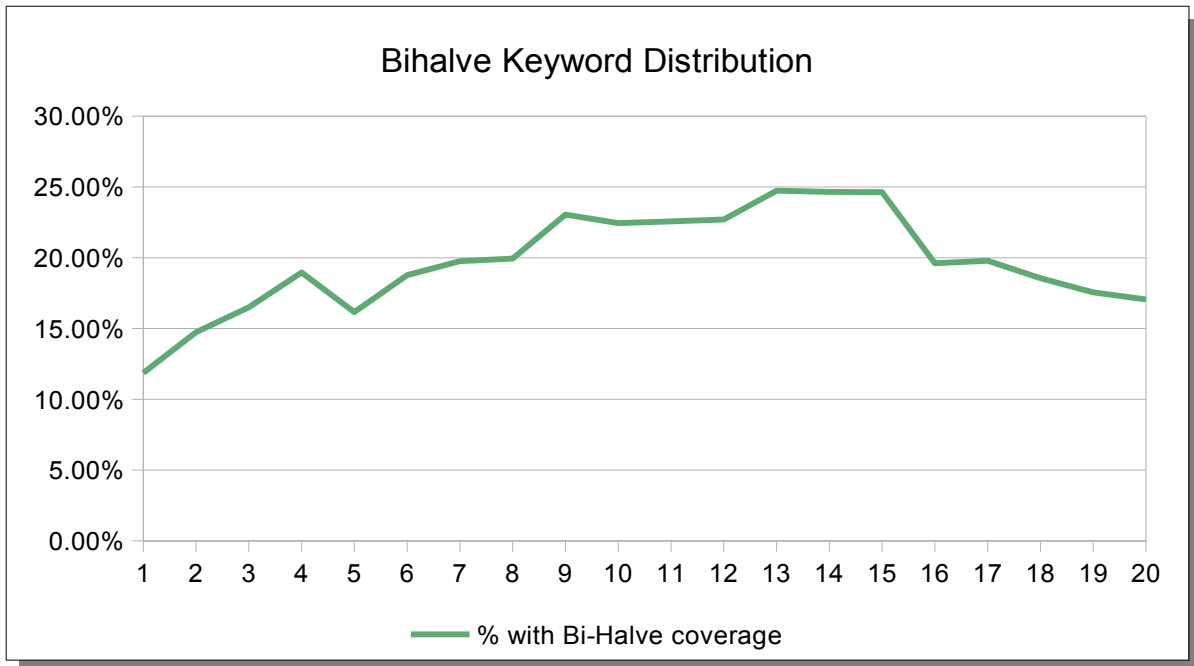


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.1	0.13	0.1	0.09
content after last kw found	53.4%	56%	52.7%	49.9%

On average, over 50% of page content follows the final appearance of the keyword, challenging the assumption that keywords frequently anchor both the beginning and end of content for Baidu's comprehension.

Bi-halve Keyword Distribution

In our analysis, we delve into the "Bi-halve Keyword Distribution" metric, examining the presence of the target keyword in both the first and second halves of the content, assessing its distribution consistency across the page span.

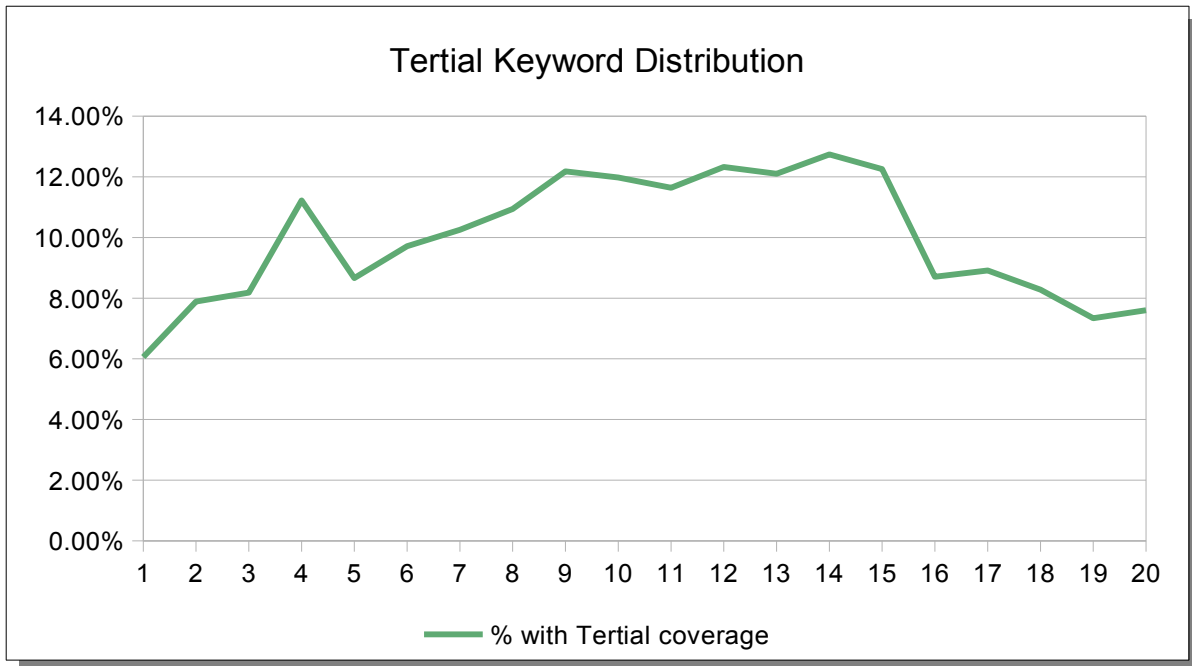


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	-0.05	-0.04	0.01
pages bi-halve usage	20%	31.3%	23.8%	9.8%

There's a subtle negative correlation between Bi-Halve keyword distribution and ranking. However, 31.3% of top-ranking short-head keywords employ this distribution, compared to 23.8% for mid-tail and a mere 9.8% for long-tail keywords.

Tertial Keyword Distribution

Expanding our scope, we introduce the "Tertial Keyword Distribution" metric, which evaluates the dispersion of the target keyword across all three tertials of the content, providing insights into its even spread throughout the page.

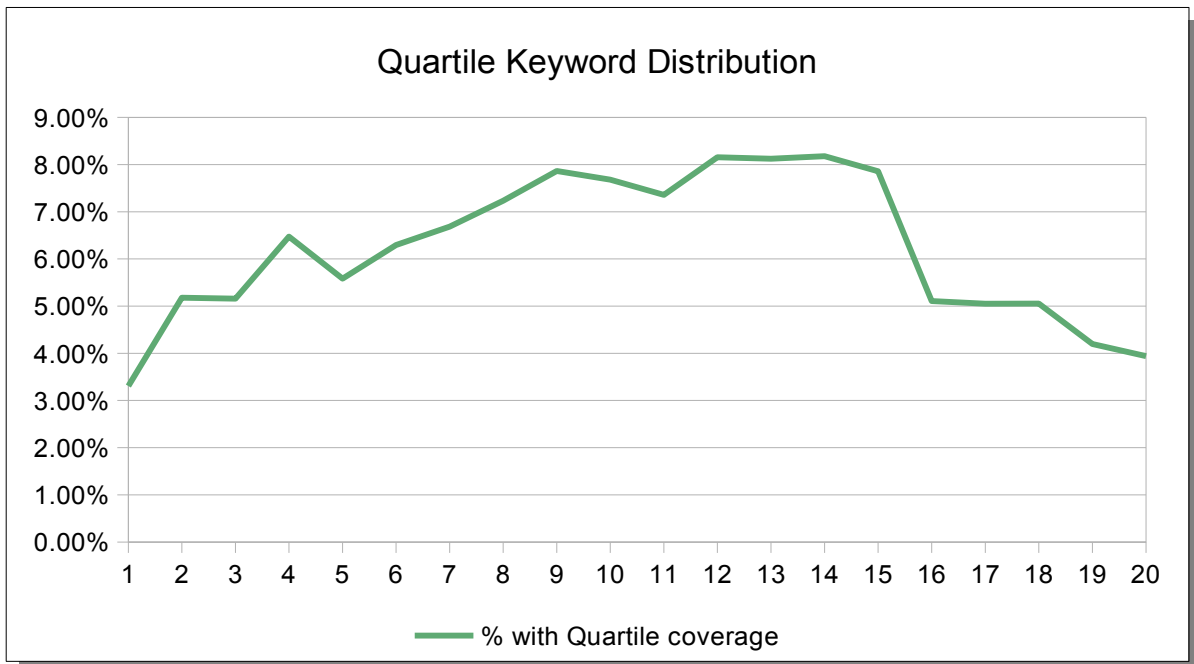


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.01	-0.01	-0.01	0.02
pages tertial usage	10.2%	18%	11.3%	3.9%

Tertial keyword distribution is observed in 10.2% of top-ranking pages. Notably, 18% of short-head keyword rankings employ this method, contrasted with 11.3% for mid-tail and a scant 3.9% for long-tail keywords.

Quartile Keyword Distribution

Further refining our analysis, we present the "Quartile Keyword Distribution" metric. This evaluates the presence of the target keyword across all four quartiles of the content, shedding light on its consistent distribution throughout the entirety of the page.

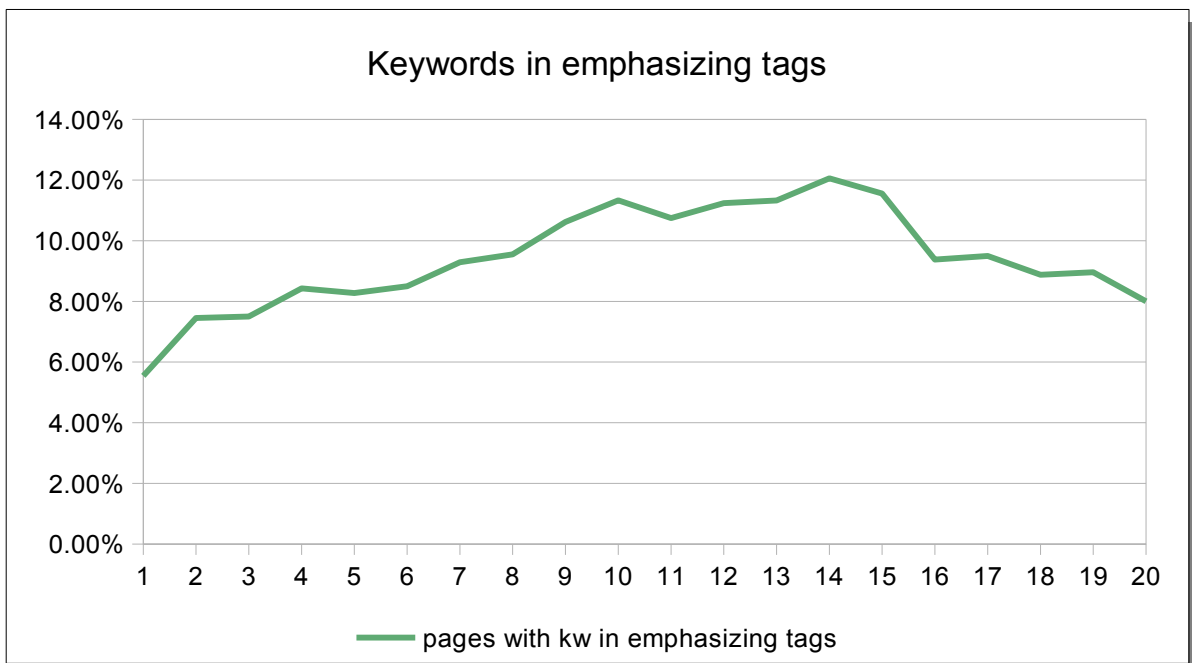


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.01	-0.003	0.01	0.02
pages quartile usage	6.4%	12.2%	6.9%	2.1%

Quartile keyword distribution appears in 6.4% of top-ranked pages. This distribution rises to 12.2% for short-head keyword rankings, while it's 6.9% for mid-tail and drops to 2.1% for long-tail keywords.

Keywords in emphasizing tags

Emphasizing tags are often employed to highlight keywords. This study examines their potential as ranking factors on Baidu, assessing whether their use can enhance visibility or influence positioning in search results. In this analysis we check on: , , <i>, , <u>, <mark>, , <ins>, <s>, ...



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.05	-0.02	0.003
pages w. emph. keywords	9.7%	13.8%	11.6%	5.3%

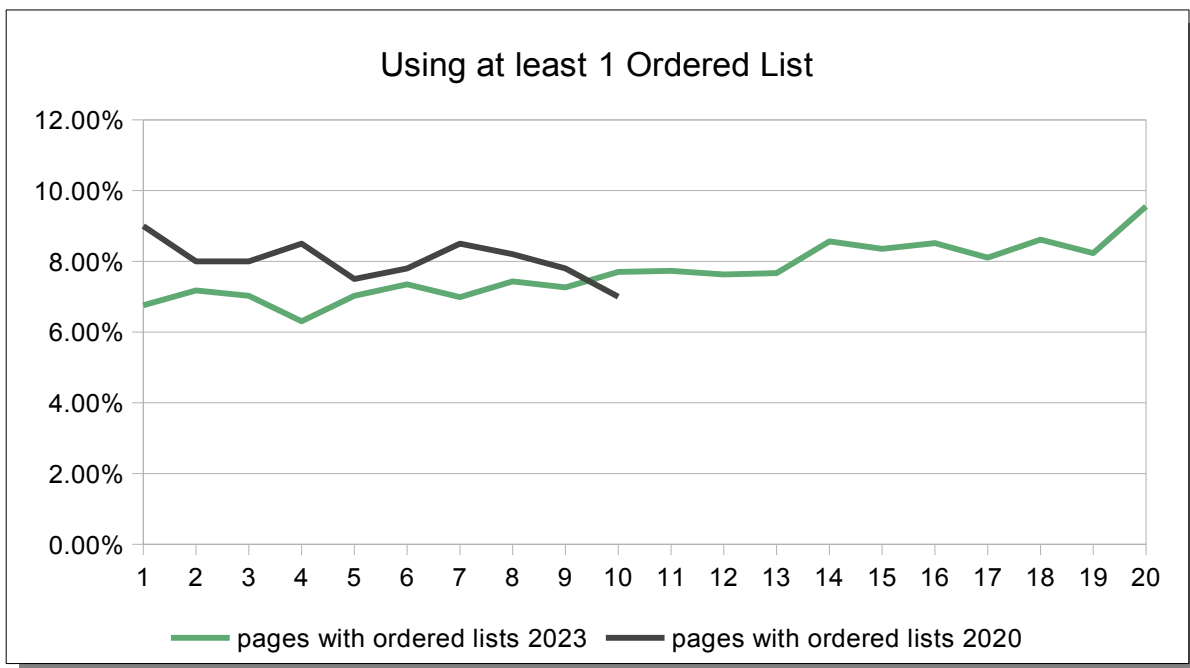
Despite a minor negative correlation between emphasizing tag usage and superior Baidu rankings, competitive keywords exhibit increased usage at 13.8%. This compares to 11.6% for mid-tail and a mere 5.3% for long-tail keywords.

Content Formats

Exploring diverse content formats, this section delves into the potential influence of images, tables, and both unsorted and sorted lists on Baidu rankings. We assess whether integrating these varied formats can enhance a page's standing in search results.

Ordered Lists

In this segment, we turn our attention to the `` tag, representing ordered lists. We evaluate its potential impact on Baidu rankings, gauging if structured, numbered lists might play a role in enhancing a page's search visibility.

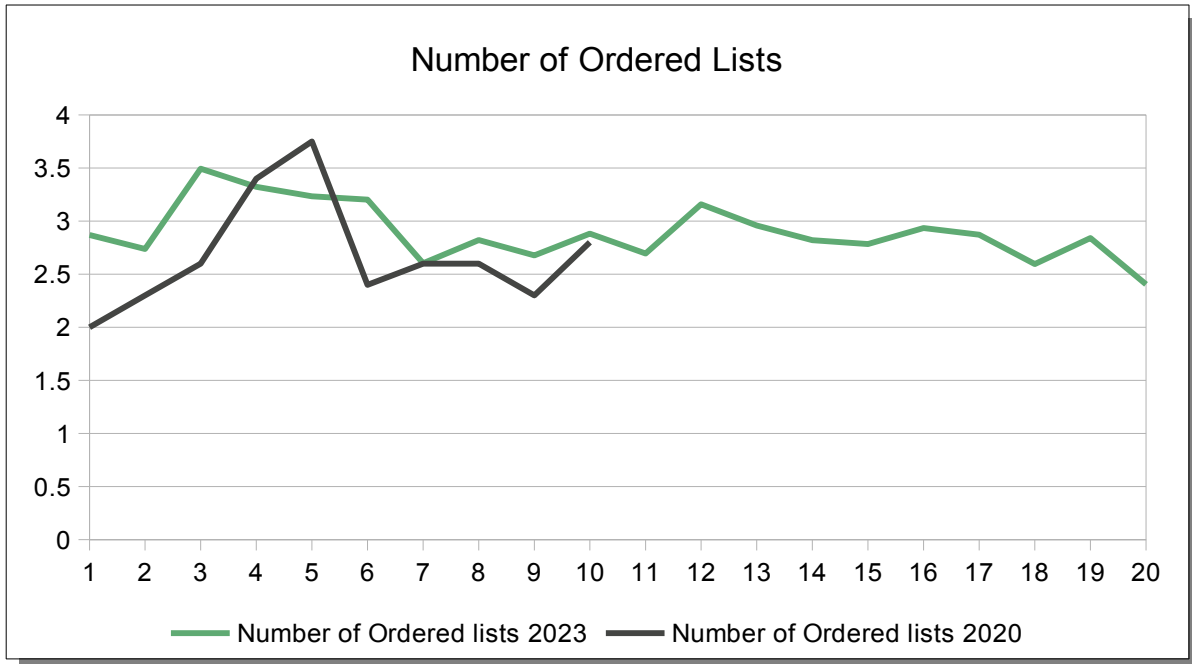


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.02	-0.03	-0.03
pages with <code></code>	7.7%	6.8%	7.5%	8.6%

A slight negative correlation exists between superior rankings and the utilization of ordered lists (). Notably, pages ranking for short-head keywords exhibit less usage at 6.8%, versus 7.5% for mid-tail and 8.6% for long-tail keywords.

Number of Ordered Lists

Among those pages utilizing ordered lists, we also liked to know how many of them are used:

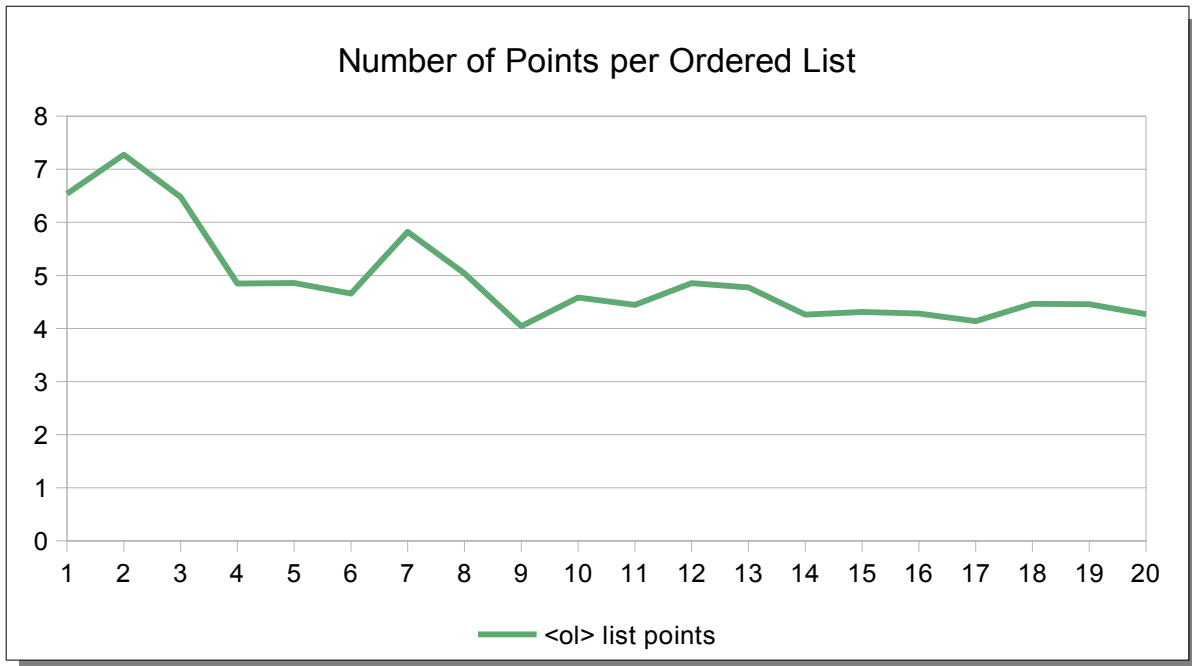


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.05	0.04	0.08	0.02
ordered lists number	2.9	2.7	2.9	2.9

A discernible trend emerges from our data: Not only is there a slight negative correlation between the presence of lists and superior Baidu rankings, but pages targeting competitive keywords also tend to feature fewer ordered lists, averaging 2.7 versus 2.9 on less competitive pages. This might suggest a cautious approach to using ordered lists for SEO.

Number of Points per Ordered List

The factor "list points per ordered list" evaluates the average number of items within each ``, assessing its potential influence on Baidu rankings. This metric delves into whether longer or shorter ordered lists may impact SEO performance.

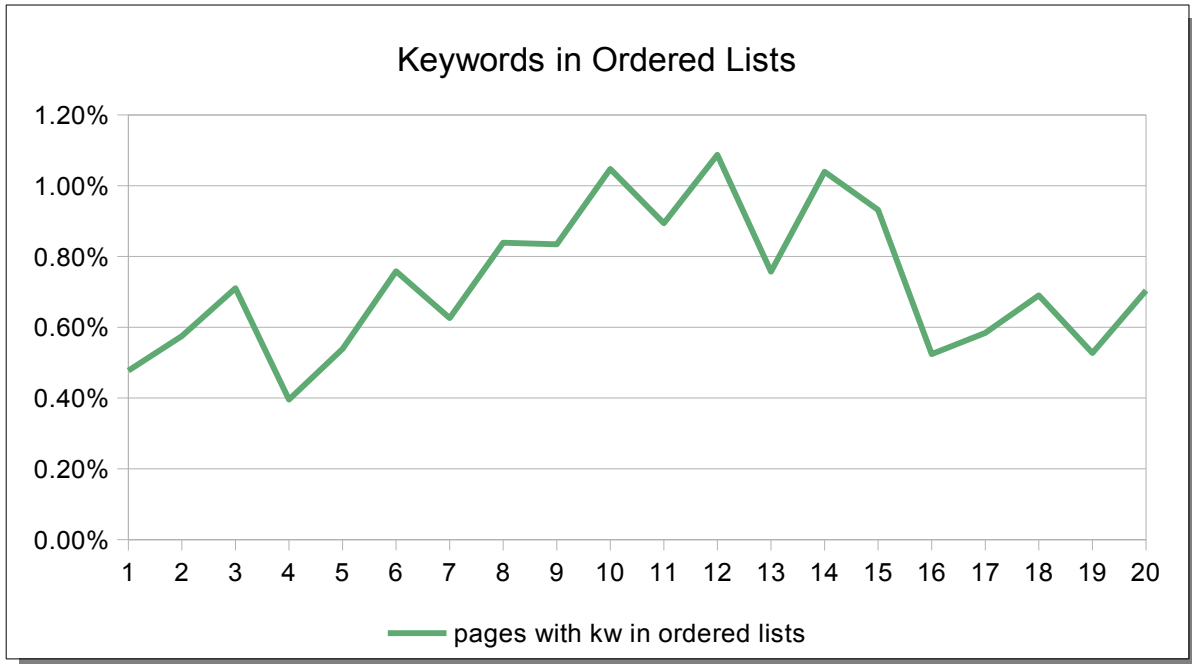


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.04	0.09	-0.01	0.05
<code></code> per <code></code>	4.7	4.5	4.9	4.7

On average, each ordered list contains approximately 4.7 items, a consistent trend observed across short-head, mid-tail, and long-tail keywords.

Keyword Usage in Ordered Lists

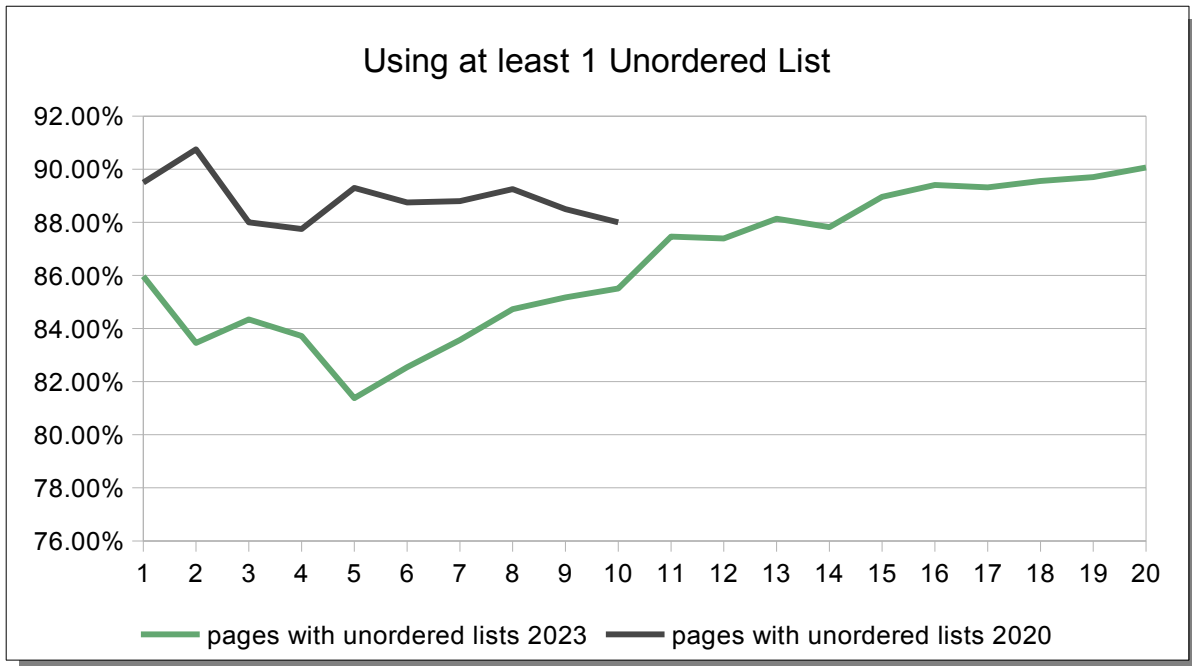
The potential factor under scrutiny is the inclusion of target keywords within list items of ordered lists, gauging its impact on Baidu rankings.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.003	-0.01	-0.003	0.003
keywords in 	0.7%	1.1%	0.9%	0.4%

Unordered Lists

In this segment, we turn our attention to the `` tag, representing ordered lists. We evaluate its potential impact on Baidu rankings, gauging if structured, numbered lists might play a role in enhancing a page's search visibility.



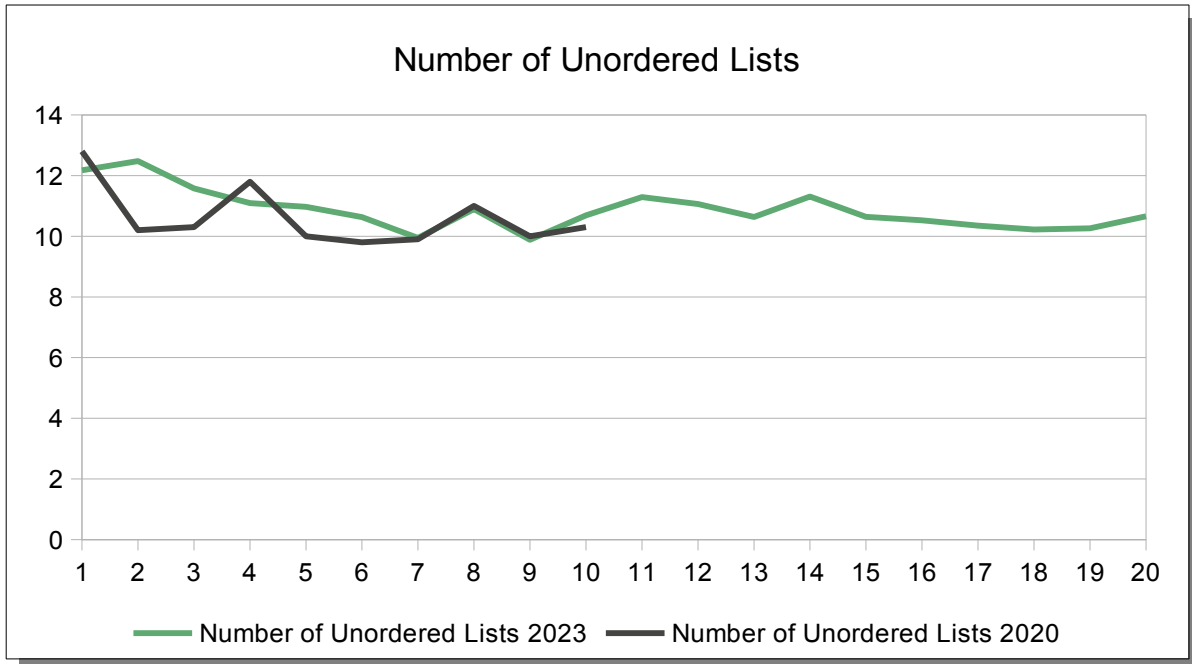
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.07	-0.06	-0.09	-0.06
pages with <code></code>	86.5%	86.2%	86.6%	86.7%

A subtle negative correlation is observed between higher rankings and the presence of unordered lists (``). Specifically, pages contending in short-head keyword categories show a minimal decrease in `` implementation, accounting for 86.2%, in contrast to mid-tail at 86.6% and long-tail at 86.7%.

Given that an overwhelming majority of 86.5% of top-ranked pages incorporate , it suggests that omitting may be perceived unfavorably by Baidu's ranking mechanisms.

Number of Unordered Lists

Among those pages utilizing unordered lists, we also liked to know how many of them are used:



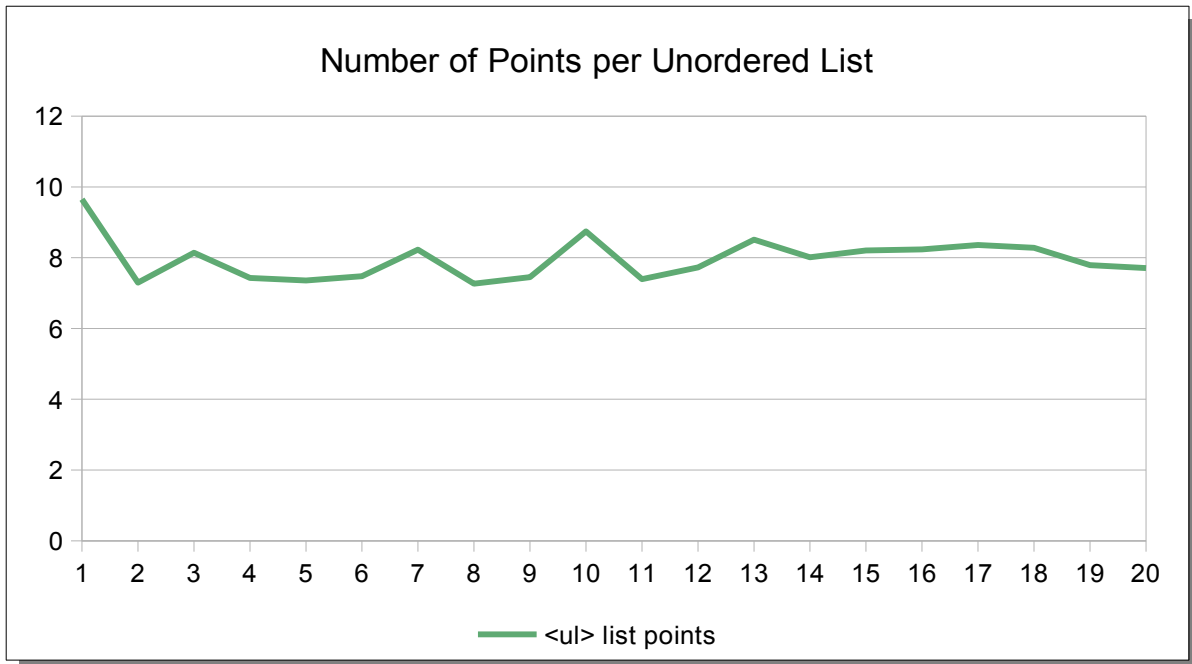
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.05	-0.05	-0.06	-0.06
ordered lists number	10.8	10.4	10.8	11.1

There's a minor negative correlation between the use of unordered lists () and higher Baidu rankings. Pages ranking for competitive keywords typically display fewer unordered lists, with an average of 10.8 compared to 11.1 for less contested terms.

This trend implies a strategic use of up to approximately 10 unordered lists could be beneficial for pages aiming to rank in competitive searches.

Number of Points per Unordered List

The factor "list points per unordered list" evaluates the average number of items within each ``, assessing its potential influence on Baidu rankings. This metric delves into whether longer or shorter ordered lists may impact SEO performance.

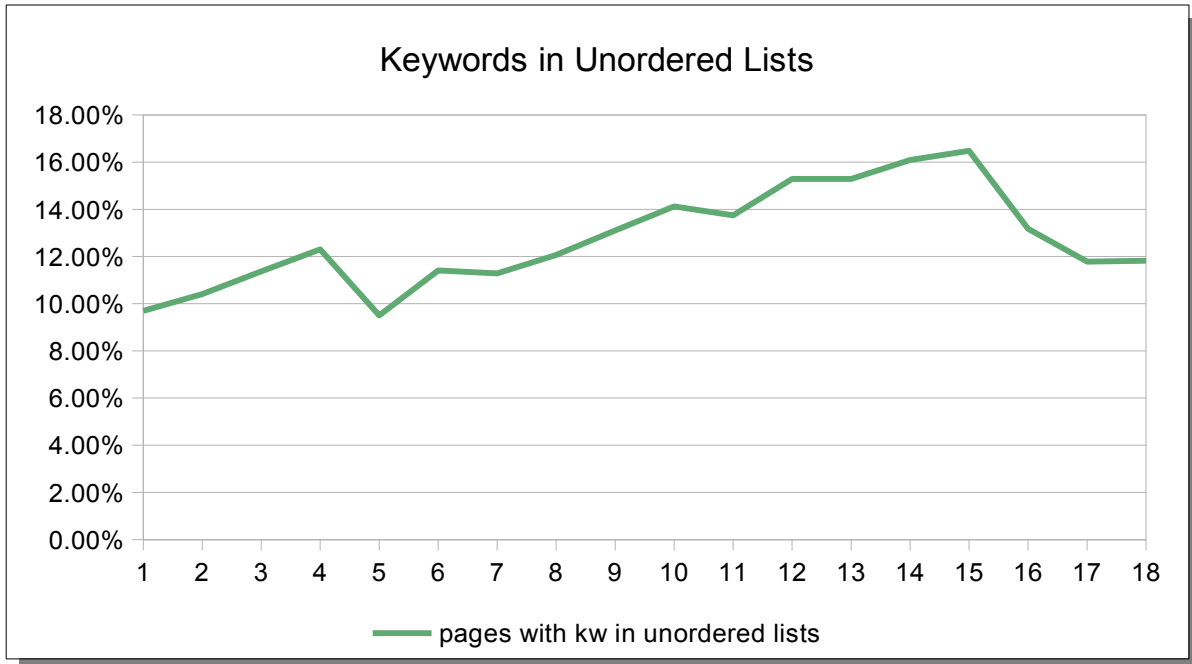


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.09	-0.1	-0.1	-0.07
<code></code> per <code></code>	7.9	8.8	7.6	7.6

On average, each unordered list contains approximately 7.9 items, with pages ranking for short-head keywords having more list points than pages ranking for mid-tail, and long-tail keywords.

Keyword Usage in Unordered Lists

The potential factor under scrutiny is the inclusion of target keywords within list items of unordered lists, gauging its impact on Baidu rankings.

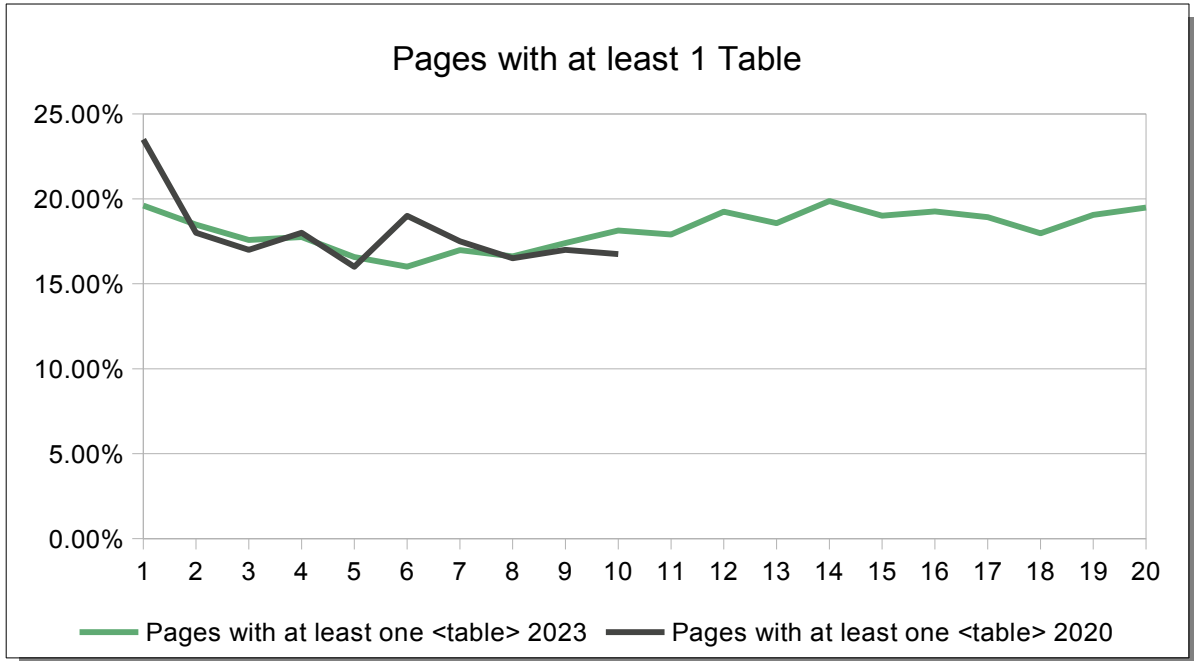


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.06	-0.03	0.02
keywords in 	12.9%	19.5%	14.9%	6.7%

Averaging at 12.9%, a modest proportion of top-ranking pages incorporate the keyword within unordered lists. Despite a faint negative correlation with higher Baidu rankings, keyword presence in unordered lists is more prevalent in pages ranking for short-head keywords at 19.5%, diminishing to 14.9% for mid-tail and 6.7% for long-tail keyword rankings. This suggests a nuanced approach to keyword integration within unordered lists, tailored to the competitive nature of the keyword.

Tables

Assessing HTML tables as a ranking factor, we examine if their presence on a page influences Baidu rankings.

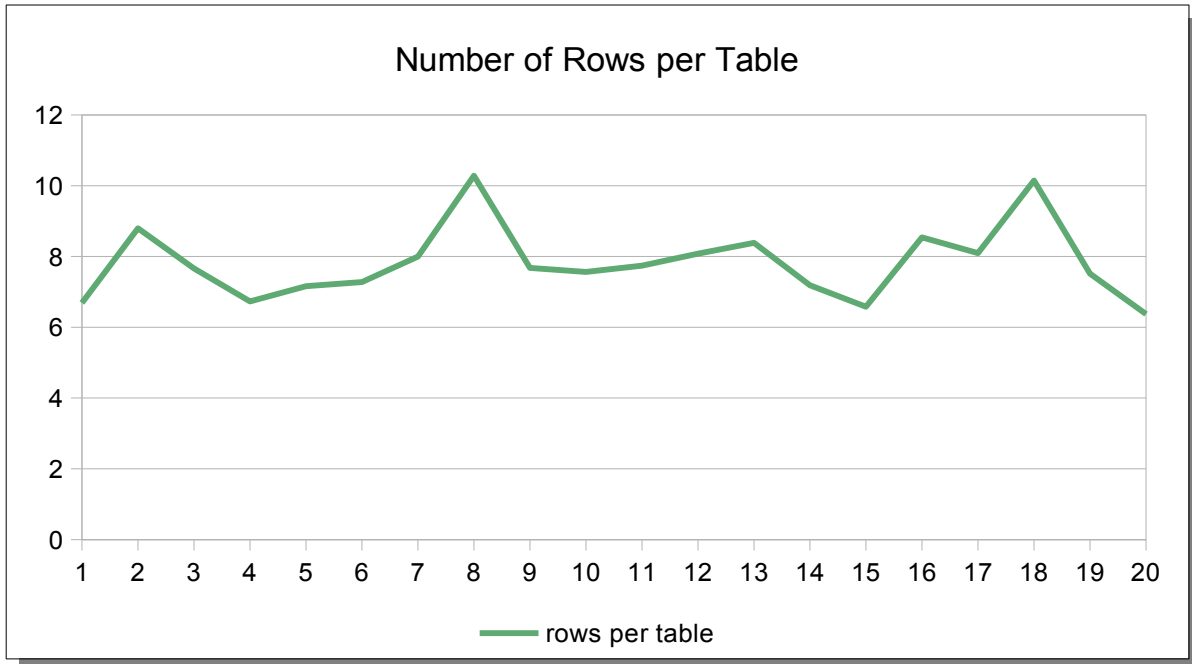


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	0.03	-0.01	-0.06
pages with <table>	18.2%	17.8%	17%	19.4%

Despite a modest negative correlation, tables appear in 18.2% of top-ranking pages, with the highest-ranked positions showing a slightly more frequent use. Short-head keyword pages incorporate tables in 17.4% of cases, indicating their not insignificant role.

Rows per Table

Among the used tables, let's identify if there are correlations between number of rows and rankings on Baidu.

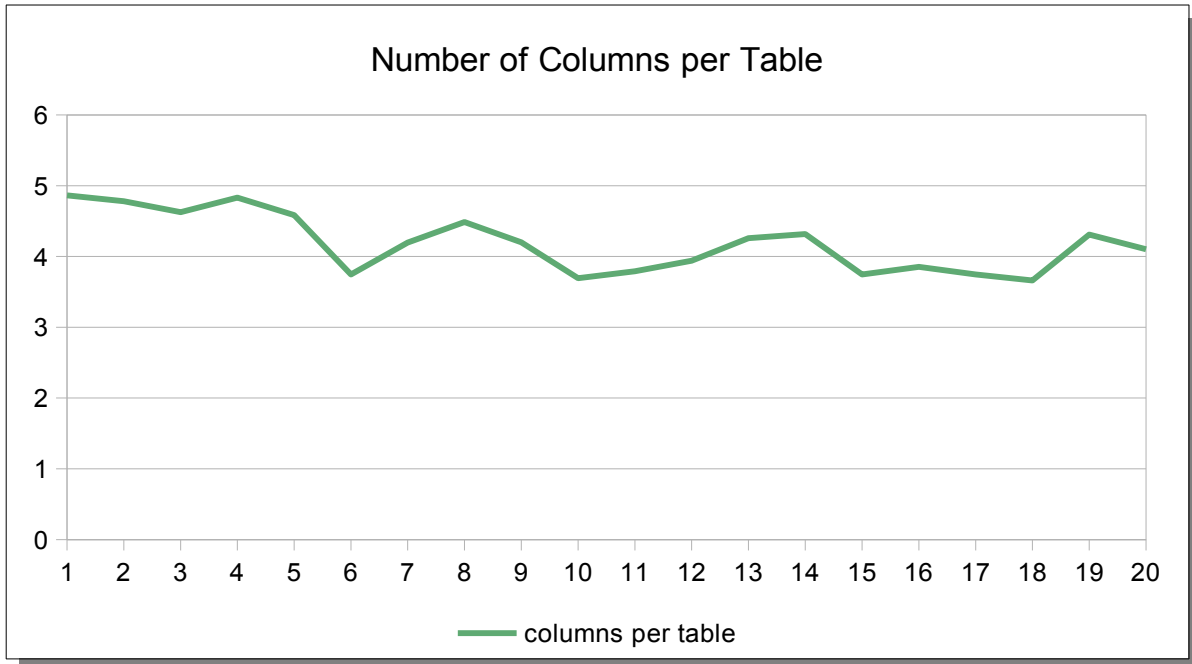


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.04	0.05	0.07	0.02
<tr> per <table>	7.8	8.3	7.5	7.9

The consistency in table structure is evident, with an average of 7.8 rows per table across the board, showing no significant variation between short-head and long-tail keyword rankings.

Columns per Table

Among the used tables, let's identify if there are correlations between number of columns and rankings on Baidu.

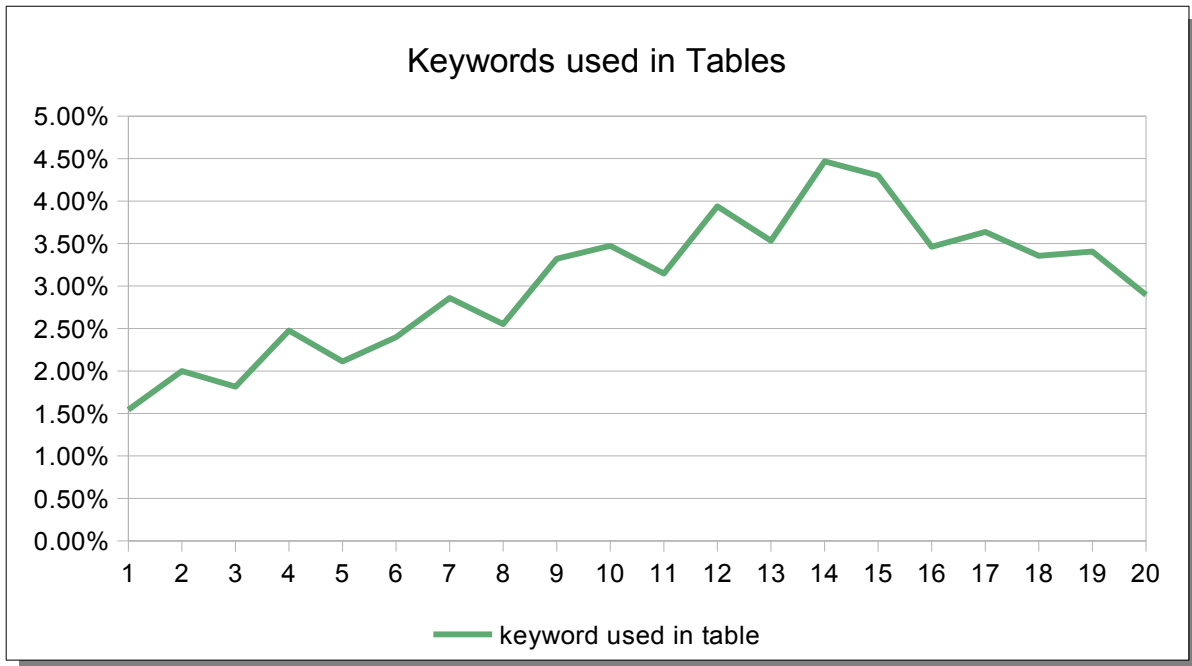


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.07	0.09	0.1	0.03
<td> per <tr>	4.1	4.8	4	3.7

The data indicates an average of 4 columns per table, with a discernible trend: higher-ranking pages and those targeting short-head keywords typically feature a greater number of columns, aligning with a slight positive correlation observed in the rankings.

Keyword usage in Tables

This metric evaluates whether incorporating target keywords within HTML table elements influences a page's ranking on Baidu. It considers the prevalence and impact of keyword deployment within tabular data on search visibility.



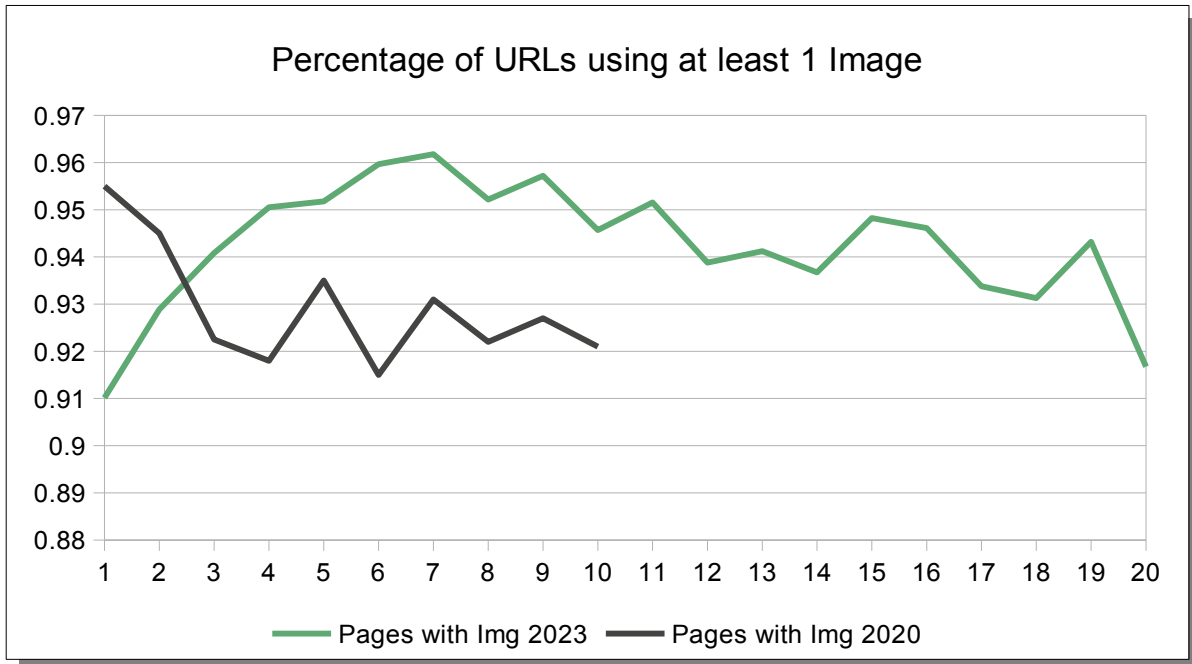
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.03	-0.04	-0.04	-0.03
keyword in tables	3.1%	3.7%	3.7%	0.23%

Of high-ranking pages, only a small fraction—3.1%—incorporate keywords in tables, with a slight uptick to 3.7% for competitive short-head keywords.

Considering that tables appear on 18.2% of these pages, the keyword presence in tables is a notable but not widespread tactic.

Images

The factor in focus is the integration of images within web content, specifically evaluating whether the inclusion and optimization of images correlate with enhanced rankings on Baidu. This analysis will discern the impact of visual elements on a page's search performance.

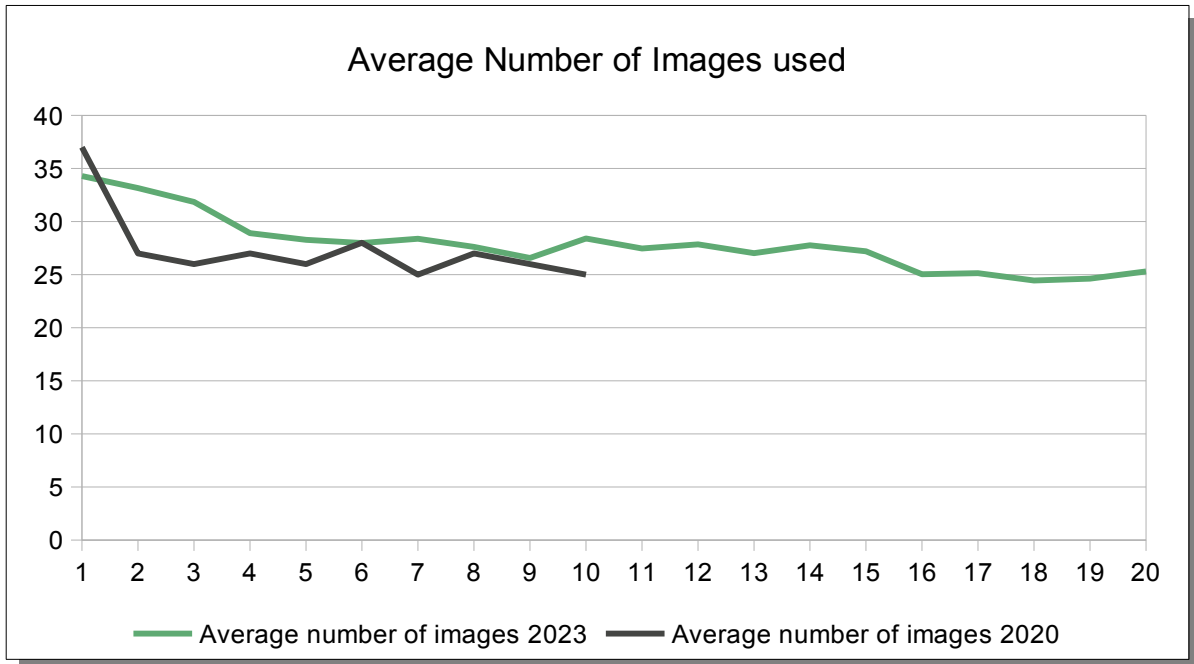


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.01	0.01	0.02	0.03
pages with 	94.4%	94%	95.2%	95.4%

In a slight uptick from 2020's findings, 94.4% of top-ranking pages on Baidu now feature images, suggesting a near-universal adoption of visual content. Despite a minimal correlation coefficient of 0.01%, the prevalence of images among high-ranking pages highlights their importance in content strategies.

Number of Images

The strategic incorporation of images across web content is considered essential for user engagement. However, from an optimization standpoint for Baidu, the pertinent question is: What constitutes the optimal quantity of images that signals a user-centric page design to the search engine?

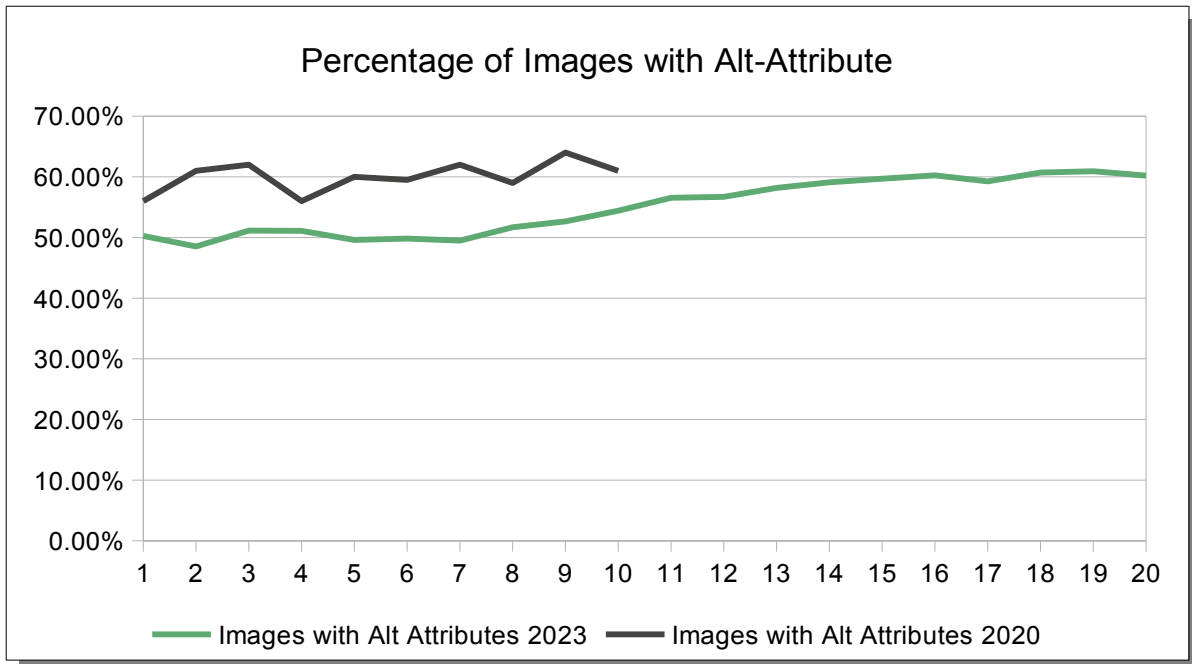


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.01	-0.03	0.02	0.06
number of images	27.5	28.6	28.1	26.4

In line with 2020 data, the average image count per page stands at 27.5. A slight positive correlation suggests better rankings with more images—top-ranked pages average 34 images, dipping to 25 for lower ranks. Short-head keyword pages often have more images (28.6) compared to mid-tail (28.1) and long-tail (26.4), indicating a trend towards visual richness correlating with higher rankings.

Images with Alt-Attributes

The integration of alt attributes in images is scrutinized as a potential ranking factor, considering its role in enhancing accessibility and providing context to search engines, potentially influencing a page's visibility on Baidu.

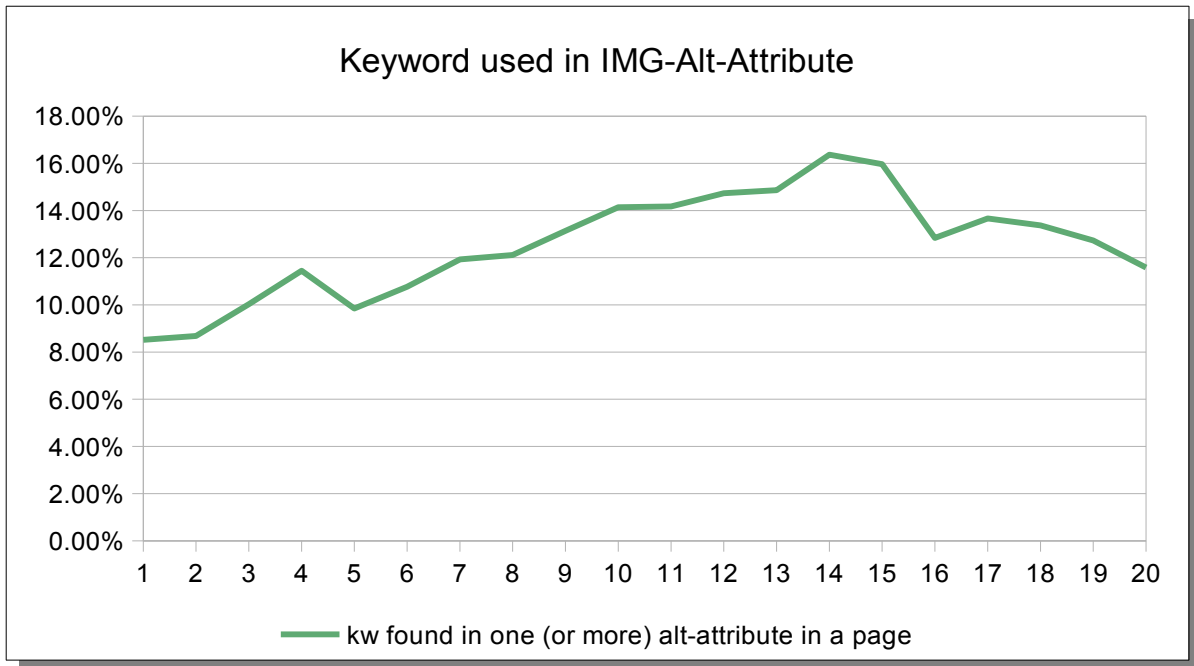


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.1	-0.1	-0.1	-0.09
images with alt	55.4%	55.5%	56%	54.8%

The investigation reveals that while 55.4% of top 20 ranking pages on Baidu incorporate alt text within image tags, the prevalent application of this attribute does not necessarily yield an association with enhanced rankings, as evidenced by a negative correlation.

Keywords in Alt-Attributes

Exploring the impact of keyword inclusion in image alt attributes on Baidu's search rankings, this analysis scrutinizes whether embedding target keywords within alt descriptions offers a tangible SEO advantage.



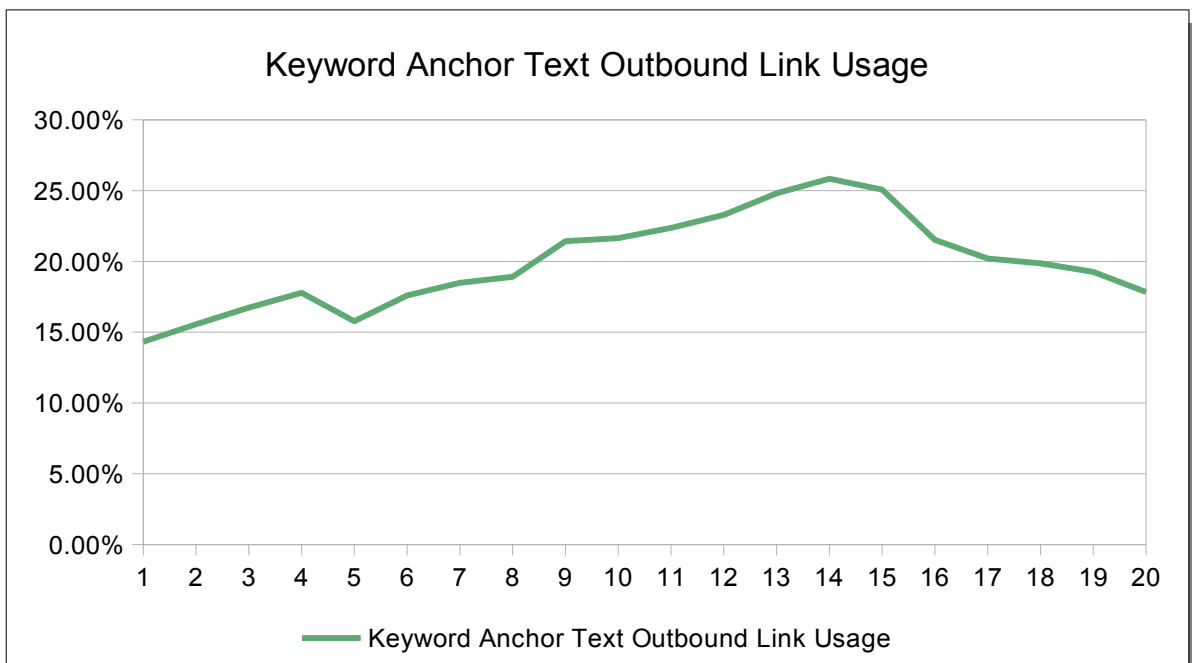
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.04	-0.1	-0.04	0.02
kw in alt	12.8%	19.7%	15%	6.4%

Approximately 12.8% of top-ranking pages on Baidu use the target keyword in image alt attributes, with a minor negative impact on rankings observed.

Nonetheless, pages with short-head keywords show higher usage at 19.7%, versus 15% for mid-tail and 6.4% for long-tail keywords, indicating competitive keywords might benefit from keyword-tagged images.

Keywords in Anchortext (within the ranking page)

This aspect turns its attention to the significance of keywords within the anchor text of links pointing to other pages, both within the same website and externally. We examine top-ranking pages to determine if their outbound links incorporate the same keywords for which they rank. Our hypothesis posits that the use of identical keywords in the anchor text of outbound links from a page already ranking for those keywords could potentially have a negative impact.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation (on not having it) kw in outbound link	0.04	0.08	0.07	-0.009
	20.3%	30.9%	23.4%	10.7%

In our analysis, we found a positive correlation between better page ranking and the avoidance of using outbound links with anchor text containing the same keyword. This aligns with our initial hypothesis. However, an intriguing pattern

emerged concerning the specificity of the keywords. As the keywords become more 'longtail' or specific, they are less likely to be part of the anchor text in outbound links. Conversely, 'shorthead' or broad keywords appear more frequently in such contexts.

One possible interpretation of this observation is that the more specific a keyword (and thus the webpage it ranks for), the less necessity there is to link out using that keyword. This could be because the page's content is already comprehensive and specific enough to satisfy a user's query. On the other hand, pages ranking for broader, shorthead keywords might benefit from linking out to more detailed subtopic pages.

These subtopic pages often incorporate the shorthead keyword within their midtail or longtail target keywords, providing a more detailed exploration of the topic at hand. Hence, the usage of shorthead keywords in outbound link anchor texts could be a strategic move to guide users towards more specific, related content.

We can try proving this hypothesis by looking at keyword length, which in Chinese is often an indicator for more broad or specific terms:

	1-2 Characters length	3-4 Characters length	5-6 Characters length	7 and more Characters length
correlation	0.03	0.04	0.05	0.06
char. coverage	55%	36%	20.3%	12.8%

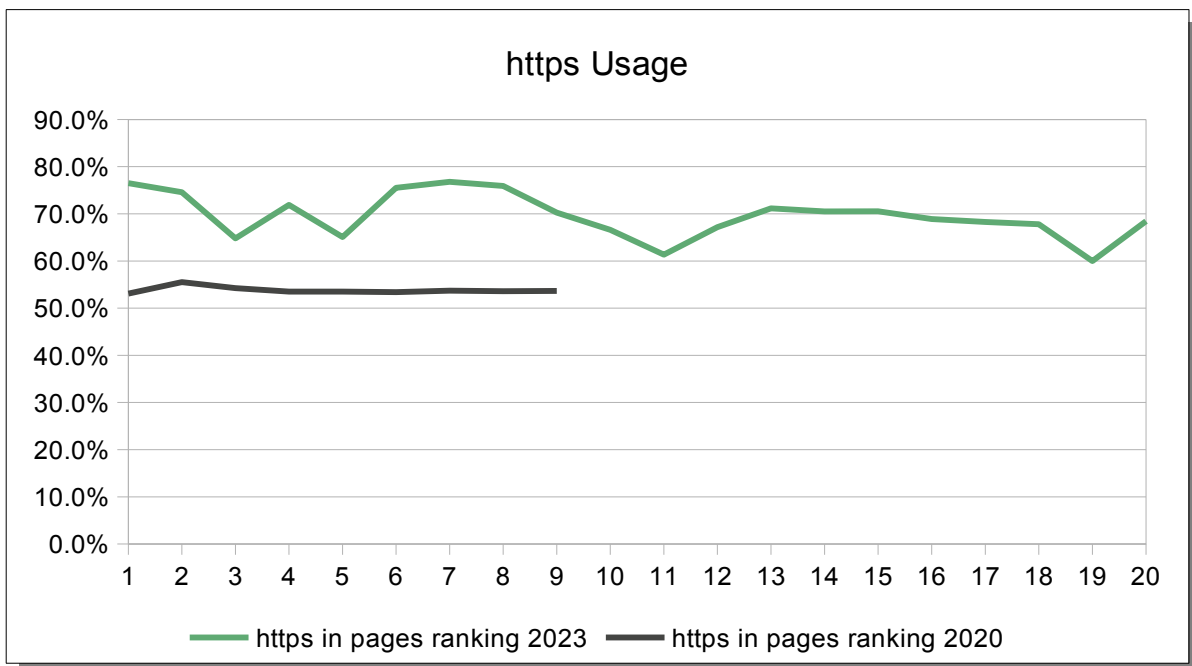
The data suggests a clear trend: as the length of the keyword (and thus its specificity) increases, the likelihood of it being used in an outbound link's anchor text decreases. This reinforces our hypothesis that more specific terms are less commonly used in such contexts, potentially due to their comprehensive and detailed nature.

Technical Considerations

In this chapter, we focus on the technical elements of Search Engine Optimization (SEO). We explore how they can influence a website's ranking. For technical correlations, we do not look at different search volume groups, as we consider them as basic optimizations, a website should follow.

https Usage

The adoption of HTTPS as a technical ranking factor has been under scrutiny since Baidu initiated crawling of HTTPS pages in 2017 (<https://www.jademond.com/magazine/baidu-algorithm-updates/#https>). This metric gauges the impact of secure socket layers on SERP positions, considering Baidu's evolving algorithmic preferences for web security.



	Whole Keyword Set
correlation	0.1
https usage	69.6%

The incremental shift from HTTP to HTTPS among Baidu's top-ranking pages is evident, with a present adoption rate of 69.6%—a substantial rise from the 53.7% observed in 2020. This trend correlates positively with ranking positions, hinting at a preference within Baidu's algorithm for SSL-secured domains, especially among the highest echelons of search results.

Expert comments:



“The shift from HTTP to HTTPS websites on Baidu shouldn't surprise anybody in the digital marketing industry. Regardless of this being a formal ranking factor or not, all websites this day in age should be securing their websites with HTTPS. Not only does it enhance security, but improves data integrity, improves performance, and can be a major trust factor for users.”

Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com

“Baidu strongly recommends the use of HTTPS for websites, signaling that HTTPS certification is a crucial factor for search rankings and visibility, as indicated in Baidu's webmaster tools.”*

Qing He, Head of Search, Jademond Digital, www.jademond.com



**editors note: “Baidu Webmaster Tools” is the term we refer to in the West, while in the original it is Baidu Ziyuan (ziyuan.baidu.com), which refers to the tool comparable to Baidu Search Console, but also to the information platform, where Baidu announces information on updates or initiatives.*



“While still far behind the rest of the web, it's encouraging to see an uptick in HTTPS usage in China over the past 3 years.”

Simon Lesser, Cofounder and CEO, Dragon Metrics,
www.dragonmetrics.com



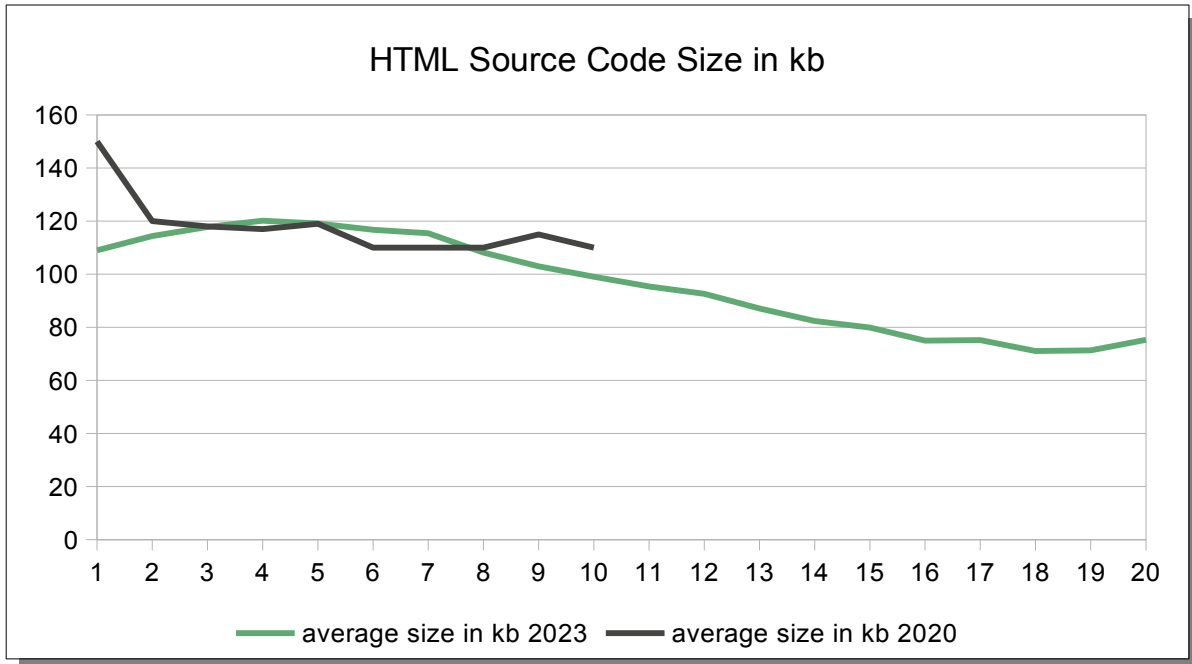
“Although Website security is a ranking factor on Baidu, many local websites, particularly B2B corporate and news sites without transactional features, continue to rely on the HTTP protocol.

While the transition to HTTPS is undoubtedly inevitable, its widespread adoption may require time to become more prevalent among diverse websites in China.”

Christina Xu, Founder of China Marketing Corp, www.chinamarketingcorp.com

HTML file size in kb

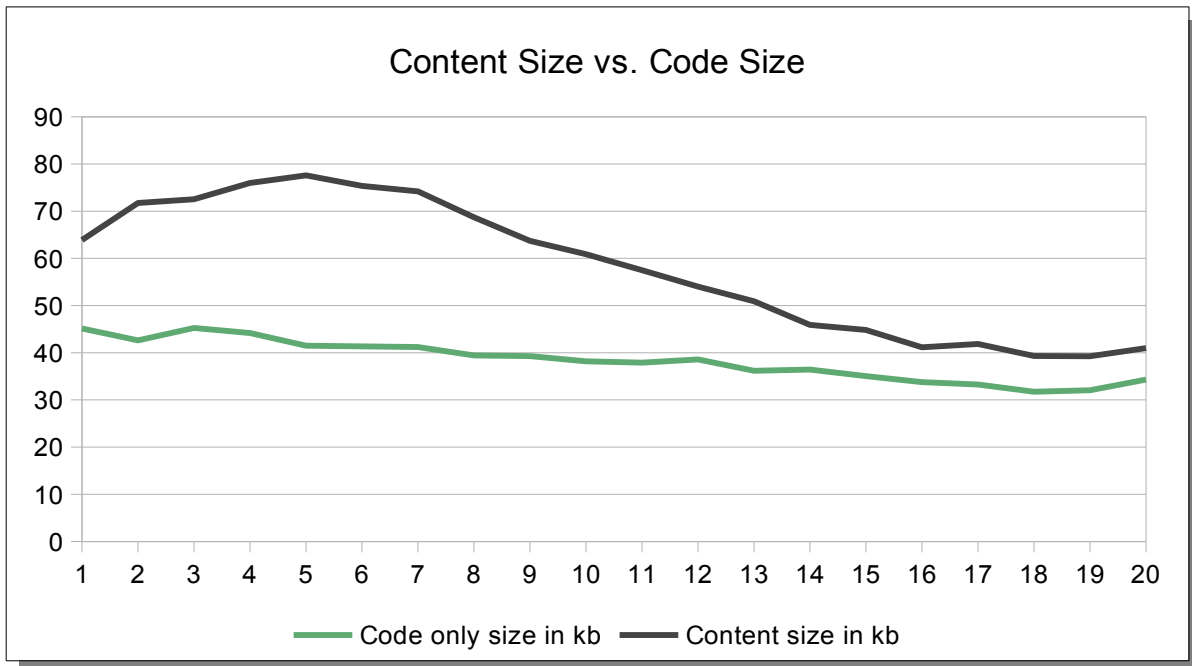
Looking at HTML size, measured in kilobytes (KB), assesses whether the data footprint of a webpage influences its search ranking on Baidu.



Whole Keyword Set	
correlation	0.2
https usage	96.4kb

In examining HTML size as a potential ranking factor on Baidu, a modest positive correlation has been observed with larger HTML files correlating to improved rankings. Current data indicates an optimal HTML size around 120KB, a reduction from the 2020 peak of 150KB. This shift prompts further investigation into whether content depth or coding complexity drives this trend.

Code Size versus Content Size in kb



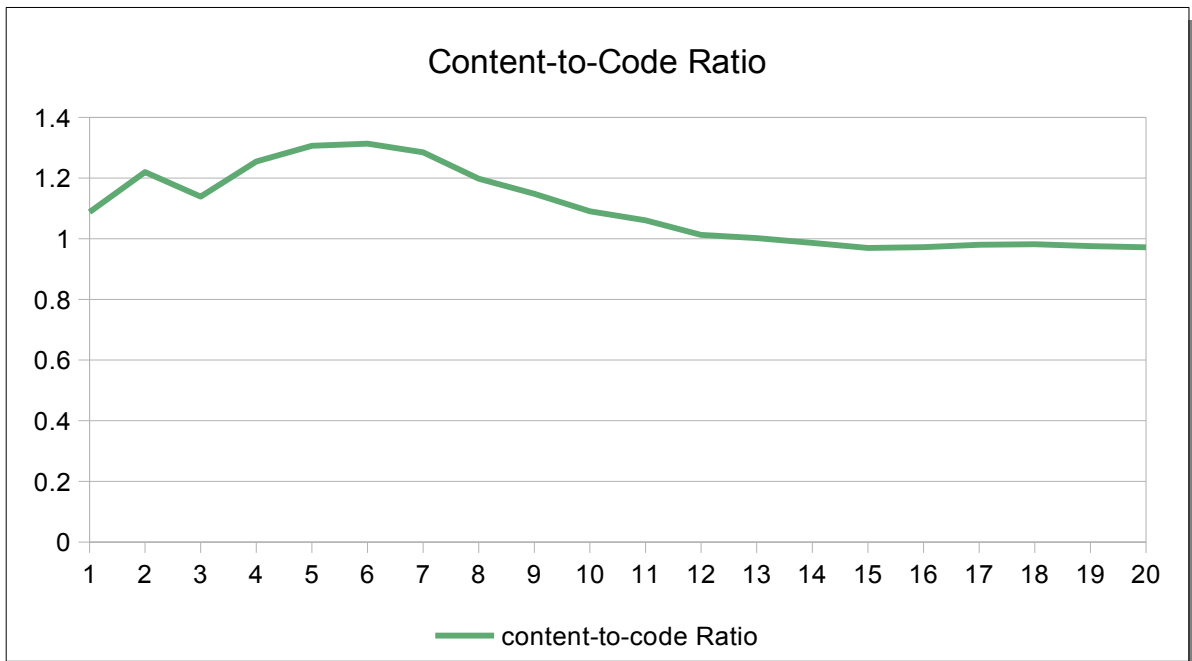
	Content Size	Code Size
correlation	0.1	0.1
size in kb	58	38.4

The data indicates that pages ranking higher on Baidu typically feature both more extensive content and more complex code.

However, the disparity in content volume between top 5 positions and those on the second page is more pronounced than that of code, hinting that while intricate templates may not significantly influence rankings, richer content appears to be a decisive factor.

Content-to-Code Ratio

The content to code ratio, as a potential ranking factor, refers to the amount of visible text (content) versus the underlying HTML code on a webpage, with a higher ratio often indicating a more content-rich site that could be favored by search engines.

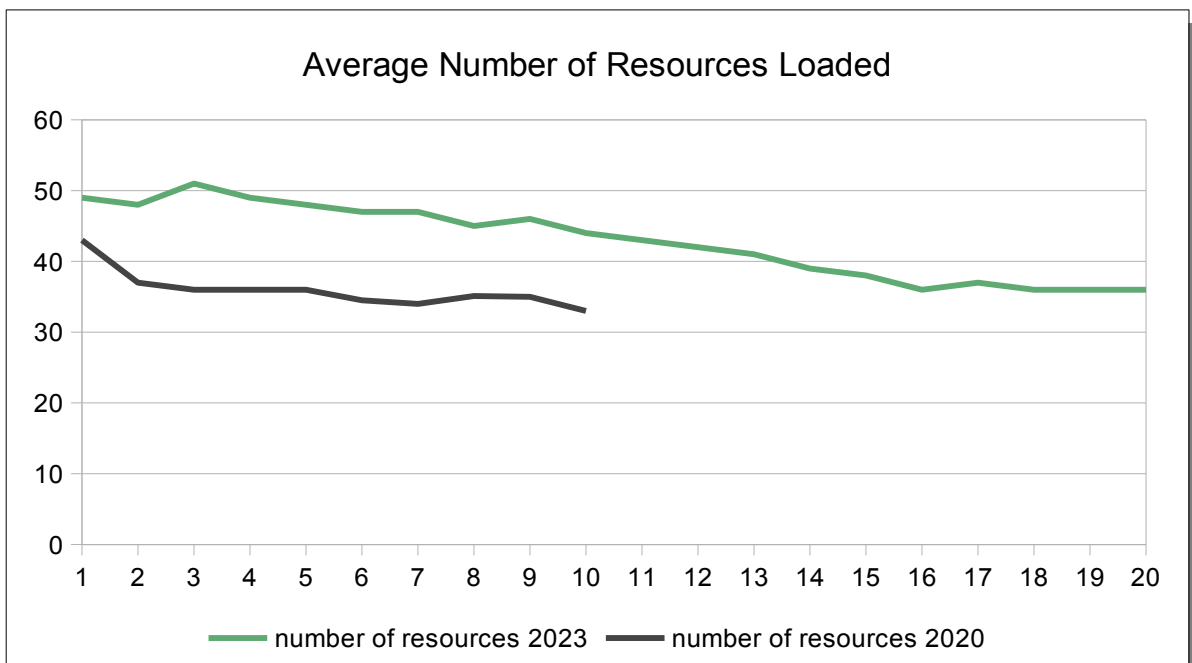


Whole Keyword Set	
correlation	0.01
content-to-code ratio	1.1

Examining the content-to-code ratio as a potential ranking factor for Baidu, our data indicates a marginal yet positive correlation with enhanced search rankings. The trend persists across keywords of varying competitive levels, suggesting that a higher content relative to code may indeed be a signal of quality in Baidu's evaluation of web pages. That basically supports what we saw earlier already.

Average Number of Resources loaded

The "Number of Resources Loaded" metric examines the quantity of individual elements fetched by the browser, such as images, CSS files, and scripts. This factor is scrutinized for its potential impact on Baidu's rankings, as it may influence page load times or indicate a media-rich user experience. Our correlation method assumes that less resources loaded result in faster loading time of the website, which is better for the user and for rankings.



Whole Keyword Set	
correlation	0.01
number of resources	53.6

In an unexpected turn, data indicates a positive correlation between the number of resources loaded and superior Baidu rankings. Contrary to initial assumptions that fewer resources might enhance page performance and thus rankings, the

average resource count has notably risen since 2020 among higher-ranking pages.

Expert comment:

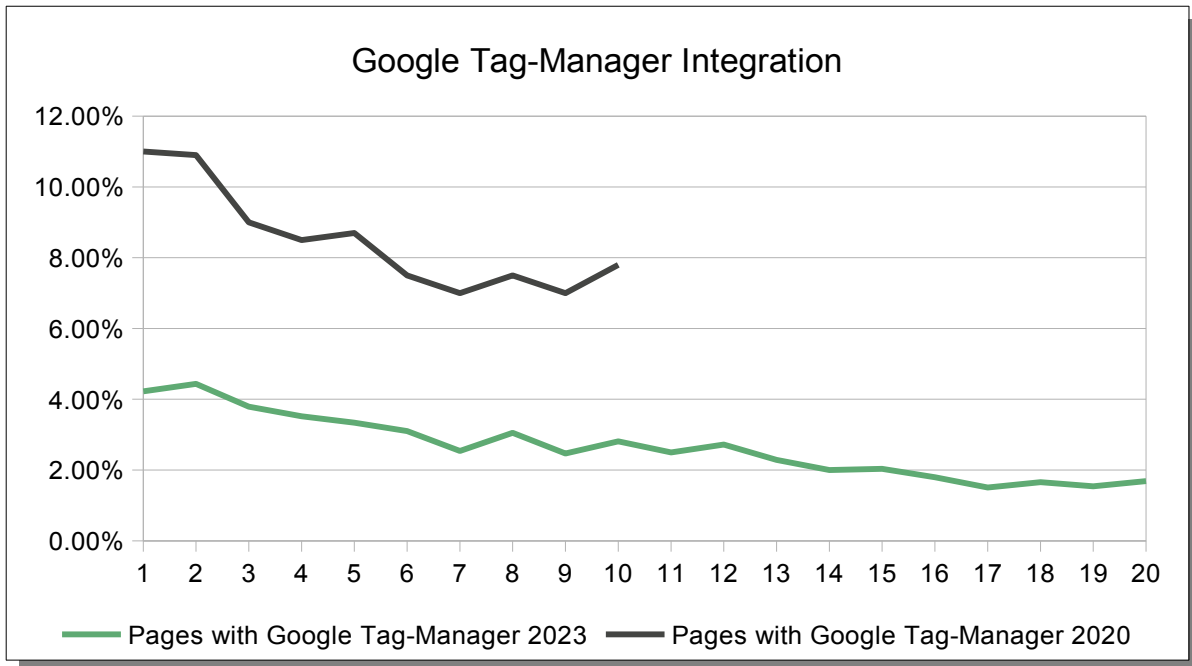


“Here we see the Chinese web following the unfortunate global trend of resource bloat. As the largest sites on the web typically also being the most bloated, I wonder if Baidu's "Winner-takes-all" approach with the dominance of the top 500 sites is responsible for the correlation between higher rankings and more resources loaded.”

Simon Lesser, Cofounder and CEO, Dragon Metrics, www.dragonmetrics.com

Google Tag-Manager Integration

The integration of Google Tag Manager emerges as a potential ranking factor amidst the discourse among China SEO experts, who debate its impact due to the perceived blockage of Google services in China.



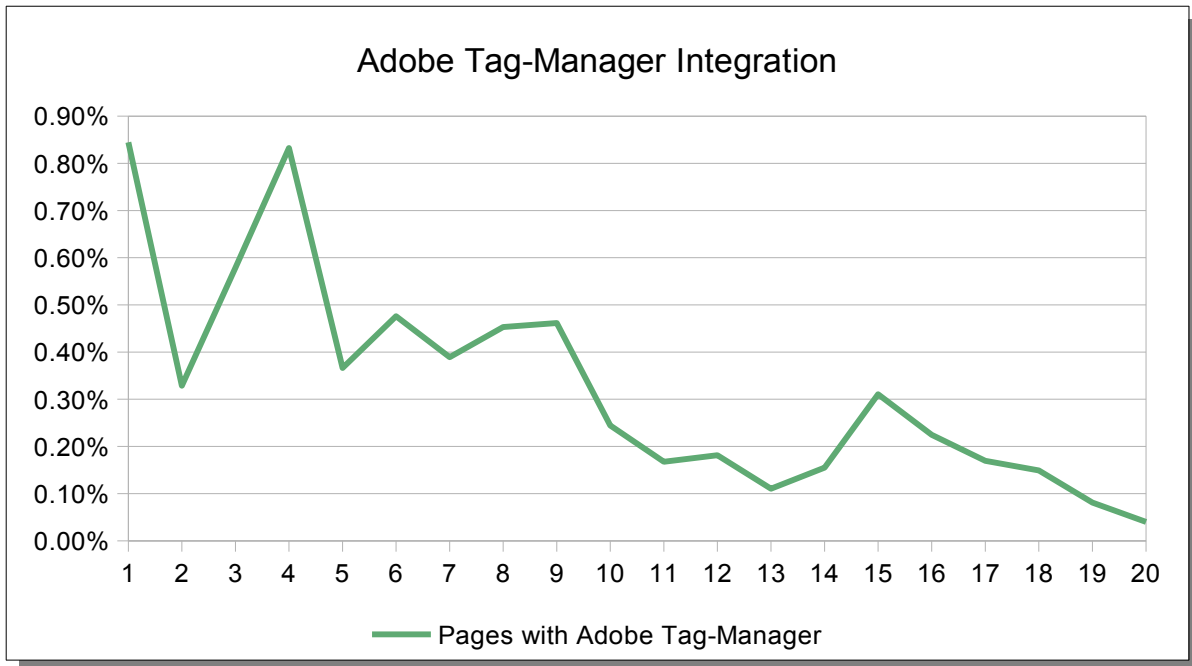
	Whole Keyword Set
correlation	0.04
pages with Google TM	2.5%

Despite the inaccessibility of Google Search, it is notable that services like Google Analytics and Tag Manager remain unblocked and could play a role in site optimization.

Despite a slight positive correlation, the reduced adoption rate of Google Tag Manager among top-ranking Baidu pages, now at 2.5% down from the previous 8%, suggests its significance as a ranking factor is likely minimal.

Adobe Tag-Manager Integration

Exploring the prevalence of Adobe Tag Manager, often preferred by established Western enterprises as an alternative to Google Tag Manager, may yield insights into its adoption relative to its Google counterpart within top-ranking Baidu pages.

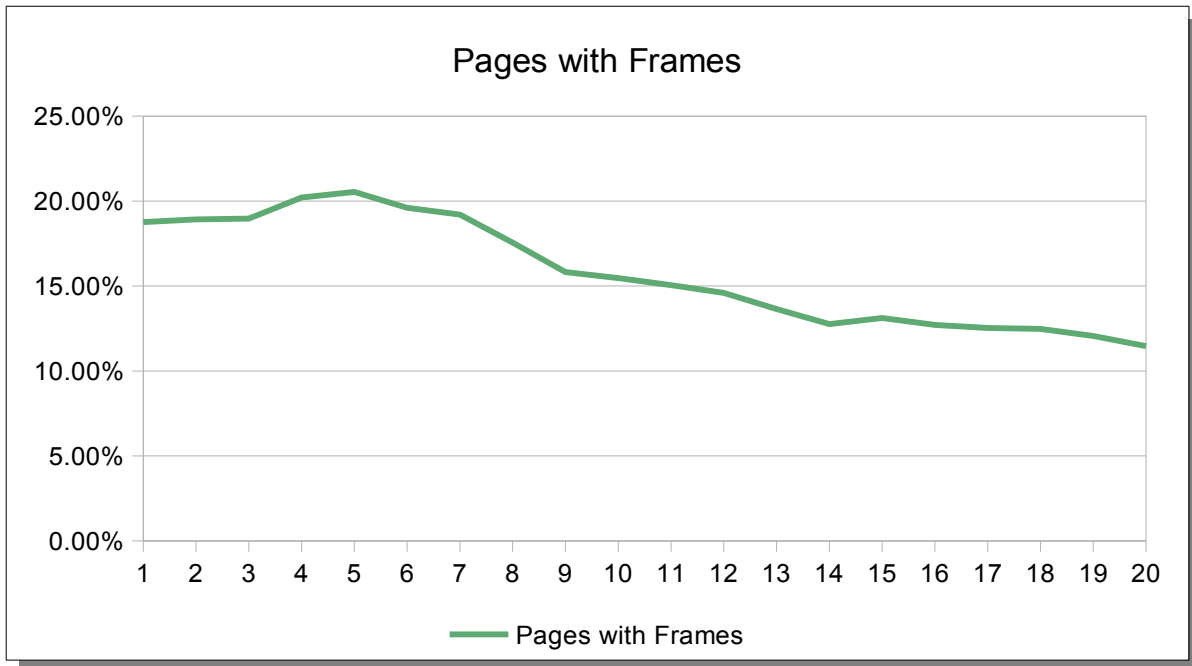


	Whole Keyword Set
correlation	0.03
pages with Adobe TM	0.3%

Despite its lower utilization compared to Google Tag Manager on top-ranking Baidu pages, Adobe Tag Manager exhibits a marginal positive correlation with superior search positions. However, this slight association is not indicative of a causal relationship for ranking improvements.

Frameset Integration

The use of frames within modern websites is often considered a deprecated practice, potentially impeding a site's ability to rank well on Baidu. We analyse the absence of frames and its correlation with better rankings.



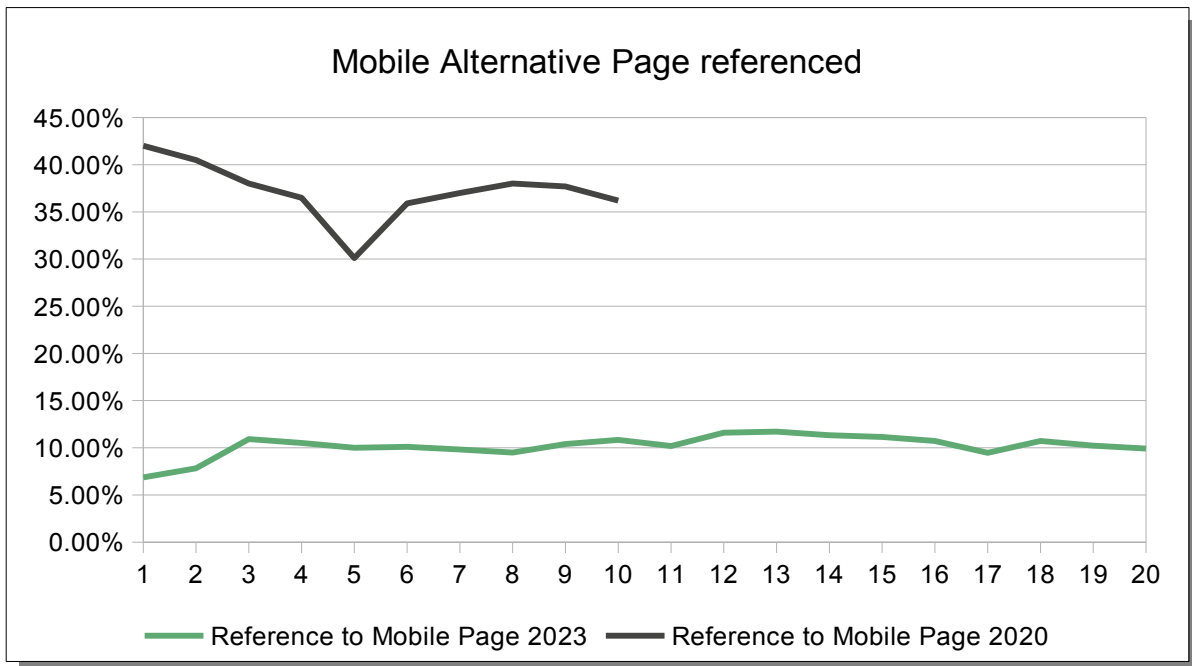
	Whole Keyword Set
correlation	-0.08
pages with frames	15.5%

Contrary to the initial hypothesis that the use of frames, an outdated web design technique, might negatively impact rankings, data reveals a higher prevalence of frames in top-ranking pages. With 20% of the top-ranking URLs featuring frames, compared to only 12% in lower-ranking pages, the findings suggest that frames may not detract from a page's ranking potential on Baidu.

The prevalence of frames within top-ranking pages may often stem from the inclusion of tag managers or analytics tools rather than serving as core design elements.

Mobile Page referenced

In China, it remains commonplace to encounter desktop and mobile website versions on separate domains or within distinct subdomains or folders. Despite this practice, Baidu's recommendations favor responsive web design as a more unified and contemporary approach. The presence of links to separate mobile versions serves as a barometer for gauging the progress of web modernization in China since 2020, reflecting the transition from traditional configurations to responsive design methodologies recommended by Baidu. Our correlation calculations assume that the existence of such a tag might rather be a negative signal.



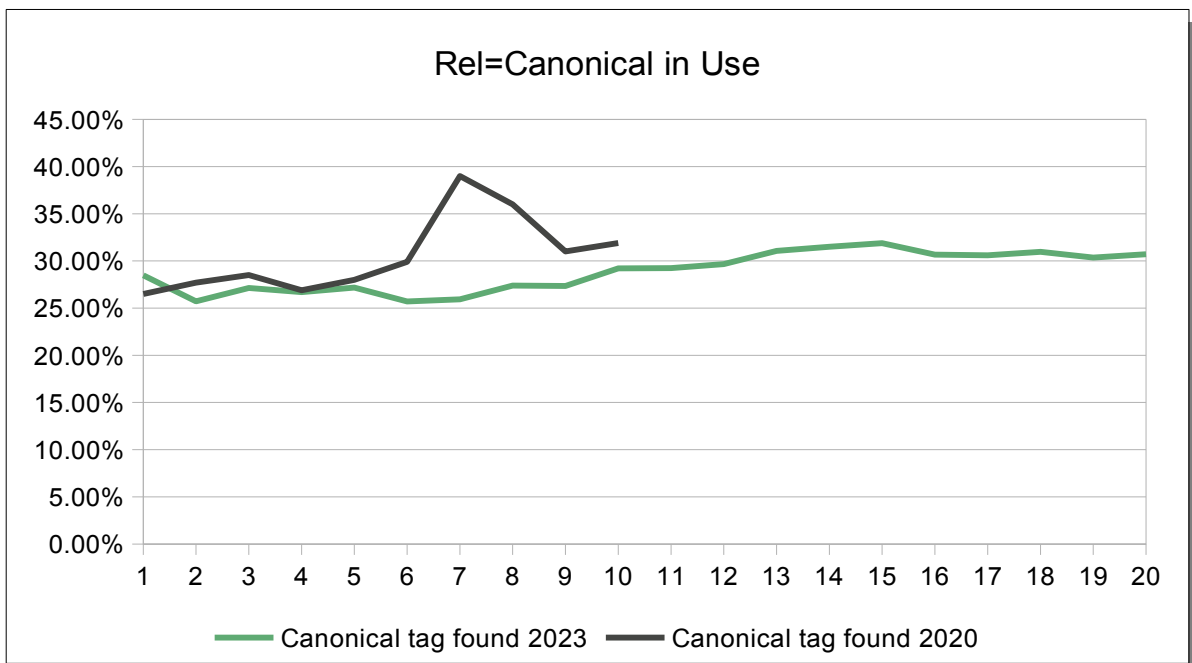
Whole Keyword Set	
correlation	0.01
pages with mob. alternative	10.3%

The data reveals a modest positive correlation between the lack of a reference to a dedicated mobile page and improved rankings on Baidu.

Coupled with a noticeable decline in such references since 2020, the trend indicates a gradual alignment with global standards, as Chinese web practices evolve away from unnecessary separate mobile pages towards more streamlined, responsive designs.

Canonical Tag Usage

Employing a canonical tag is considered best practice for SEO, even for pages intended for indexing. It provides search engines with clarity on preferred URLs, ensuring that search visibility is concentrated on the authoritative page and mitigating the risk of diluting ranking signals across duplicates.



Whole Keyword Set	
correlation	-0.04
canonical found	29%

Baidu recognizes the rel=canonical tag within single domains, yet its sparse utilization among Baidu's Top 20 pages, coupled with a slightly negative correlation score, implies a negligible effect on ranking.

Alternatively, this could indicate that top-ranking sites maintain such pristine code structures that the necessity for rel=canonical is significantly diminished—a conjecture we offer for your contemplation.

Expert comment:

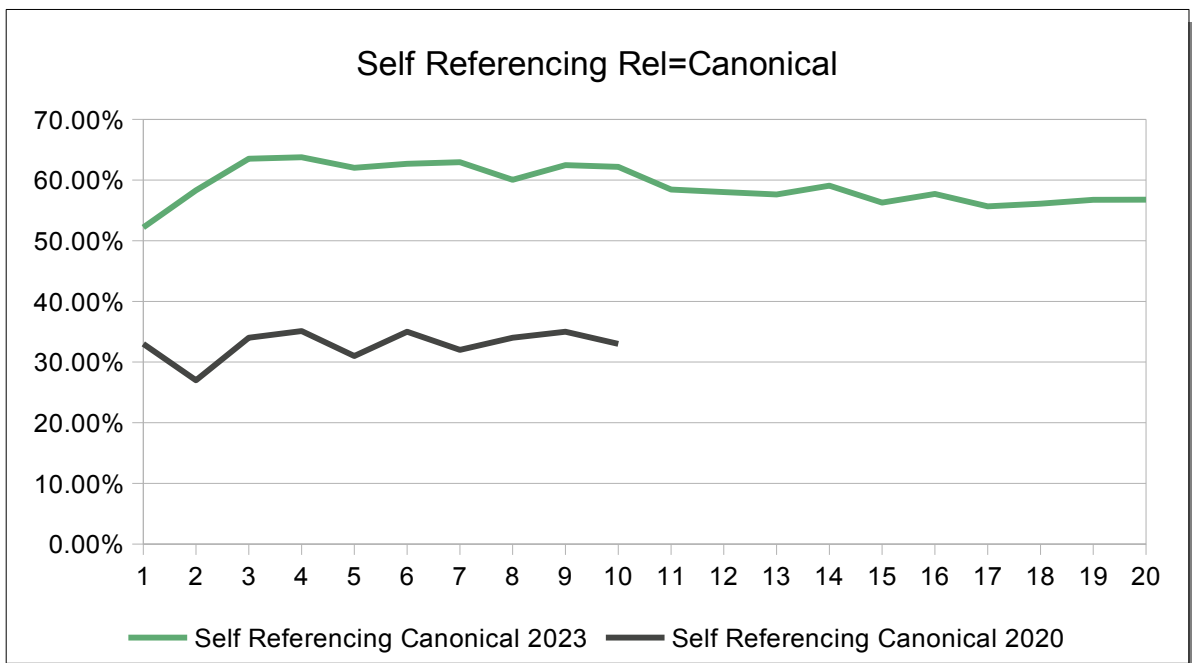


“Baidu only uses rel=canonical for mobile version directives. Confirmed with Baidu.”

Hermes Ma (马骏), Head of Performance Marketing,
Merkle, Dentsu, www.dentsu.com/cn

Self Referencing Rel=Canonical

The presence of self-referencing canonical tags on top-ranking pages may serve as a ranking signal, suggesting either Baidu's ambivalence towards canonical references or that these tags often point to irrelevant pages. This factor's potential influence on rankings warrants close examination.

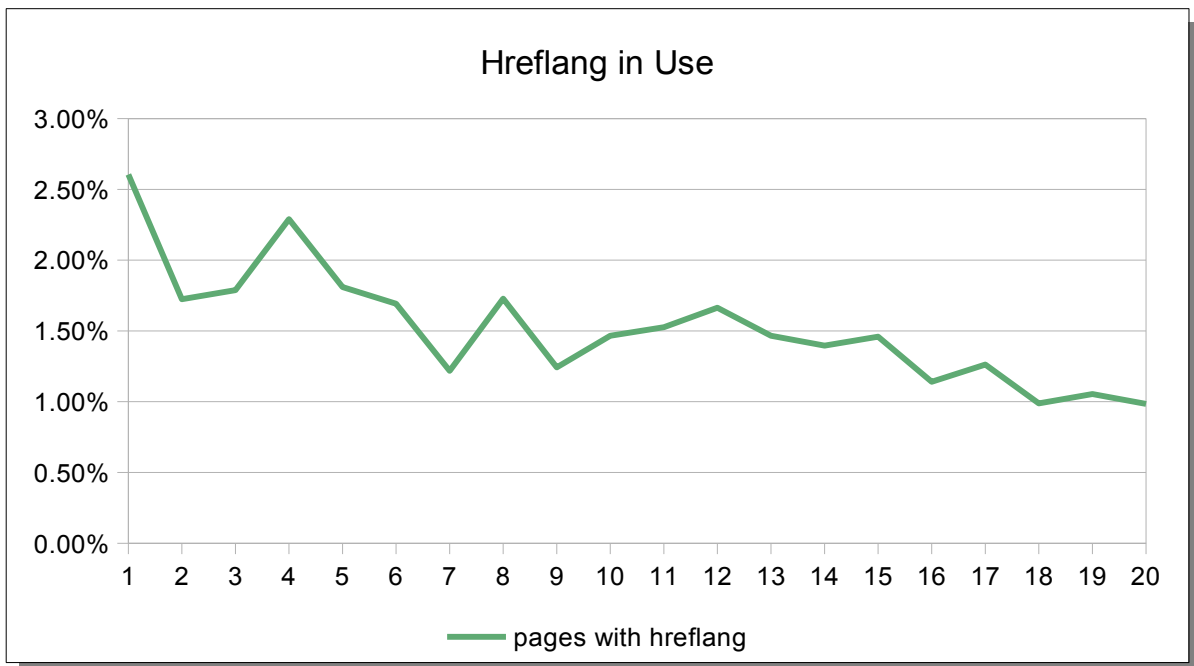


Whole Keyword Set	
correlation	0.04
canonical self referencing	59%

The shift from a negative to a positive correlation regarding self-referencing canonical tags suggests an evolving landscape. In 2020, only 33% of top-ranking pages employed self-referencing canonicals, whereas now, that figure has risen to 59%. This significant increase indicates that Baidu may now be placing greater emphasis on the presence of canonical tags as part of its ranking criteria.

Hreflang in Use

The hreflang attribute is typically overlooked in Baidu SEO, as the engine caters specifically to Chinese users. This study examines whether hreflang could negatively affect rankings on Baidu, despite its intended use for global audience targeting (positive correlation means, we see less hreflang, negative correlation means we see more hreflang).



	Whole Keyword Set
correlation	0.02
pages with hreflang	1.5%

While present in a modest 1.5% of Baidu's top 20 ranking pages, the hreflang tag is not considered a direct ranking booster for Baidu SEO. Its increased appearance in top-positioned sites, peaking at 2.5% for the number one spot, may hint at a growing presence of internationally-oriented websites on Baidu's search results.

Expert Comments:

“I’m not surprised to see this number being so low given Baidu’s lack of support for Hreflang. From experience, and brand or website entering the Chinese market by creating a subdomain or subfolder of an existing domain, I would look to restrict Baidu’s crawling access to your non-Chinese market intended content.”

Dan Taylor, Editor, ChineseSearchNews.com / Head of Technical SEO at SALT.agency



“We’ve known for a while that Baidu doesn’t pay much heed to Hreflang implementation, with most studies suggesting that the attribute isn’t supported. This is no surprise given that Baidu exists almost solely for Simplified Chinese speakers in Mainland China. That being said, I’m interested in keeping an eye on Hreflang use and its correlations with the performance of internationally-oriented sites with Simplified Chinese versions ranking on Baidu. This may signal to the search engine taking a more global approach in its aspirations.”

Owain Lloyd-Williams, Independent SEO Consultant and Owner at Son of Jack Ltd, www.owainlloydwilliams.com

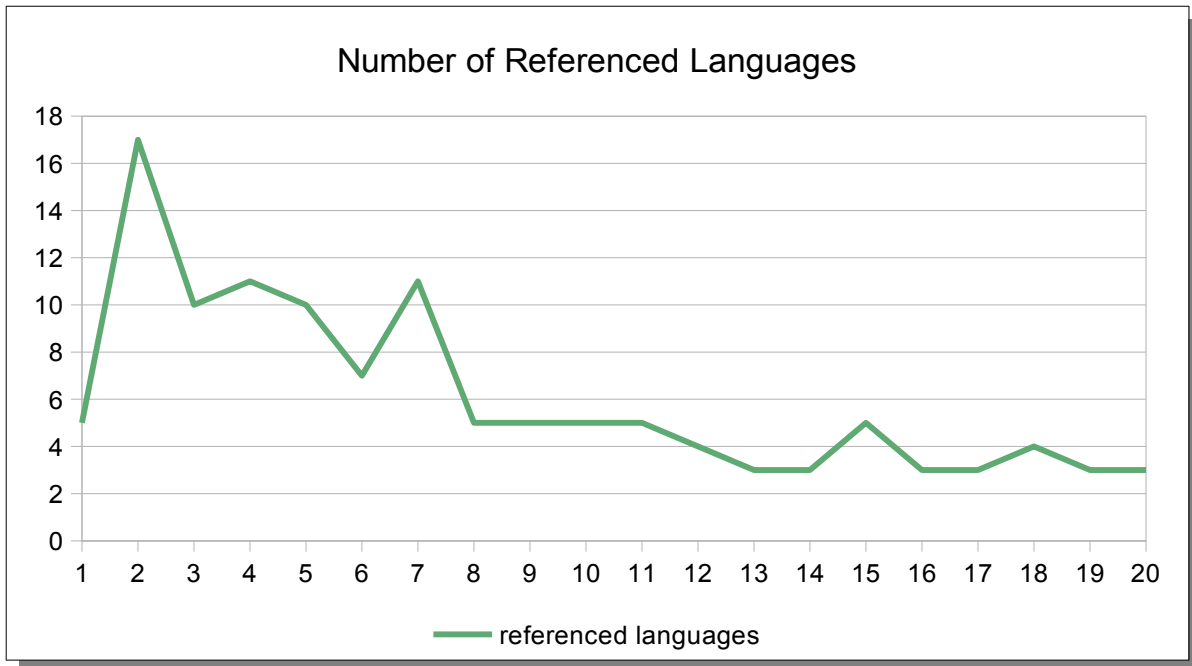


“This chart, perhaps even more so than all others shows how much Baidu is dominated by domestic Chinese-only focused sites. While not all foreign sites will use hreflang, as it's unlikely for a local Chinese site that only focuses on China to use hreflang, this 1.5% representing foreign sites is still stark.”

Simon Lesser, Cofounder and CEO, Dragon Metrics, www.dragonmetrics.com

Number of Hreflang References

The presence of multiple languages referenced through hreflang is seen in about 1.5% of Baidu's top-ranking pages. We will scrutinize whether a broader linguistic range correlates negatively with ranking efficacy.



	Whole Keyword Set
correlation	0.2
median number of referenced languages	5

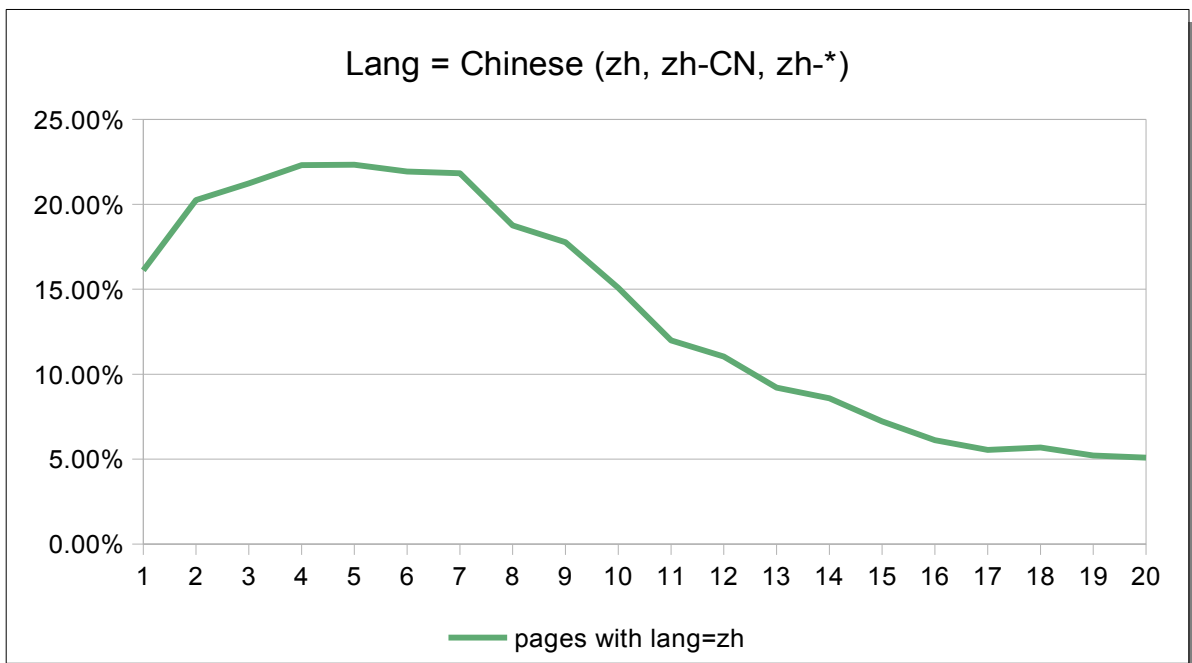
	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.2	0.2	0.2	0.2
number of ref. languages	5	3	5	6

While top-ranking pages on Baidu that employ hreflang attribute reference a greater number of languages on average, an interesting pattern emerges for

competitive short-head keywords, where leading pages tend to reference fewer languages. This nuanced observation hints at a complex interplay between language diversity and search visibility.

<html> Lang-Attribute = Chinese

Shifting focus to the lang attribute within HTML tags, we anticipate "zh," "zh-CN," or "zh-Hans" to predominate for Mainland China-targeted pages. The ensuing analysis will reveal whether this attribute aligns positively with enhanced Baidu search rankings.



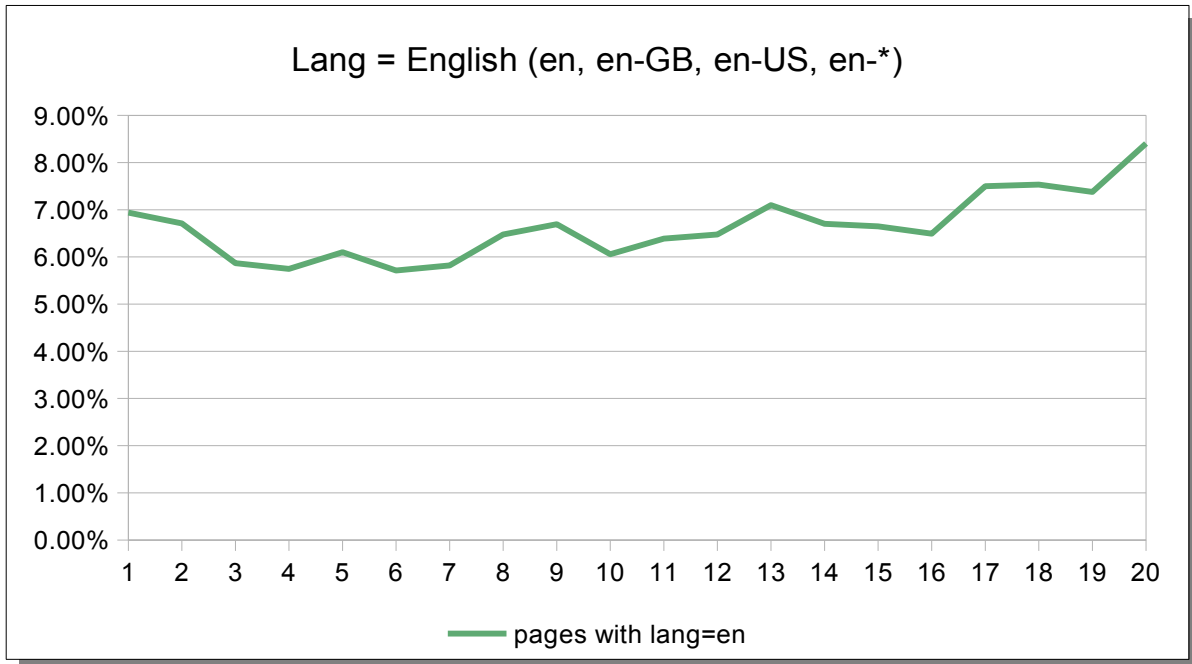
	Whole Keyword Set
correlation	0.2
html lang contains zh	13.3%

A notable positive correlation emerges between the specification of Chinese (zh) in the lang attribute and improved Baidu rankings.

However, with only 13.3% of pages utilizing "zh," further investigation into the language settings of the remaining pages is warranted.

<html> Lang-Attribute = English

The prevalence of "en" as the default language setting in templates, coupled with a mere 13.3% of pages declaring "zh," prompts an examination for a potential negative correlation between English language specification and ranking performance on Baidu.



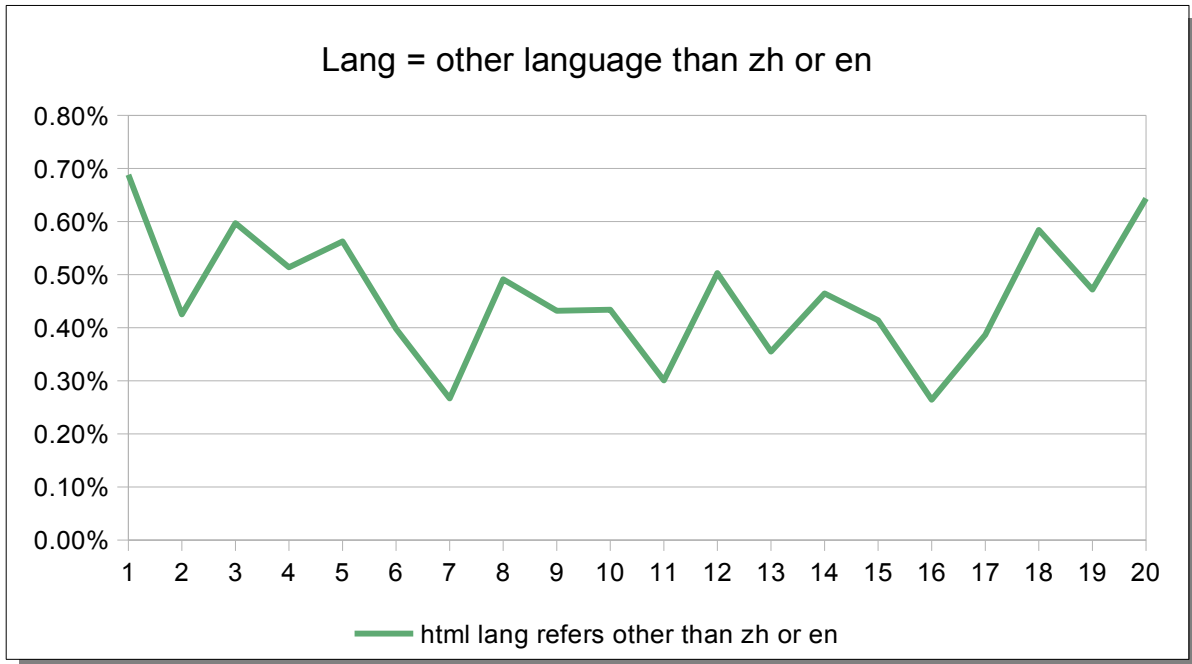
	Whole Keyword Set
correlation	0.02
html lang contains en	6.6%

A subtle positive correlation emerges when English is omitted as the HTML language attribute, suggesting a potential ranking advantage on Baidu.

However, English language tags constitute a scant 6.6%, failing to bridge the disparity left by the low 13.3% utilization of Chinese language references.

<html> Lang-Attribute = any other language

Given the minimal use of Chinese and even fewer instances of English as content language tags, one might speculate about the prevalence of other languages. Do these alternative language references correlate positively with poorer rankings on Baidu?



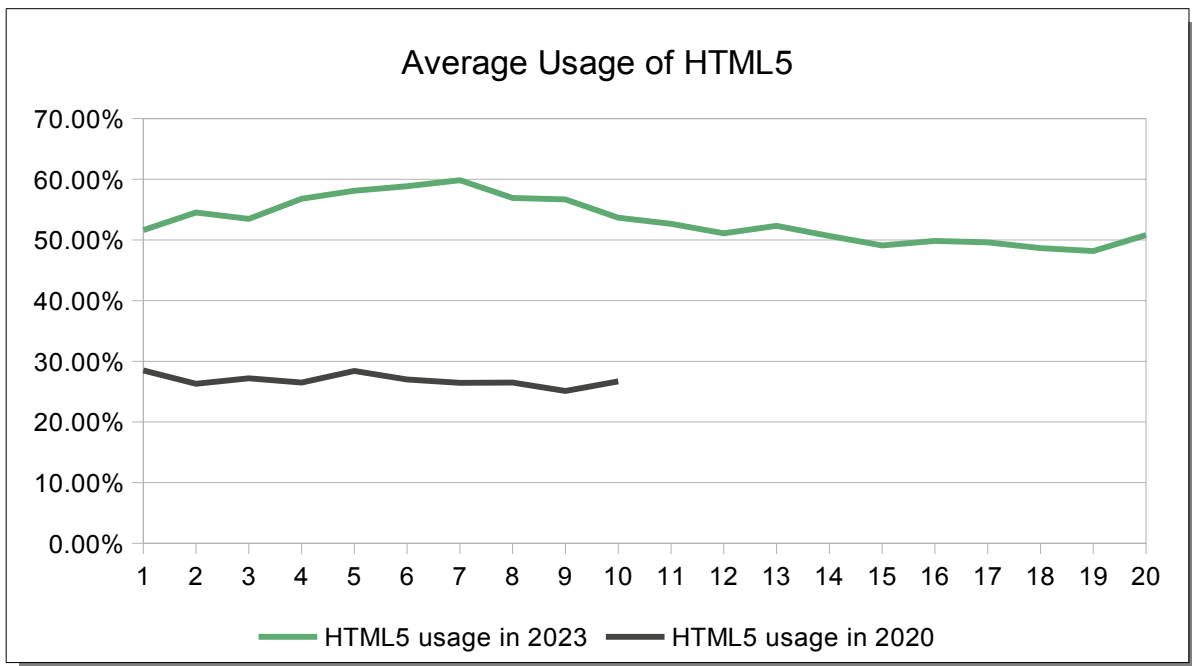
	Whole Keyword Set
correlation	0.001
html lang not zh or en	0.4%

The rarity of alternative language tags, at a mere 0.4%, coupled with the modest figures for Chinese (13.3%) and English (6.6%), suggests that the norm for webmasters is to omit the content language attribute.

Given the distinctiveness of Simplified Chinese, search engines may readily bypass this need for explicit language identification.

HTML5 Adaption

In 2020, HTML5, the latest standard in web development, is employed by approximately 26% of top-ranking pages. As the internet evolves, this uptake is expected to increase. Given the semantic clarity provided by HTML5 tags, which potentially enhances search engine comprehension, we might anticipate a correlation between HTML5 implementation and improved rankings on Baidu.



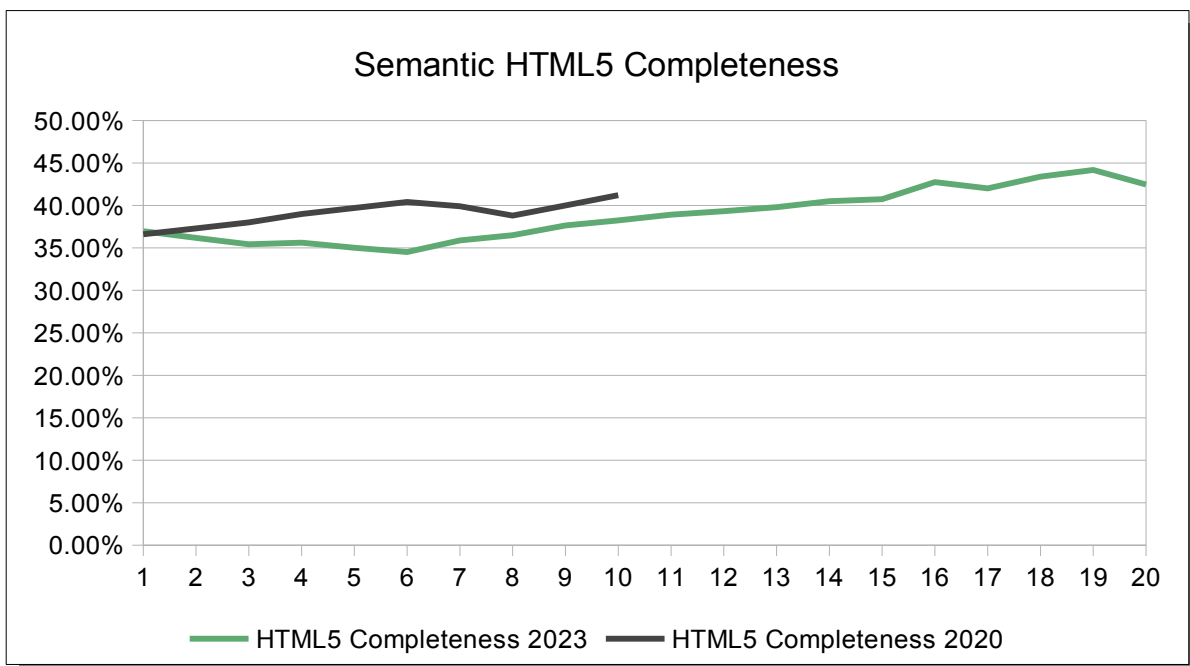
	Whole Keyword Set
correlation	0.06
adaption of html5	53.2%

Over the past three years, adoption of HTML5 has surged by approximately 27%. A modest positive correlation has been noted between the application of HTML5 standards and enhanced ranking positions on Baidu.

Semantic HTML5 Completeness

Considering the observed minor positive correlation between HTML5 implementation and improved Baidu rankings, we turn our focus to the extent of semantic HTML5 tag usage—such as 'header', 'main', 'footer', 'aside', and 'nav'—and its association with ranking performance.

This analysis will discern whether a more comprehensive application of these structure-defining tags aligns with higher ranking positions.

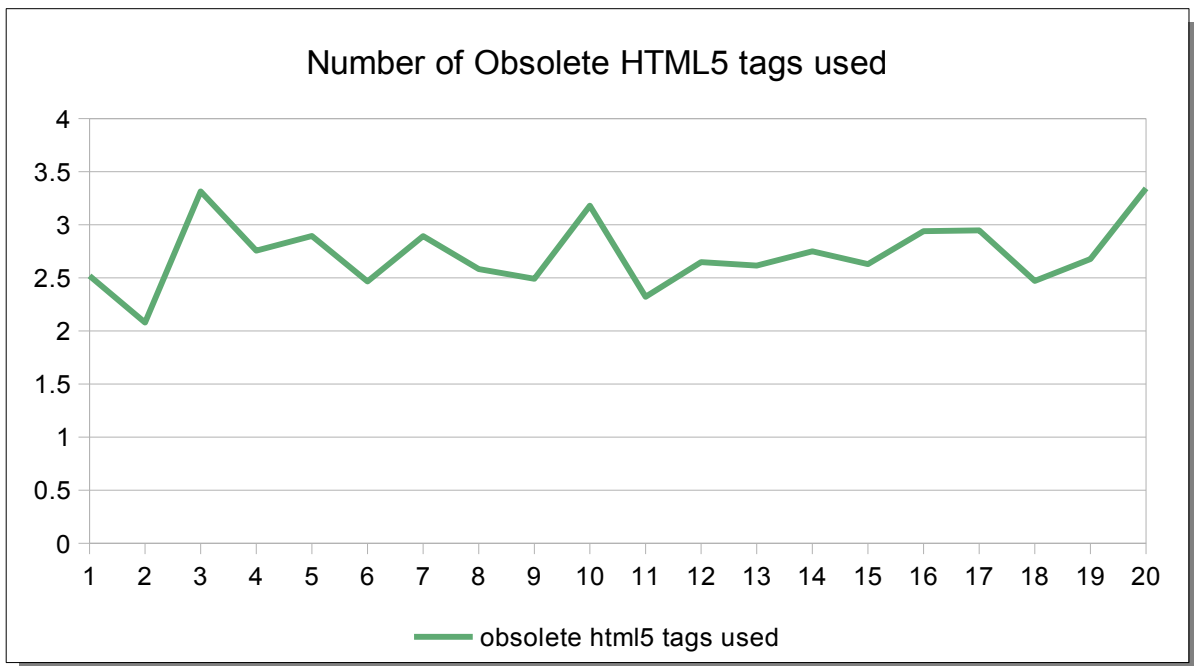


	Whole Keyword Set
correlation	-0.08
semantical html5 adoption	38.8%

The variance in the use of distinct semantic HTML5 tags among Baidu's top-ranking websites is marginal, with the top 10 positions demonstrating a slight decrease. This pattern suggests that Baidu may not significantly weigh HTML5 structure in its ranking algorithm.

Number of Obsolete HTML5 tags used

Utilization of outdated HTML5 tags might negatively impact website performance on modern browsers and could potentially hinder search engine optimization efforts. Given the evolution of web standards, the presence of such deprecated tags could be interpreted by search engines as a sign of an unmaintained or outdated website, which may adversely affect rankings.

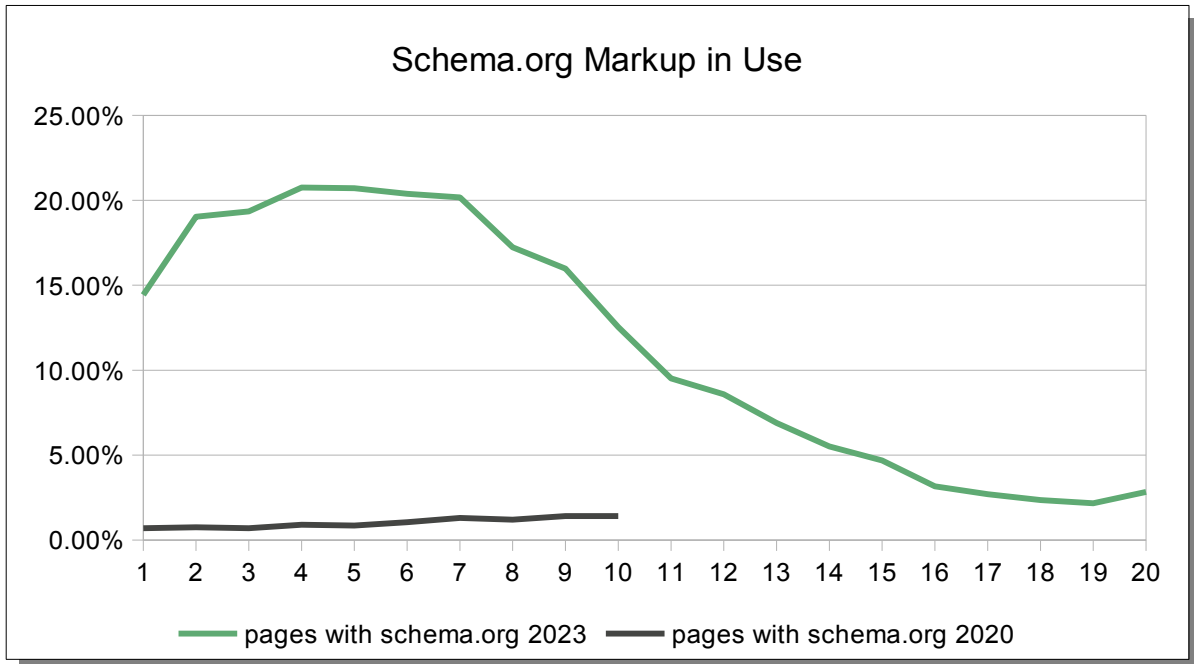


	Whole Keyword Set
correlation	0.02
obsolete html5 tags used	2.7

In fact, a slight positive correlation is observed between the avoidance of obsolete HTML5 tags and improved rankings, suggesting an increasing awareness among web designers of the importance of utilizing current web technologies for optimization.

Schema.org

The incorporation of structured data, particularly Schema.org, remains unsupported by Baidu according to official statements. Nonetheless, this analysis seeks to ascertain whether webmasters persist in embedding such markup, possibly as a consequence of its inclusion in contemporary templates.



	Whole Keyword Set
correlation	0.2
pages with schema.org	11%

Despite Baidu's official stance of not endorsing Schema.org, a significant uptick in its application is evident from 2020 to 2023. Notably, pages that achieve superior rankings on Baidu display a markedly higher implementation rate—exceeding 15% within the top 10—contrasting with a sub-10% adoption among pages ranked 11 to 20.

Expert comments:

“As a marketer that works in several international search engines, implementing schema becomes almost second-nature on pages like the homepage, about us page, product pages, etc. This can yield some pretty cool results on Google or Bing SEPRs. But with Baidu still not supporting schema, I still urge sites to forgo implementing schema for their sites designed to rank on Baidu. By avoiding schema, you not only save yourself time for other optimizations, but also cut down on unnecessary code.”

Adam Di Frisco, Senior Digital Marketing Strategist, Atigro, www.atigro.com



“It would be interesting to see who are those top 10 sites with schema.org to understand if this reflects the causation or just ranked giant sites just implemented this since 2020.”

Hermes Ma (马骏), Head of Performance Marketing, Merkle, Dentsu, www.dentsu.com/cn



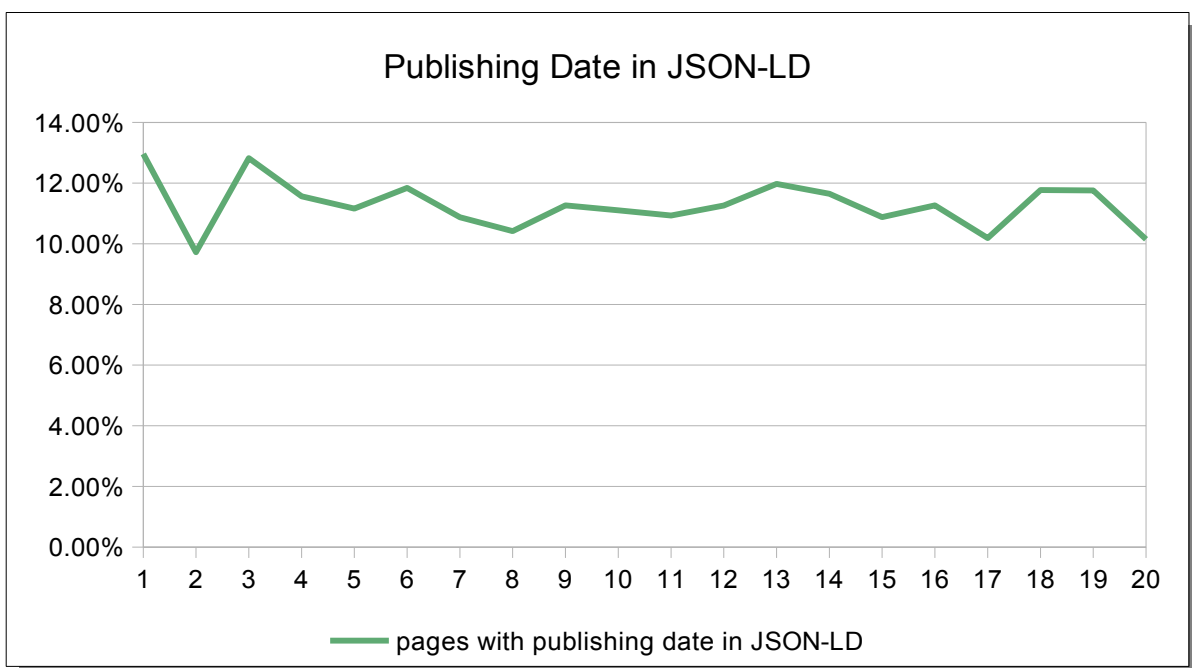
“I would encourage businesses to not shy from Schema implementation in their China SEO ventures even if it is officially unsupported by Baidu. This study shows that there has been an uptick in its usage from 2020 to 2023, with a clear correlation pages that achieve superior rankings. Given Baidu’s tendency for favouring such a visual SERP, you may be getting ahead of the game by adoption of Schema markup on your content.”

Owain Lloyd-Williams, Independent SEO Consultant and Owner at Son of Jack Ltd, www.owainlloydwilliams.com

Publishing Date in JSON-LD

In the pursuit of optimizing for Baidu's search algorithms, the inclusion of a publishing date within the JSON-LD structured data format emerges as an official ranking factor

(<https://www.jademonD.com/magazine/baidu-algorithm-updates/#aurora>). This particular element serves as a signal to Baidu, enhancing content freshness and relevancy, which are critical parameters in its ranking calculus.

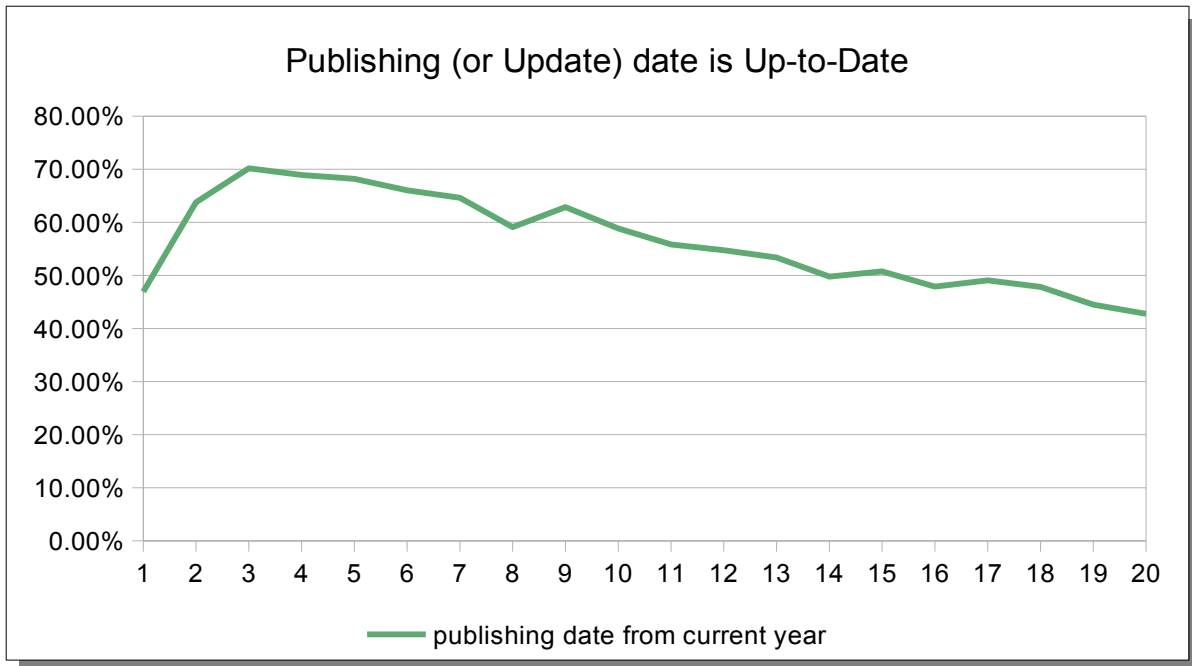


	Whole Keyword Set
correlation	0.01
pages with publishing date in JSON-LD	11%

A marginal correlation exists between higher rankings and the implementation of publishing dates in JSON-LD format across Baidu's top 20 ranking sites. With adoption rates hovering between 10 to 12%, the consistent implementation across the spectrum suggests an untapped opportunity for optimization in Baidu SEO.

Publishing Date (or Update date) in JSON-LD is Up-to-Date

Analyzing Baidu's algorithm, we're probing whether a recent publishing date correlates with improved rankings, hypothesizing that Baidu may prioritize freshly updated information to deliver the most current content to its users.



	Whole Keyword Set
correlation	0.14
publishing date is up-to-date	56%

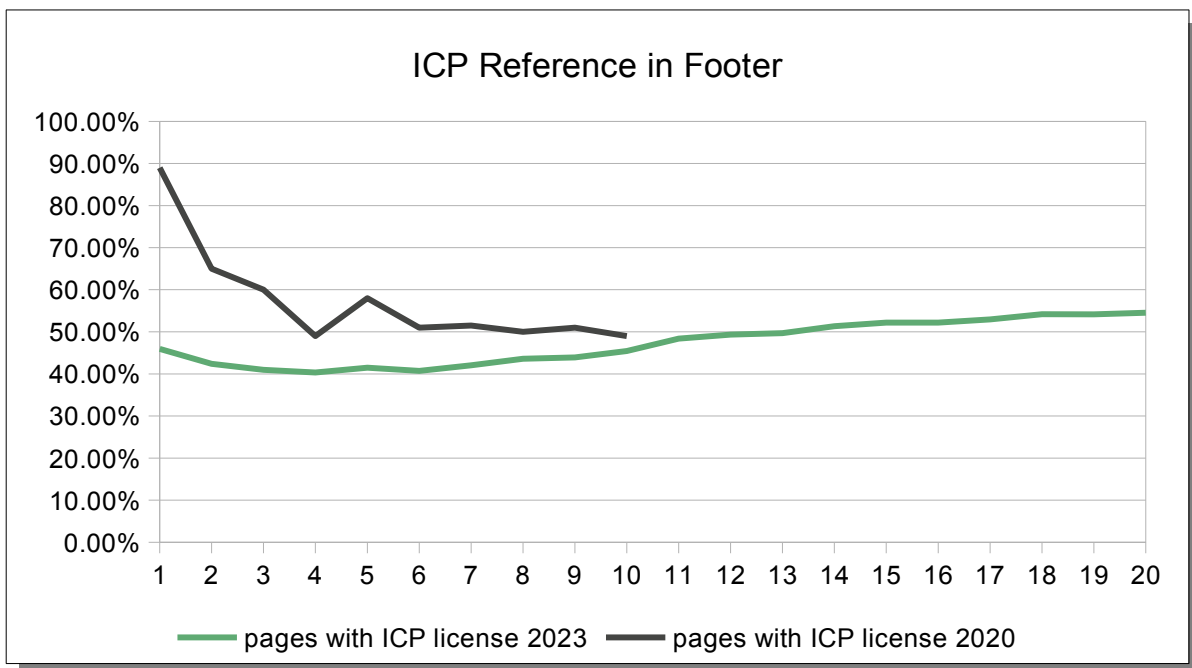
A discernible positive correlation emerges between higher-ranked pages and publication dates within the current year, suggesting Baidu's algorithm does give credence to its stated preference for recently updated content.

Website Legal Requirements

In our analysis of legal adherence, we recognize that while there are no explicit legal mandates for ranking on Baidu, compliance with local Chinese legislation remains a critical underpinning for visibility within the search engine's results.

ICP License

In evaluating Baidu's ranking landscape, attention is drawn to the ICP license—a mandatory(?) credential for China-based websites. Its presence or absence may subtly influence a site's standing within Baidu's search results, warranting a closer examination as a potential ranking factor.



	Whole	Keyword Set
correlation		-0.09
pages with ICP license		48%

The prevalence of ICP licenses among top-ranking Baidu pages, at 48%, has dipped slightly from 51.9% in 2020. While securing an ICP license is a trusted best practice that may enhance credibility, it is not a prerequisite for high rankings on Baidu, countering the overstated claims of some SEO narratives.

Expert Comment:



“Whilst having an ICP license isn’t a prerequisite to rank in Baidu - if you’re a Western brand expanding into the Chinese market and wanting to rank in Baidu, I have witnessed instances where a random website targeted the brand of a well-known Western DevOps brand in Baidu, and just linked to random software (affiliate links). But when the Western brand acquired the ICP, and published evidence, within days Baidu ranked their Chinese .com.cn domain for the brand ahead of the random website.”

Dan Taylor, Editor, ChineseSearchNews.com / Head of Technical SEO at SALT.agency



“It is refreshing to hear that the ICP License doesn’t present a clear correlation between improved rankings and visibility on Baidu, despite the narrative that it is a mandatory requirement for China-based sites. While it’s still recommended to obtain one, getting an ICP License is often a bureaucratic headache for many businesses, particularly for those without physical entities in China.”

Owain Lloyd-Williams, Independent SEO Consultant and Owner at Son of Jack Ltd, www.owainlloydwilliams.com

“In September 2023, Baidu’s webmaster tools announced a cleanup to remove or downgrade many low-quality, unqualified websites (see: <https://www.sinoanalytics.com/seo/baidu-tightens-control-on-risky-resources-in-search-platform-735.html>). Possessing an ICP license is essential for accessing more advanced features in Baidu’s webmaster tools, such as sitemap submission, site logo customization, and more.”*

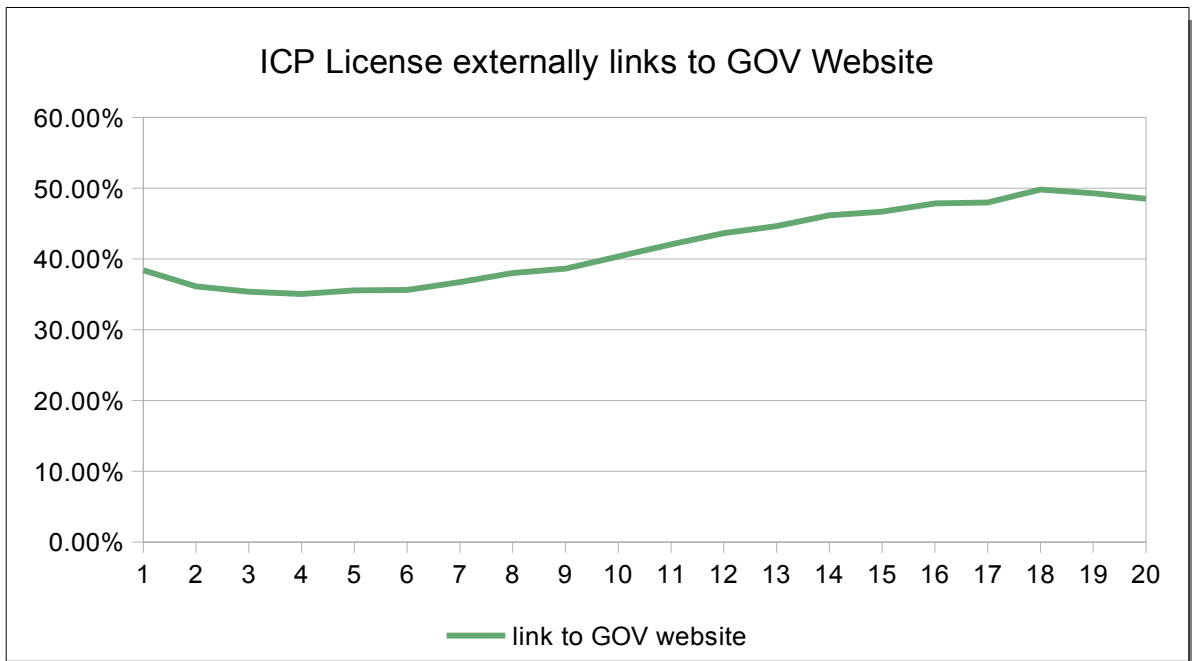


Qing He, Head of Search, Jademond Digital, www.jademond.com

**editors note: “Baidu Webmaster Tools” is the term we refer to in the West, while in the original it is Baidu Ziyuan (ziyuan.baidu.com), which refers to the tool comparable to Baidu Search Console, but also to the information platform, where Baidu announces information on updates or initiatives.*

ICP License externally links to GOV Website

The presence of a link from the ICP license notation to the official government verification site in the footer is a marker of authenticity and transparency. While this practice establishes trust and confirms the validity of the ICP license, its direct influence as a ranking factor on Baidu remains to be seen.



	Whole Keyword Set
correlation	-0.1
Link to GOV website	42%

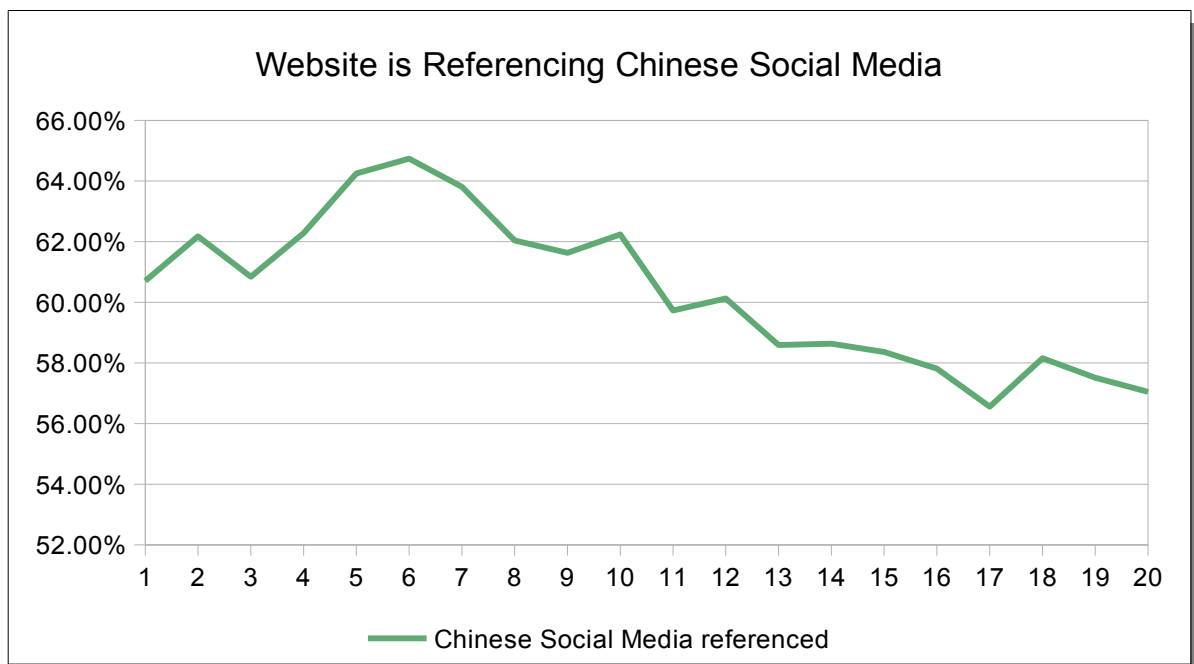
Less than half of the top-ranking websites on Baidu display an ICP license, and of these, only 42% provide a verification link to the government's official site, underscoring the varied adherence to this aspect of website authentication. It is still good practice to do so.

Social Media Integration

In assessing the integration of social media on top-ranking Baidu websites, this segment of the study examines the prevalence and potential influence of outbound social media links and widgets as factors in search visibility on China's leading search platform.

Referencing Chinese Social Media

This analysis probes the potential impact of referencing or incorporating Chinese social media platforms on website rankings within Baidu's search results, exploring the hypothesis that social connectivity may contribute to enhanced visibility and relevance in the digital space.

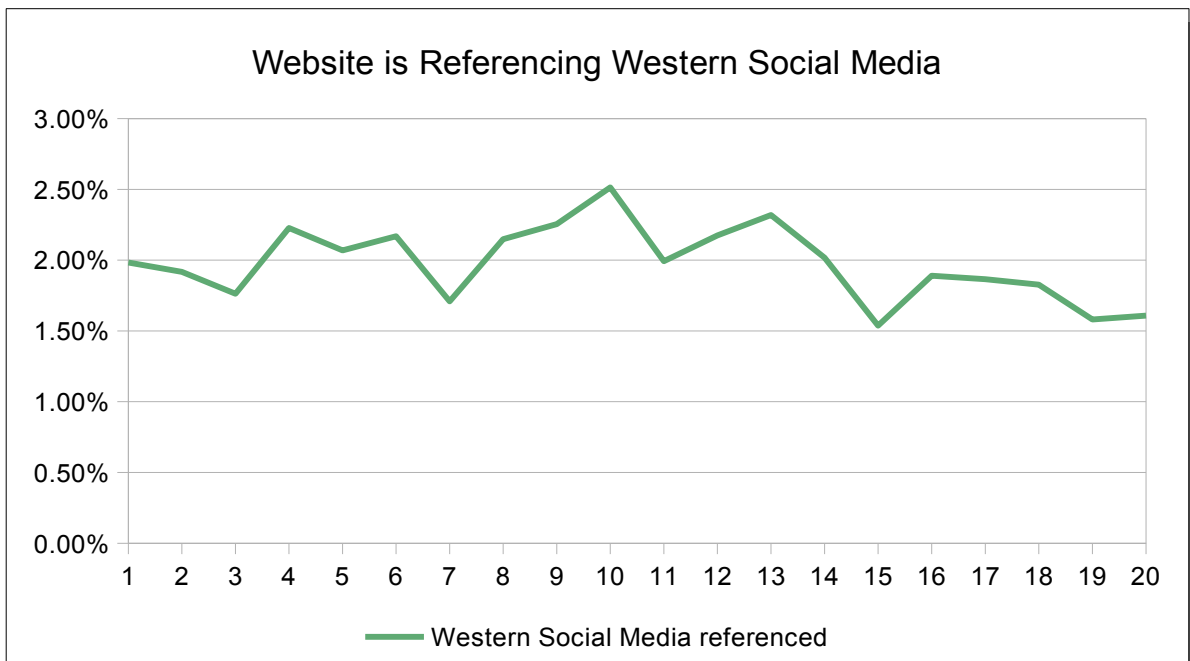


	Whole Keyword Set
correlation	0.04
pages with reference to chinese social media	60%

A correlation analysis indicates that top-ranking pages on Baidu often reference Chinese social media platforms, with over 60% of the top 20 pages integrating such connections, suggesting a potential link to improved search visibility.

Referencing Chinese Social Media

The integration of Western social media platforms on websites may potentially be viewed negatively in terms of Baidu search rankings, given the inaccessibility of these platforms in China, which could detract from the user experience for the local audience. When looking at this factor, we correlate the absence of such integrations against better rankings.



	Whole Keyword Set
correlation	-0.01
pages with reference to western social media	2%

A mere 2% of Baidu's top-ranking pages reference Western social media platforms like Facebook, Twitter, or YouTube. Although we see a slight negative correlation observed between top rankings and the absence of these links, it's reasonable to infer that Baidu does not regard the integration of such elements favorably. Best practices recommend tailoring content to Chinese users,

suggesting that references to inaccessible Western social media may be an oversight by international sites failing to fully localize for the Chinese market.

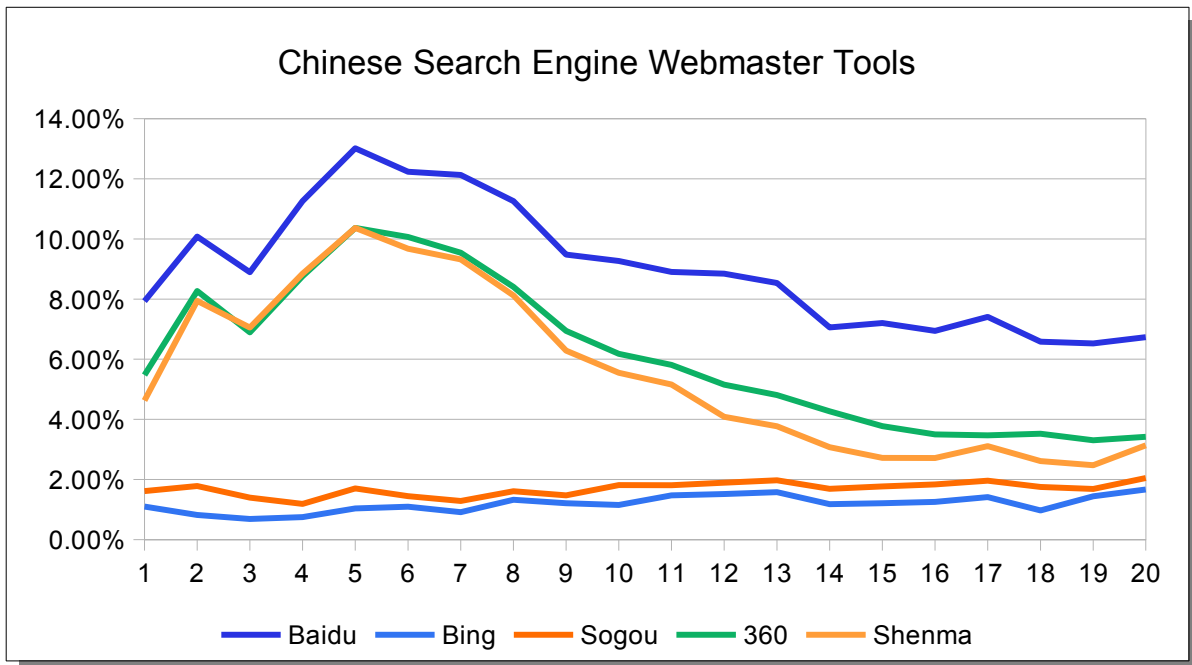
Webmaster Tools

In the forthcoming section of our study we scrutinize the adoption of various search engines' or SEO tools' webmaster tools by high-ranking websites, shedding light on the search engine optimization strategies and preferences of webmasters.

Note that our investigation is confined to direct HTML meta tag verification, excluding any verification implemented via JavaScript or tag managers.

Chinese Search Engine Webmaster Tools

The accompanying chart provides a succinct overview of the percentage of top-ranking Baidu pages that have verified ownership with webmaster tools of search engines prevalent in China, offering insights into the SEO practices and focus areas of these high-performing websites.

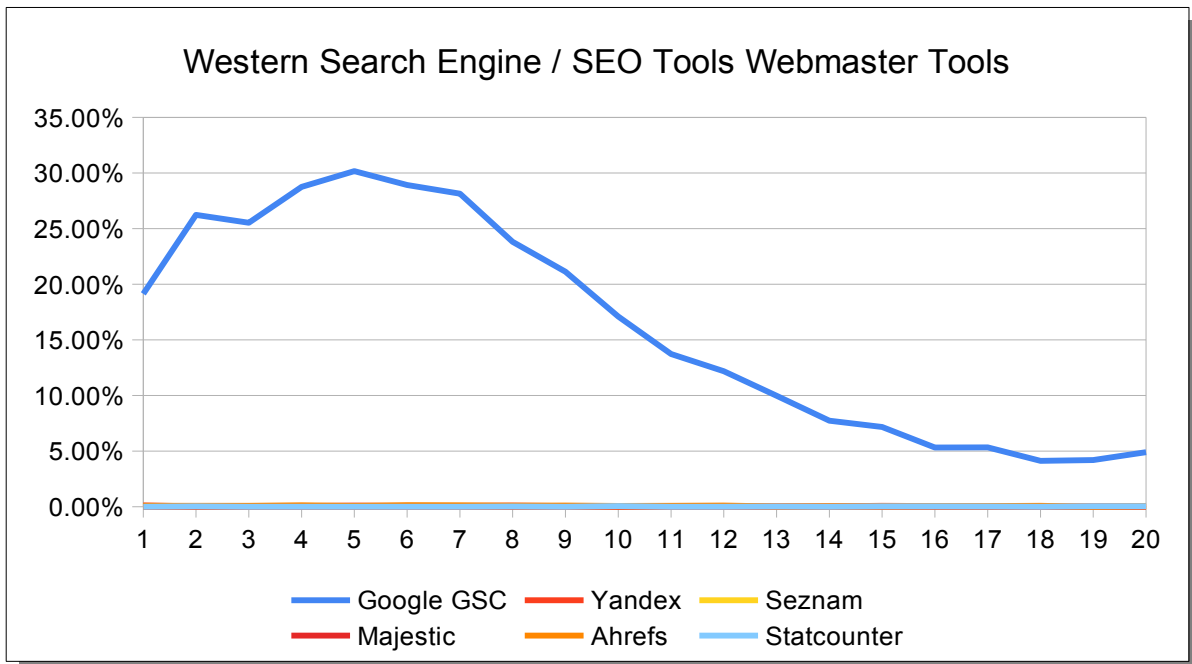


	Baidu	Bing	Sogou	360	Shenma
correlation	0.06	-0.02	-0.01	0.09	0.01
pages implementation	9%	1.2%	1.7%	6%	5.4%

Interestingly, besides Baidu Webmaster Tools, we still see many websites verifying 360 (so.com) and Shenma (sm.cn) webmaster tools, although these search engines are not on the top of Chinese user's preference anymore.

Western Search Engines' and SEO Tool Webmaster Tools

The chart delineates the proportion of Baidu's top-ranking pages that validate their site with webmaster tools from Western search engines like Google or Yandex, as well as Western SEO or analytics tools such as Ahrefs or Statcounter, highlighting the cross-regional SEO strategies employed by these leading sites.



	GSC	Yandex	Seznam	Majestic	Ahrefs	Statcounter
correlation	0.24	0.006	/	/	0.008	-0.007
percentage	15.6%	0.05%	0%	0%	0.07%	0.01%

The 15.6% adoption of Google Search Console among Baidu's top-ranking sites, exceeding Baidu's own webmaster tools usage, indicates a preference for Google's advanced SEO analytics, which can inform cross-search engine optimization.

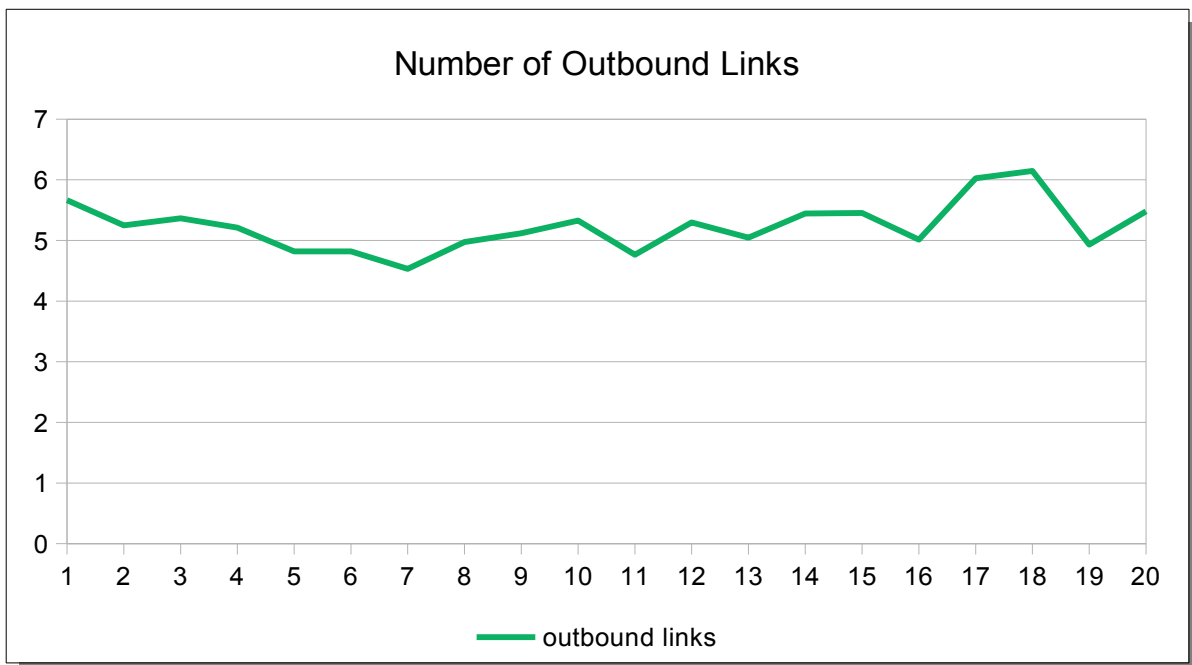
The minimal use of Statcounter suggests its market share data for China may be unreliable (see: <https://www.jadmond.com/magazine/search-engine-marketshare-china/>).

Links

In this segment of our study, we examine internal and external linking practices among Baidu's top-ranking pages. We assess the prevalence of jumpmark links for internal navigation, the proportion of links to non-Chinese content, and the balance between inbound and outbound links, to determine their influence on rankings.

Outbound links count

Exploring the web's interconnected nature, we turn to the potential ranking influence of linking to external sites. As digital ecosystems thrive on interlinkage, we investigate whether Baidu rewards pages that contribute to this network by linking to relevant external content.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.07	-0.07	-0.07	-0.07
outbound links	5.2	4.9	5	5.6

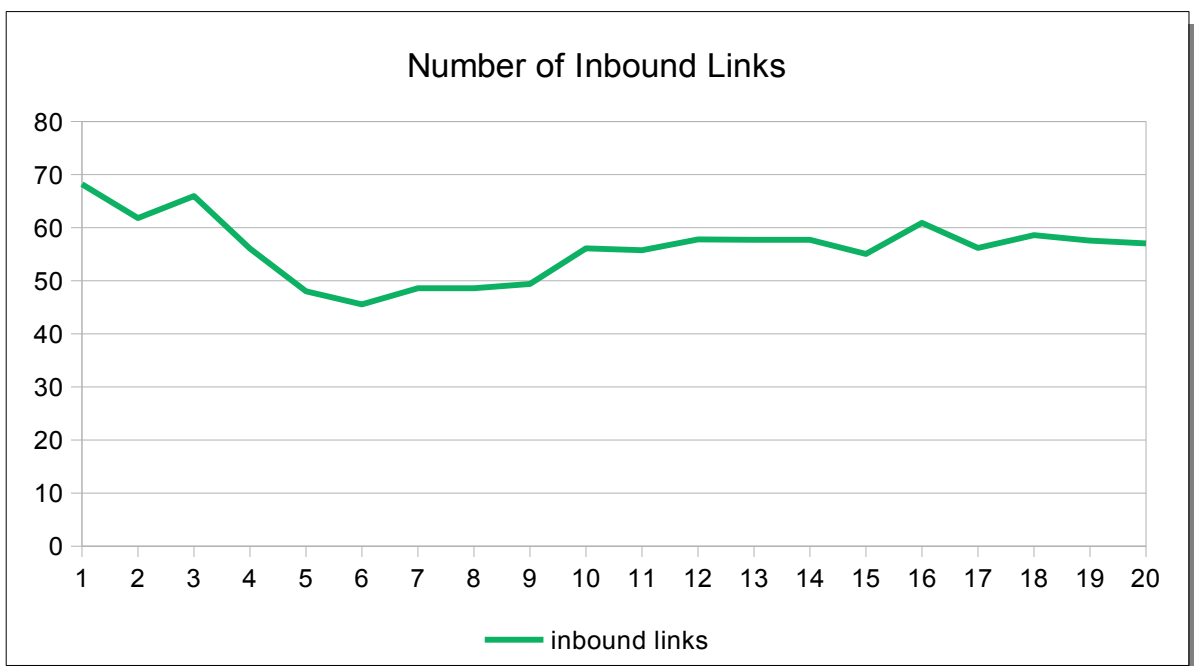
A minor negative correlation is evident between outbound links and higher rankings on Baidu.

Closer scrutiny reveals that top-ranked pages for more competitive, short-head keywords tend to feature fewer outbound links, suggesting restraint may be beneficial, with around five external links emerging as a common benchmark.

Internal Inbound links count

Exploring the impact of internal linking, we assess whether a robust network of inbound links within a website correlates with superior Baidu rankings. This potential ranking factor merits examination for its role in site navigation and hierarchy.

We look at the number of links pointing to further pages of the same website (not at the number of links pointing to the ranking page).



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.08	-0.08	-0.09	-0.08
inbound links	55.5	62	55.8	50.6

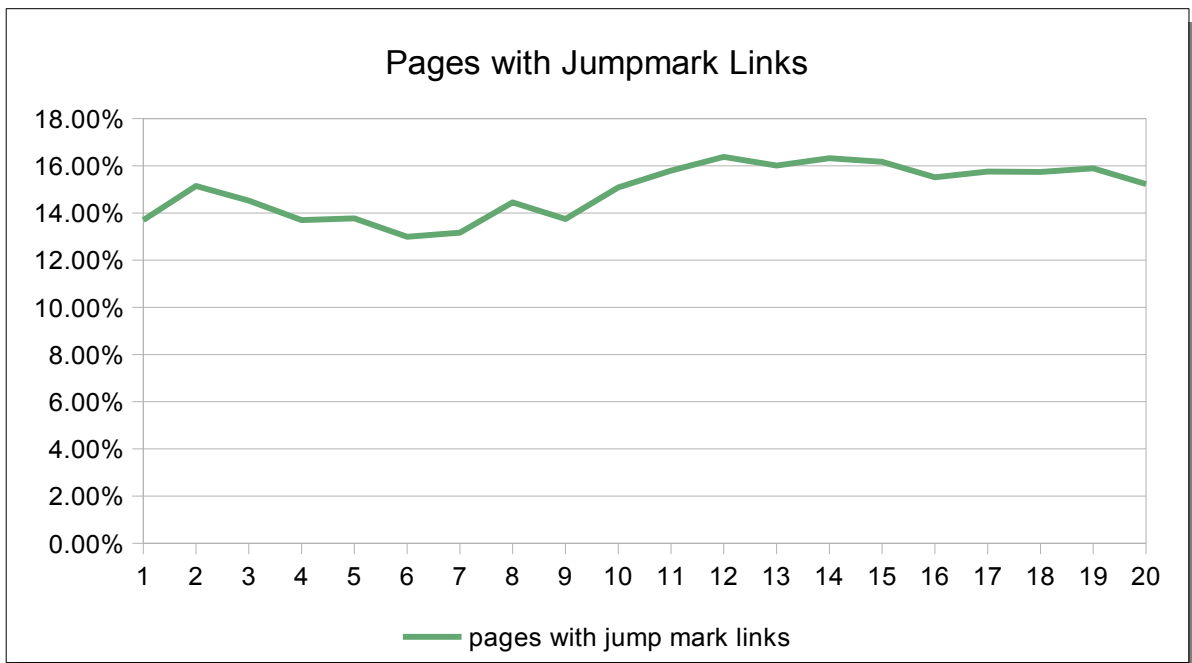
A marginal negative correlation exists between the quantity of outbound internal links from top-ranking pages on Baidu and their rank position. However, pages

that excel in competitive keyword arenas (short-head) demonstrate a higher propensity for internal linking compared to their midtail and longtail counterparts.

This pattern could reflect the inherent nature of short-head keywords necessitating more expansive content, thereby requiring additional internal pages to elaborate on the subject, or it might suggest that Baidu favors websites with a comprehensive internal linking structure.

URLs with Jumpmark Links

Investigating the influence of internal navigation aids, we turn to jumpmark links as a possible ranking signal. Such links typically serve as a table of contents, facilitating user access to specific sections of long-form content. We will examine whether this user-centric design feature correlates with enhanced visibility in Baidu's search results.

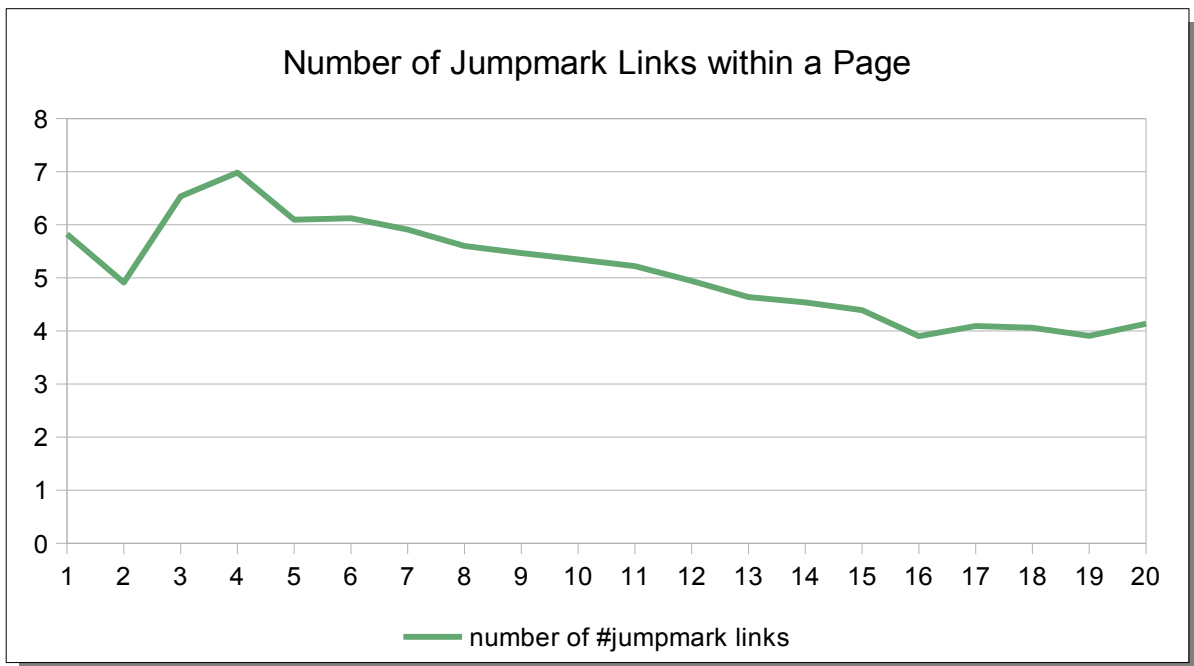


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.03	-0.02	-0.01
pages with jumpmark links	15%	13%	15%	16%

Contrary to expectations, the presence of jumpmark links within top-ranking pages shows a slight negative correlation with their search positions on Baidu indicating, that top ranking pages on Baidu might make less use of internal TOC links.

Number of Internal Jumpmark Links

Despite the previous negative correlation, we (JademonD) still consider the usage of jumpmark links a good practice to help users navigate within a page. Reason enough to investigate the potential influence of jumpmark links on search rankings, we assess whether a higher count of these navigational aids correlates with improved visibility on Baidu.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.12	0.1	0.2	0.01
number of jumpmark links	5	4.6	4.9	5.4

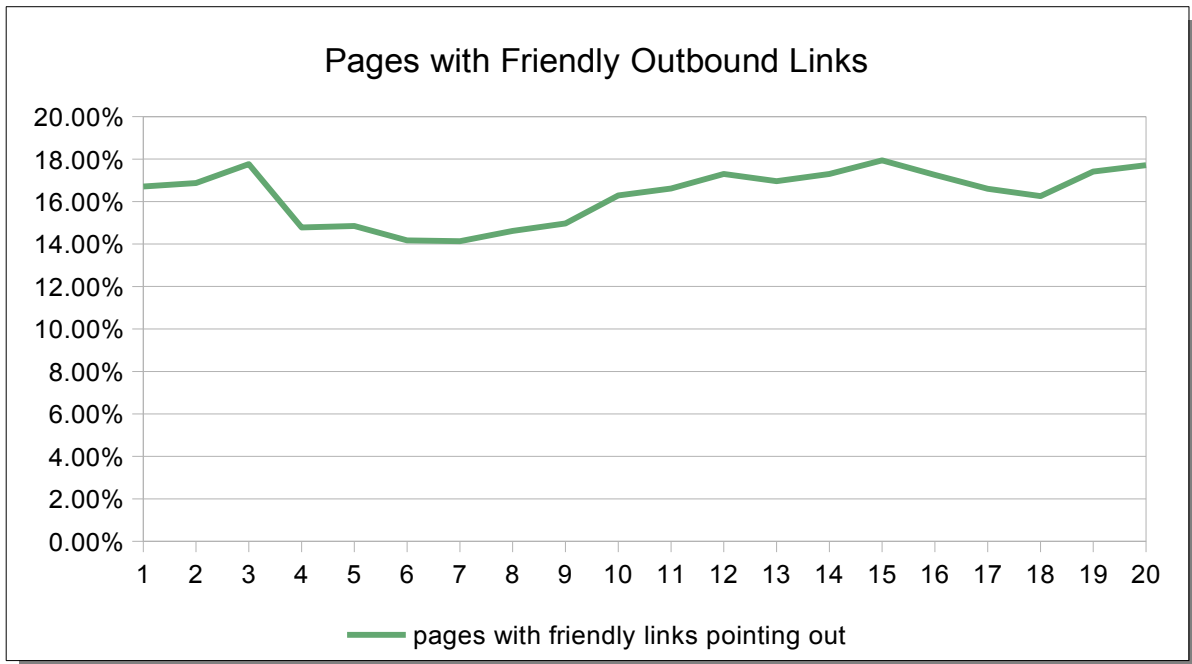
The optimal range for internal jumpmark links on well-ranking pages appears to hover around 4 to 5, suggesting a balance between complexity and navigability is preferred.

“Friendly Links”

The analysis focuses on whether top-ranking pages in Baidu's search results are extending "friendly links" or "partner links" to others, a practice still prevalent in China despite Baidu's "Moneyplant" update against link purchasing.

This aspect examines if such pages are participating in linking out to associated or potentially compensated partnerships.

We consider such partnerships (linking out) as potentially negative, so a positive correlation means the absence of such links, and a negative correlation the existence.

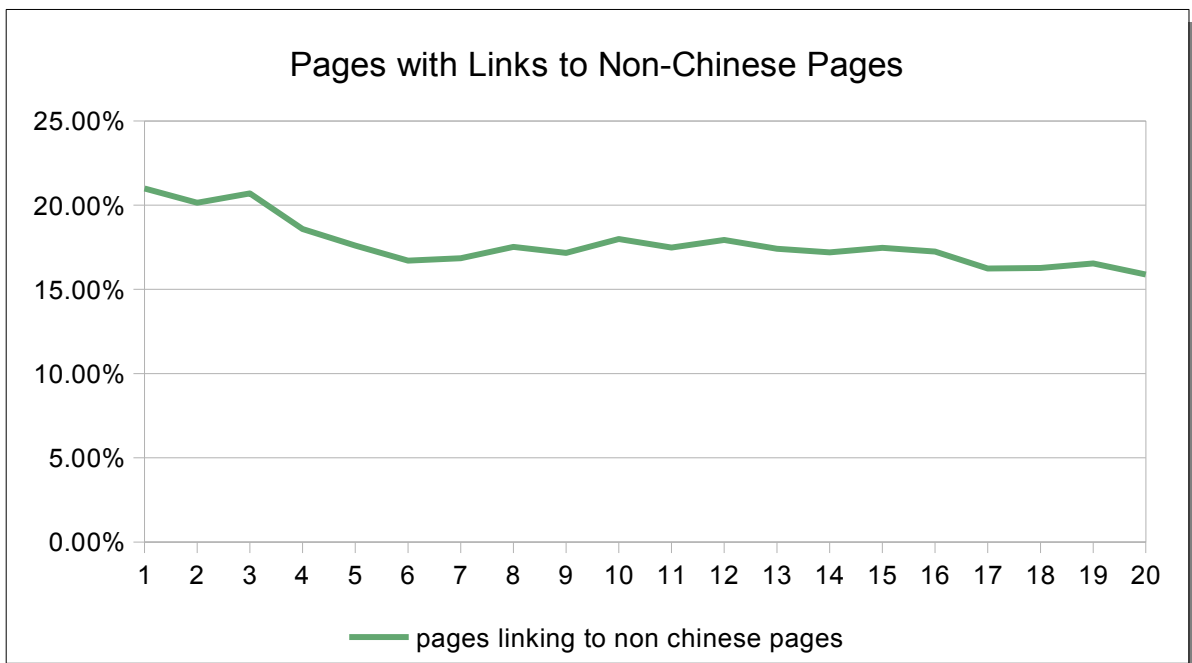


	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.02	0.002	0.007	0.04
pages with friendly links	16.3%	16.3%	16.5%	16.1%

Despite a uniform 16% prevalence of "friendly links" across varying keyword difficulty levels and minimal correlation with rankings, it appears Baidu does not currently penalize this link-building tactic – not the friendly link selling websites, and most likely not friendly links buying websites.

Links Pointing to Non-Chinese Pages

Evaluating outbound links to non-Chinese pages as a possible detrimental ranking factor, we consider the user experience impact, as navigating to foreign-language pages could potentially confuse and frustrate Chinese-speaking users, affecting site performance.



	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	-0.02	-0.05	-0.03	-0.003
links to on-Chinese pages	17.5%	16%	17.4%	18.5

Our analysis reveals a surprising 17.5% of top-ranking pages link to non-Chinese content, with a negative correlation suggesting higher-ranked pages often include such links. Despite this, competitive keyword rankings feature fewer non-Chinese links, and a dominant 82.5% of pages exclusively link to Chinese-language content, underscoring a strong preference for domestic links.

Backlinks and Domain Strength

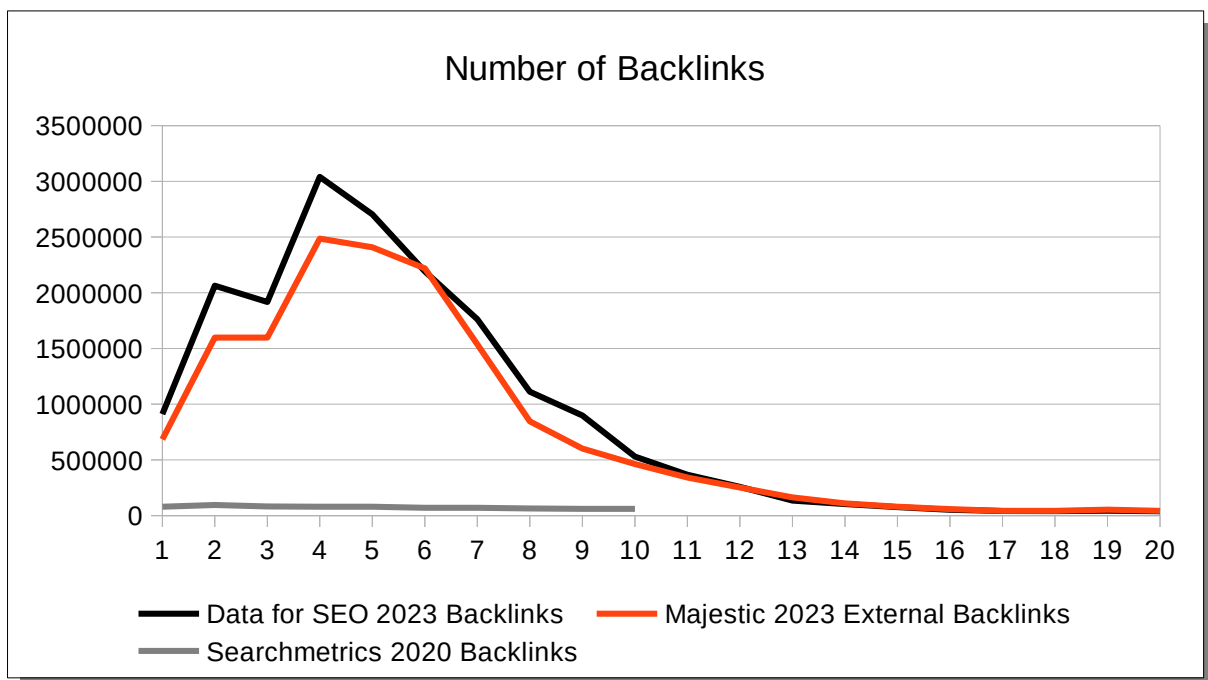
Backlinks have historically held a position of significance in Baidu SEO. As foundational components of the digital ecosystem, backlinks serve not just as navigational aids, but as endorsements of content quality and relevance. Their impact on search rankings is a topic of ongoing debate and study. Within this context, we aim to elucidate the precise role and importance of backlinks in Baidu's ranking algorithm, drawing from empirical data and comparative analyses.

In the quest to understand the backlink landscape for Baidu SEO, we face inherent limitations, primarily due to the lack of clarity on which backlinks Baidu acknowledges. Moreover, the absence of a dedicated backlink tool from Baidu further compounds the challenge.

To navigate this, we have harnessed the datasets of two esteemed SEO data vendors: Data-for-SEO and Majestic. While both platforms utilize their proprietary crawlers and may share some common data points, they each offer a unique set of backlink data. For the purpose of this analysis, our emphasis will be on discerning overarching trends rather than fixating on precise backlink counts. Metrics such as Majestic's Citation Flow and Trust Flow, or Data-for-SEO's Rank and Backlinks-Spam-Score, present intriguing insights into the backlink dynamics influencing Baidu SEO.

Number of Backlinks

The factor of "number of backlinks" presents nuanced challenges in its assessment, given Baidu's potential discretion in counting repetitive site-wide links. Furthermore, juxtaposing data from *Majestic* and *DataForSEO* with 2020's *Searchmetrics* data necessitates caution, as each employs distinct crawling methodologies and approaches to identifying backlinks (and dealing with sitewide links).



DataForSEO	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.24	0.26	0.25	0.24
average	59,210,490	61,422,890	50,713,755	64,599,733
median	355,082	342,657	329,973	387,488
avg. low50	1.3	8.1	5.2	3.3
med. low50	1	6	4	3

In the data table above we introduce 2 new values: “avg. low50” and “med. low50”. They represent the median and the average value of the lowest 50 values. The idea is to provide insights, that although the overall median and average values might be high, there are pages ranking for keywords of all difficulties, ranking with low numbers of backlinks identified by *DataForSEO*.

Majestic	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.25	0.25	0.25	0.24
average	18,951,316	18,953,593	17,197,245	20,409,982
median	319,525	375,772	323,391	304,484
avg. low50	1.3	3.8	3.4	2.5
med. low50	1	3	3	2

While the 2020 study, using *Searchmetrics* data, revealed a robust correlation of 0.9 between backlink quantity and superior rankings, the recent data paints a slightly nuanced picture.

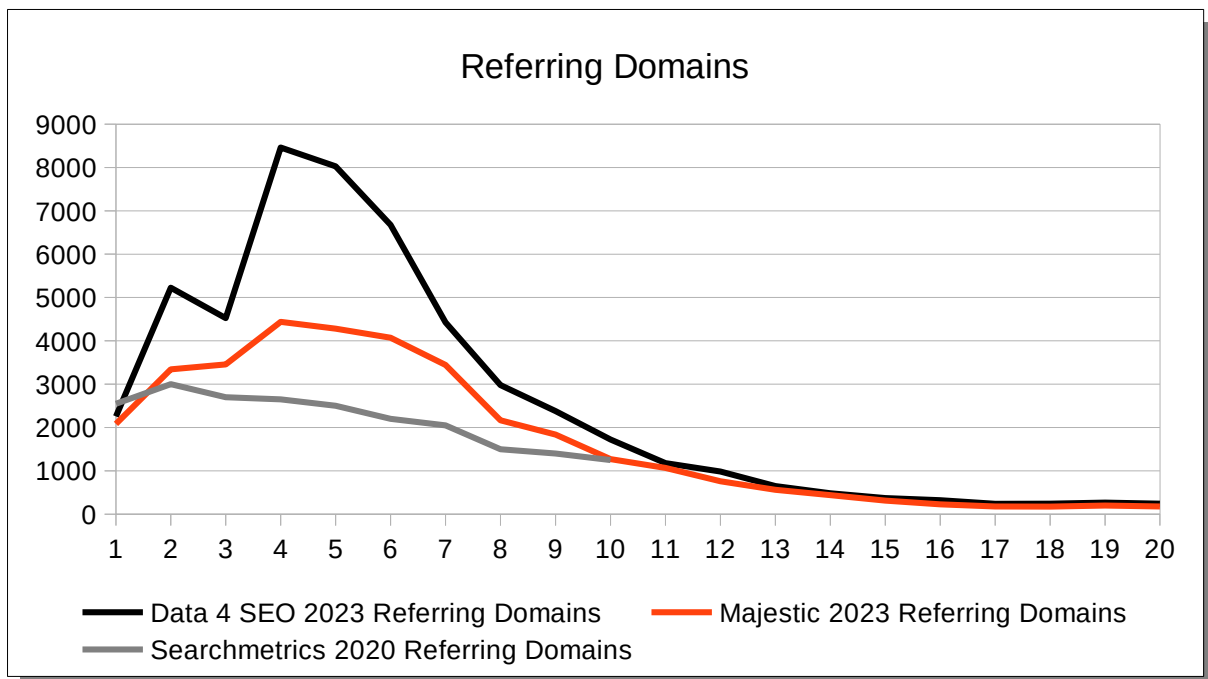
Both *Majestic* and *DataForSEO* indicate a correlation of approximately 0.25, suggesting that while the presence of numerous backlinks can be advantageous, its influence is moderate.

Notably, the median number of backlinks for top-ranking pages has seen an increase (comparing apples with oranges), with *Majestic* reporting a median of 319,525 and *DataForSEO* at 355,082 for Baidu's top 20.

However, it's imperative to understand that while these numbers provide insights into backlink trends, they underscore that backlinks are but one component in a multifaceted ranking algorithm.

Number of Referring Domains

The number of referring domains serves as a potentially more insightful metric than sheer backlink volume. However, again, drawing parallels with the 2020 study, which utilized Searchmetrics data, and comparing numbers from Majestic and DataForSEO, might get criticized by some readers of this study.



DataForSEO	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.28	0.29	0.28	0.26
average	39,781	36,799	40,205	41,663
median	1,169	1,141	1,098	1,293
avg. low50	1.1	2	2	1.4
med. low50	1	2	2	1

Majestic	Whole Keyword Set	Shorthead Keywords	Midtail Keywords	Longtail Keywords
correlation	0.27	0.28	0.28	0.26
average	17,019	15,519	16,842	18,296
median	1,038	1,057	999	1,066
avg. low50	1.3	1.6	1.4	1.2
med. low50	1	2	1	1

The analysis underscores a consistent correlation between the number of referring domains and superior rankings on Baidu.

The 2020 study, based on *Searchmetrics* data, highlighted a median of 2,341 referring domains for the top 10 positions with a robust correlation score of 0.9.

Transitioning to the present data, both *DataForSEO* and *Majestic* findings resonate with this trend. Specifically, *DataForSEO* presents a median of 1,169 referring domains for the top 20 positions, while *Majestic* reports a closely aligned median of 1,038 for the same range.

These figures, in tandem with respective correlation scores of 0.276 and 0.272, reaffirm the continued significance of diverse referring domains in achieving commendable rankings on Baidu, albeit with a slight variation across different datasets.

Expert comment:

“Link Building: Each external website that points to a website is like a vote for it. When a website receives lots of referring domains and external links, this same website will be seen and considered as very popular for Baidu and will rank it better in its organic search results. Baidu still considers massive link building as a good ranking factor.”

Veronique Duong, International SEO & Marketing Digital Expert, veronique-duong.com

Domain Authoritiveness

Baidu likely employs an instrument analogous to Google's Pagerank to ascertain a domain's authoritiveness, subsequently integrating this measure into their ranking algorithms.

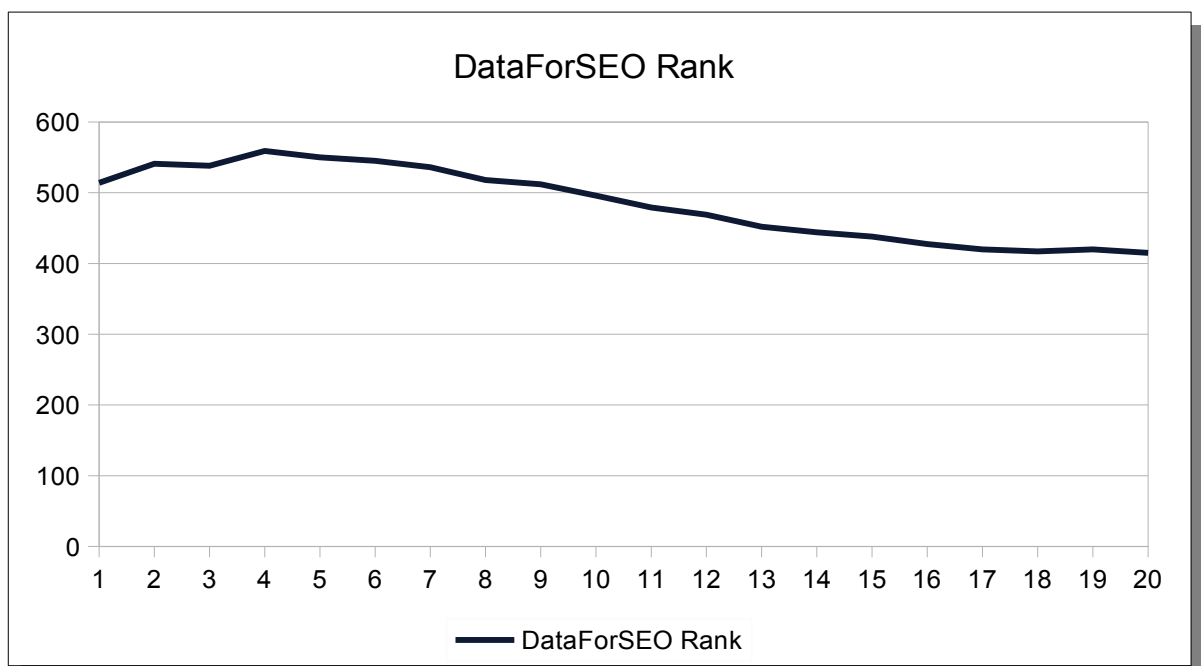
The 2020 study harnessed Searchmetrics' "Page Strength" as such a surrogate, revealing a notable negative correlation of 0.7. Moreover, an average Median Page Strength of 5.17, on a scale from 0 to 10, was observed, underscoring the nuanced nature of domain authority in the current digital landscape.

DataForSEO Rank

In assessing domain authority, DataForSEO introduces the "Rank" metric, a measure primarily grounded in a domain's backlink profile. Domains with extensive authoritative backlinks typically achieve scores above 500.

Those with a substantial backlink foundation hover between 200-300, while emerging domains, characterized by limited backlinks, generally fall within the 25-35 range.

This metric provides a structured approach to gauge a domain's standing in the vast digital ecosystem. For this analysis we are not looking at the individual Rank of the ranking URLs, but at the Rank of the ranking domains on the different positions.



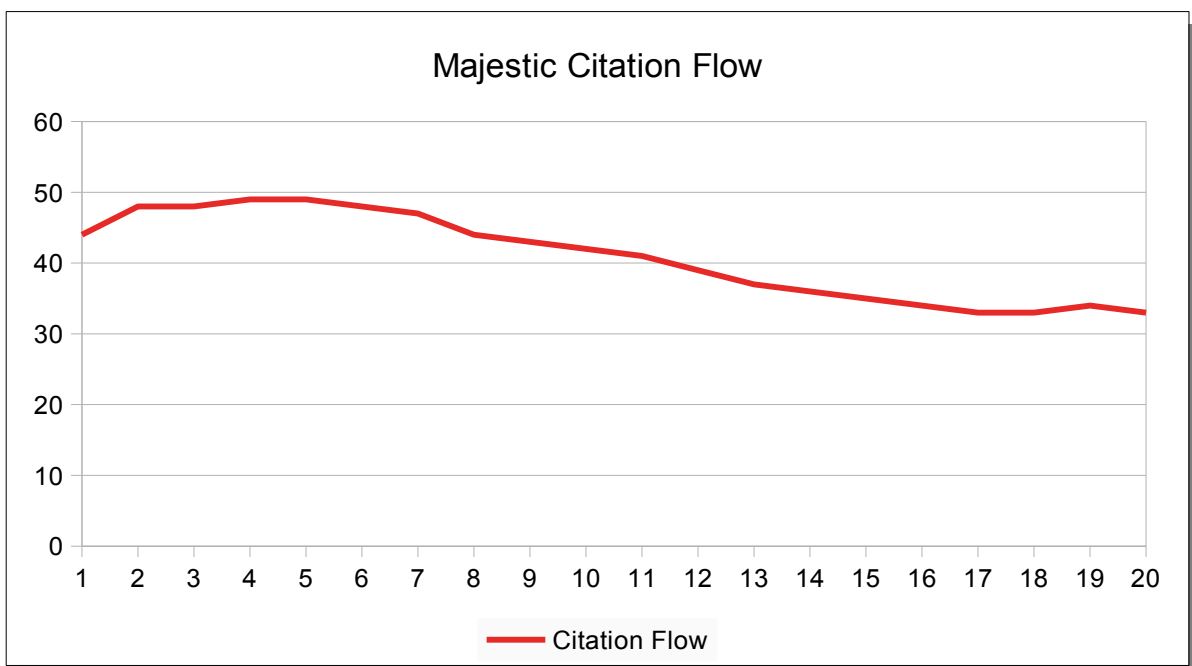
Based on DataForSEO's "Rank" metrics, there's a notable correlation of 0.2303, suggesting a relationship between a domain's authoritiveness and Baidu rankings. The median value for domains within Baidu's Top 20 stands at 480. A more granular inspection shows domains at the pinnacle, positions 1 to 4,

generally score above 540, while those towards the tail end, positions 17 to 20, hover around the mid-400s. This distribution underscores the potential influence of domain authority on Baidu's ranking mechanism.

Majestic Citation Flow

Majestic's Citation Flow serves as an equivalent to Google's Pagerank, quantifying the link equity or "power" inherent in a website or specific link.

With a score range of 0-100, Citation Flow operates in tandem with another Majestic metric, Trust Flow. For this analysis we are not looking at the individual Citation Flow of the ranking URLs, but at the Citation Flow of the ranking domains on the different positions.



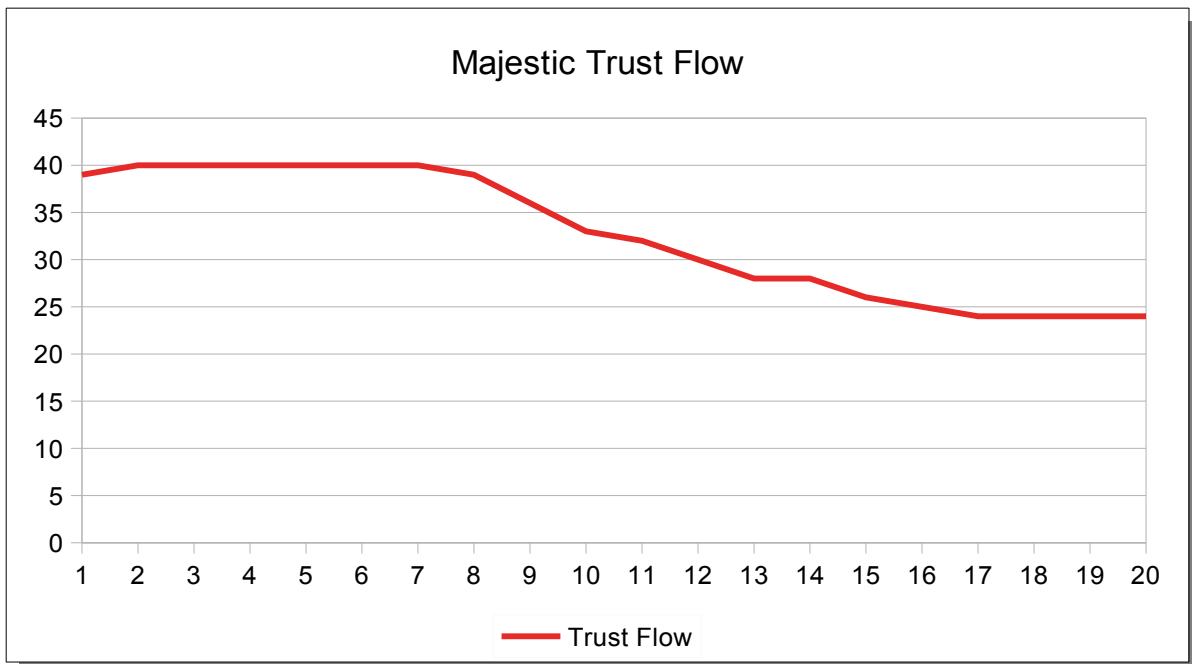
The data indicates a correlation of 0.246 with Baidu rankings, suggesting a moderate link between domain authority and ranking position. The median score for Baidu's Top 20 is 41, with higher ranks typically having slightly elevated scores. This highlights domain authority's potential role in Baidu's ranking algorithm.

Majestic Trust Flow

Majestic's Trust Flow measures the quality of a site's link profile on a 0-100 scale. Derived from a manual review of trusted seed sites, it gauges the proximity of a site to these seeds.

Higher scores indicate linkage to trusted sites, while lower scores can suggest potentially questionable links. As we can expect, that these manually selected “trusted seed sites” are rather of English (or western) origin and rather are not trusted sites from Mainland China, we might expect less correlation to rankings on Baidu.

For this analysis we are not looking at the individual Trust Flow of the ranking URLs, but at the Trust Flow of the ranking domains on the different positions.



We see an unexpected correlation of 0.224 with Baidu rankings. With an average score of approximately 34.58 across websites, the median Trust Flow for the top 20 positions on Baidu stands at 32. Notably, sites ranking in the top 8 positions tend to have higher Trust Flow values, often exceeding 39. As positions descend

from 9 to 20, there's a visible decline in the median Trust Flow, underscoring the potential significance of link quality in Baidu's ranking considerations.

Expert comment:



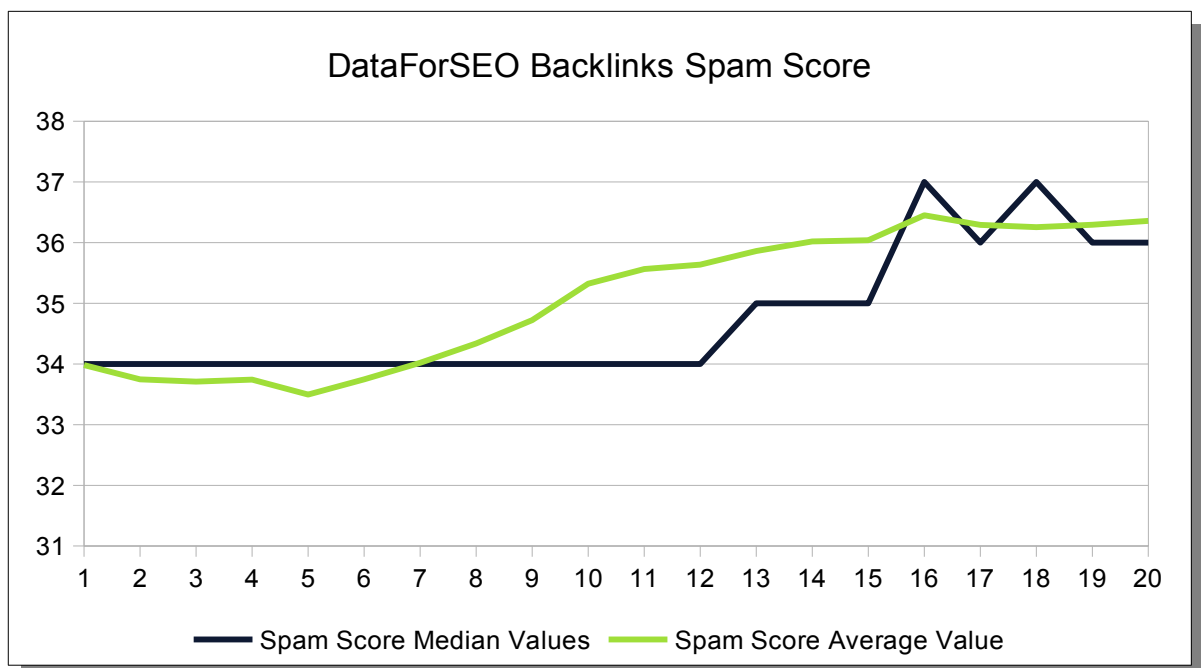
“While backlinks undeniably influence rankings on Baidu, link building in China is more intricate. In Chinese digital PR and content marketing, the emphasis lies on exposure and awareness, making it uncommon to acquire dofollow links through these methods. However, it is crucial to note that purchasing links is strictly discouraged.”

Christina Xu, Founder of China Marketing Corp, www.chinamarketingcorp.com

DataForSEO Backlinks Spam Score

DataForSEO's "Spam Score" is a proprietary metric denoting the perceived spam level of web content, ranging from 0% to 100%. Derived from 18 distinct signals, including the assessment of specific TLDs like .cn, this score categorizes sites as having low (0%-30%), medium (31%-60%), or high (61%-100%) spam levels.

While TLDs such as .cn are flagged by this system, within the context of Baidu SEO, these are viewed as valid local TLDs. Thus, direct correlations between this score and Baidu rankings should be approached cautiously. For this analysis we are not looking at the individual Spam Score of the ranking URLs, but at the Spam Score of the ranking domains on the different positions.



Because the Median value chart looked a little strange to us (we double checked, the data is right) we added the average values for the top 20 positions to the chart. The correlation score is still based on the whole set of roughly 100K median values.

The DataForSEO's Spam Score demonstrates a weak correlation (0.08) with Baidu rankings. The average score across analyzed domains is 35.14, with a consistent median score of 34 for the top 12 positions. However, positions 13 to 20 exhibit a slight increase, indicating that domains with marginally higher spam scores can still rank within Baidu's top 20.

About our Data Suppliers (alphabetically)

Our special THANK YOU goes to our data providers, who are helping to demystify the secrets of SEO for China.

Thank you!

DataForSEO

A leading data solutions provider for the marketing technology industry, DataForSEO offers reliable, granular, and up-to-date data on every aspect of digital marketing, web analytics, advertising, and SEO. Its convenient pay-as-you-go pricing model, variety of APIs, and customer-oriented approach make DataForSEO the obvious destination for anyone interested in SERPs, backlinks, keywords, and other digital marketing data.

Among other solutions, DataForSEO offers both real-time and historical search engine results data for Google and Bing - which is available through its powerful [SERP API](#) and proprietary databases. On top of that, DataForSEO Backlinks API provides [high-quality backlink data](#) that is on par, and in some respects even superior to other prominent vendors.

Web: <https://dataforseo.com/>

Dragon Metrics

Dragon Metrics is an all-in-one SEO platform based in Hong Kong. The platform was originally created in 2010 as internal tool at an APAC-focused SEO agency. While there were plenty of tools on the market focusing on Western and English-speaking markets, there was nothing available that delivered the data the agency needed to provide value for our clients in regional markets and non-Google search engines like Baidu, Naver, and Yahoo! Japan.

English | Login

DRAGON METRICS | Product Tour | Why Dragon Metrics? | Pricing | Support | About | Contact | Get a Demo | **Free Trial**

Stop the Busywork. Get Results.

Improve Ranking & Automate Reporting
with the First Truly Global SEO Platform

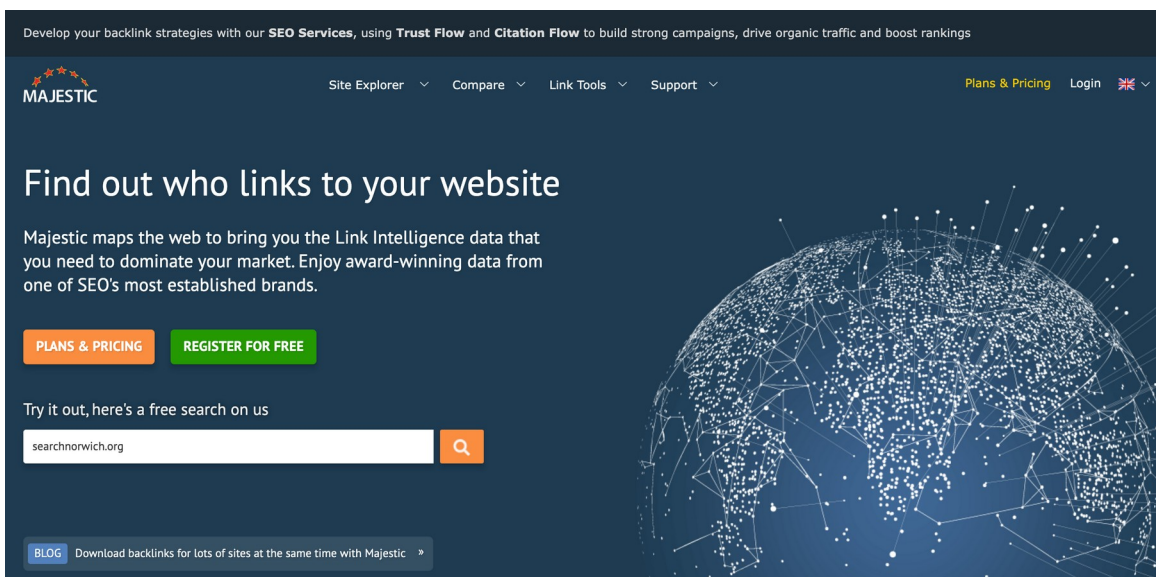
[TAKE THE TOUR](#) | [START FREE TRIAL](#)

While their legacy is still that they're the most comprehensive SEO tool that covers non-Google search engines, they've built a name for themselves as an innovative solution not just in Asia, but the world over.

Web: <https://www.dragonmetrics.com/>

Majestic SEO

Launched in 2004, Majestic (previously known as MajesticSEO) is a pioneer in surveying and mapping the world wide web, creating one of the largest Link Intelligence databases available commercially. This internet map is used by SEOs, New Media Specialists, Affiliate Managers and online Marketing experts for a variety of uses surrounding online prominence including Link Building, Reputation Management, Website Traffic development, and Competitor Analysis.



Develop your backlink strategies with our **SEO Services**, using **Trust Flow** and **Citation Flow** to build strong campaigns, drive organic traffic and boost rankings

MAJESTIC

Site Explorer ▾ Compare ▾ Link Tools ▾ Support ▾

Plans & Pricing Login 🇬🇧 ▾

Find out who links to your website

Majestic maps the web to bring you the Link Intelligence data that you need to dominate your market. Enjoy award-winning data from one of SEO's most established brands.

[PLANS & PRICING](#) [REGISTER FOR FREE](#)

Try it out, here's a free search on us

[Q](#)

[BLOG](#) Download backlinks for lots of sites at the same time with Majestic ▶

As link data is also a component of search engine ranking, understanding the link profile of your own, as well as competitor websites can empower rational study of Search Engine positioning. Majestic is constantly revisiting web pages and sees billions of URLs a day.

Based in the United Kingdom, Majestic has won multiple SEO awards and has an international client base, attracted by its focus on backlink intelligence, transparency of operation, reputation for quality, and value of offering.

Web: <https://majestic.com/>

About Jademond Digital

Founded in 2012 by Kun Tang, Jademond Digital began its journey in Guangzhou, China, as a Baidu SEO-focused agency catering primarily to foreign companies. Quickly recognizing the broader needs of its clients, the agency expanded its offerings to include SEA/PPC services, Social Media Advertising, and E-Commerce platform management for major platforms like Tmall, Taobao, and JD. This evolution was supported by assembling a team of specialists in these diverse fields, ensuring top-tier service for our clients.



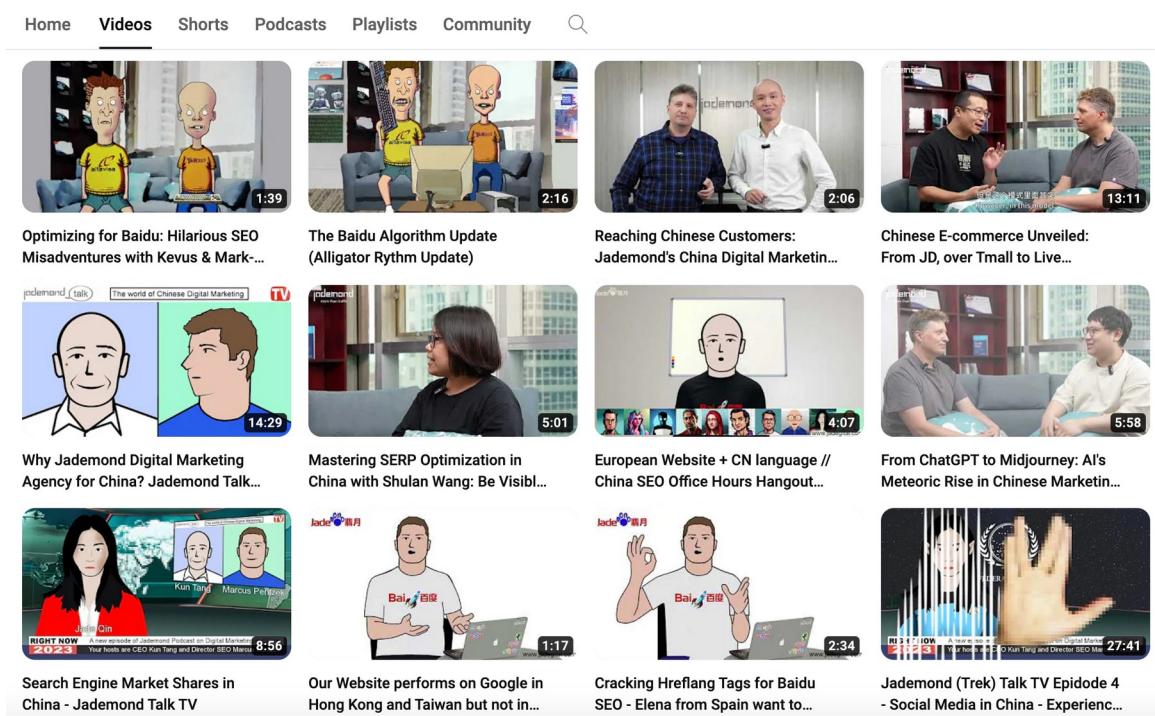
Kun Tang, a Chinese native with an educational and early professional background in Germany, founded Jademond Digital with a vision to bridge Western and Eastern business cultures. This cross-cultural understanding is encapsulated in the company's name - 'Jademond', combining the Chinese symbol of heritage, 'Jade', with the German word 'Mond' (meaning moon). This name reflects not only Kun's personal journey but also signifies the cultural significance of the moon in Chinese mythology, further linking back to the company's roots.

In December 2023, Jademond Digital strengthened its international ties by welcoming Marcus Pentzek, a German native, as Director of SEO and partner. With Marcus's joining, the company expanded its presence by opening an office in Tianjin, near Beijing. This move marked a significant step in reinforcing Jademond's connection to Germany, complementing its existing roster of German clients.

Today, the 50-strong team at Jademond Digital is committed to delivering exceptional digital marketing solutions and services. We specialize in helping Western businesses navigate and succeed in the complex and vibrant Chinese digital marketplace, blending the best of both Western and Eastern digital marketing strategies.

Web: <https://www.jademond.com/>

Jademond Digital's Youtube Channel is filled with information and fun



Starting with our animated Video-Podcast discussing questions sent to us from our audience, over discussing questions in Google's John Mueller's style

Webmaster Hangouts – featuring animated Kun Tang and focusing on Baidu Search – discussing questions in Google’s Matt Cutts’ video style – featuring animated Marcus Pentzek and focusing on Baidu SEO – Expert interviews letting you get to know our many Jademond experts and their focus topics – over to the parody of Kun and Marcus in a style that reminds of Beavis and Butthead, the Jademond Youtube channel has infotainment for everyone interested in Digital Marketing for China.

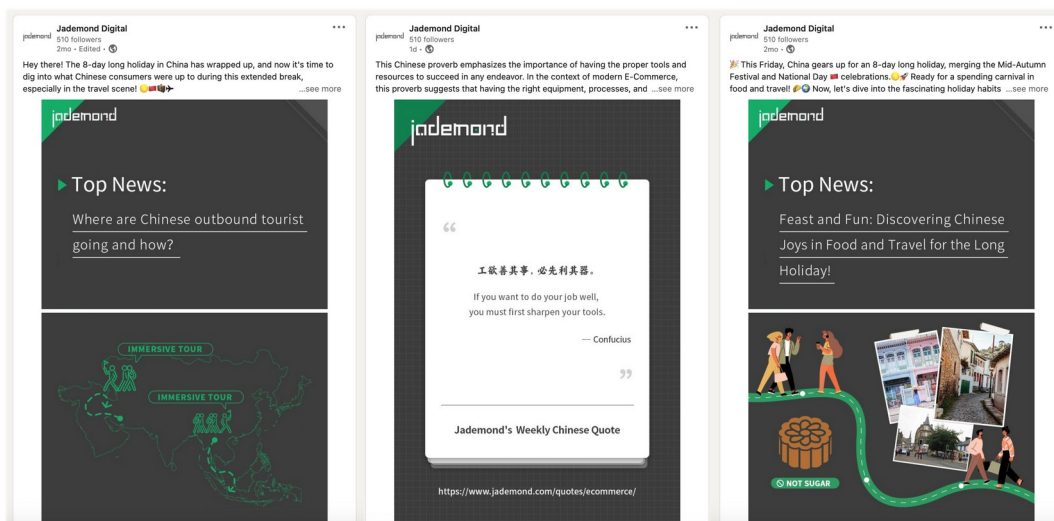
Youtube: <https://www.youtube.com/@jademonddigital2271>

Find regular insights into Digital Marketing for China in our magazine:

In our Jademond Digital Magazine you can find tutorials on how to set up Baidu Webmaster Tools, information about the real search engine marketshare in China, a list of all Baidu-Search-Algorithm-Updates and how they affect your China-SEO-Strategy, hand-on-tips for successful SEO for China and much more.

Magazine: <https://www.jademond.com/magazine/>

China News and Culture on our LinkedIn page:



LinkedIn: <https://www.linkedin.com/company/jademond-digital/>

Glossary of Terms

Backlinks

Backlinks, also known as "inbound links" or "incoming links," are links from one website to a page on another website. They are considered vital for SEO because they signal to search engines that others vouch for your content, which can improve a site's ranking on search engine results pages (SERPs). The quality and relevance of backlinks can significantly affect a website's visibility and authority in its niche.

Baidu-Property

Refers to sub-domains under Baidu.com, representing various services provided by Baidu. An illustrative example is Baike.baidu.com, Baidu's counterpart to Wikipedia. Observations from this study indicate a trend of Baidu prioritizing and ranking its own services higher than other search results.

Correlation Score

This score quantifies the potential relationship between a specific signal or factor and top Baidu rankings, based on observed trends. A positive or negative value represents the direction and strength of the correlation, with extremes in the data influencing the score's polarity.

Friendly Links

"Friendly links," also known as "partner links," refer to hyperlinks placed on a website that lead to business associates or partner sites. They are often exchanged between websites to promote mutual visibility or as part of a link network strategy, although they must be used judiciously to avoid penalties from search engines for manipulative link practices. In China it is a widely seen link building practice.

Median Baidu Top 20

The "Median Baidu Top 20" refers to the middle value derived from the rankings of Position 1 to Position 20 in the Search Engine Results Pages (SERPs) for a dataset of 10,000 keywords. To ensure precise and representative data interpretation, the median, rather than the mean or average, has been employed as the principal metric in the majority of our charts.

Organic Rankings

Positions a website occupies on a SERP that are not influenced by paid advertising. These rankings are determined by a search engine's algorithm based on various factors like content quality, site structure, and backlinks. It's noteworthy that paid listings can appear above these organic rankings on the results page.

Position 1

Refers to the topmost "organic" result on a SERP. Due to its prominence and potential for high user engagement, Position 1 is highly sought after by SEO professionals and business leaders.

Ranking

In the context of SEO, "ranking" denotes the specific position a website occupies on the SERP. For instance, a website situated in Position 1 holds the topmost rank.

Ranking Factor

A discernible signal correlated with enhanced search engine rankings. While search engine companies do not officially enumerate definitive ranking factors, analytical deductions can be made regarding how various website elements influence their position in organic search results.

SERP (Search Engine Results Page)

A page presented to users following a search query. It displays a list of relevant results, with the topmost result denoted as Position 1.

TLD (Top Level Domain)

A domain name classification signifying quality and authority. Predominantly, TLDs are country-specific or generic domain names such as .com, .co.uk, .de, and .ch. For clarity, www.website.com utilizes a TLD due to its .com extension. Conversely, www.website.info, employing the .info extension, is not considered a TLD as .info domains are not typically linked to high-quality websites.

Short Keyword

Distinct from “short head keyword”, a "Short Keyword" denotes search queries of just 1 or 2 characters. Predicated on empirical observation, such keywords more commonly manifest as "exact match" in SEO-relevant positions in contrast to extended queries. The label "short" strictly relates to character count and doesn't infer search volume, although a correlation with higher volumes is often noted. Nonetheless, a 2-character keyword might align with long-tail characteristics, attracting minimal search interest.

Still Short Keyword

A search query comprised of 3 or 4 characters. The designation "still short" is based purely on the keyword's character count, focusing on its concise nature.

Common Length Keyword

Refers to search queries that span 5 to 6 characters. This character range is frequently observed and represents the typical length for many keywords. While there's often a correlation with mid-tail and long-tail keywords, it's not a definitive association.

Long Keyword

Refers to search queries that span 7 or more characters. The descriptor "long" is solely tied to the keyword's character count, not implying other qualities. While these often correlate with long-tail keywords, exceptions exist; some lengthy keywords can register high search volumes, positioning them as short head keywords.

Short Head Keywords

Terms or phrases that capture a broad search intent. While these keywords might often appear concise, the designation "short" in "short head" does not allude to their physical length. Rather, it signifies their general nature and high search volume, contrasting with more specific, long-tail keywords.

Mid Tail Keywords

Search queries that sit between the broadness of short head keywords and the specificity of long-tail keywords. They capture a moderate search volume and present a balanced competitive landscape for SEO efforts.

Long Tail Keywords

Search queries characterized by their specific nature, often targeting niche topics or audiences. While they can be lengthier in composition, the term "long" in "long tail" doesn't denote their physical length. Even concise queries, such as those with just two characters, can be defined as long tail if they address a highly specialized search intent with only low search volume.

UGC

UGC (User-Generated Content) refers to any form of content created by users rather than publishers or brands, such as comments, posts, images, videos, and reviews, typically shared on websites, social media platforms, and online forums.

This content can influence SEO as search engines consider it as a signal of engagement and relevance.

URL

A URL (Uniform Resource Locator) is the address used to access a resource on the Internet, providing the means to retrieve a web page or file. It consists of a protocol (like HTTP or HTTPS), domain name, and path, and may include additional components like query parameters or anchors. URLs are fundamental to navigation and interaction on the World Wide Web.