

# **Comprehensive Guide to Automated Keyword Selection for SMEs as a Part of an AI-Driven Textual Content Generation**

**Author:** Anton Shelepov

August 13, 2024

## **Abstract**

This document outlines an AI-driven approach to automated keyword and keyword cluster selection as a pre-step for AI-based content generation for small and medium-sized enterprises (SMEs), with a focus on the coaching industry. By leveraging large language models (LLMs) and data analysis techniques, the study identifies high-performing keywords and keyword clusters, providing crucial insights into the competitive landscape.

# 1. Introduction

## 1.1 Purpose

Generating meaningful and engaging content represents a significant challenge for numerous small and medium-sized enterprises (SMEs). Despite recognizing the importance of high-quality content for enhancing search engine rankings, increasing web traffic, and engaging potential customers, many SME website owners grapple with developing effective textual content strategies. This struggle is further exacerbated when efficiency and cost-effectiveness are of critical importance. The purpose of this document is to explore an automated approach for keyword/keyword cluster selection as a prerequisite for AI-based textual content generation tailored for SMEs, with a focus on the personal coaching industry as a case study.

## 1.2 Importance

High-quality textual content is a crucial driver of digital marketing success, offering benefits such as enhanced visibility, increased engagement rates, and improved conversion rates. However, SMEs often lack the resources and expertise to consistently produce competitive content. Automated keyword/keyword cluster selection and AI-driven content generation present promising solutions that bridge this gap, providing SMEs with scalable and cost-effective strategies to enhance their online presence and competitive edge.

## 1.3 Overview of Document Structure

The document is organized as follows:

- **Introduction:** Discusses the document's purpose, importance, overall structure, current limitations, and scope.
- **Methodology:** Outlines the approach for keyword selection and AI-based content generation, including business field selection, keyword generation, data collection, and AI's role.
- **Data Preparation and Dataset Description:** Details the steps for data collection, keyword and domain clustering, and preparation processes.
- **Keyword and Content Analysis:** Describes the analytical framework used to evaluate and refine keyword selection and content generation.
- **Results and Discussion:** Presents the analysis results, key findings, implications for SMEs, and discusses these in the context of the personal coaching industry.
- **Keyword Choices Based on the Analysis:** Provides the final keyword set, strategies, and recommendations for future research.
- **Conclusion and Recommendations:** Summarizes the study and offers final thoughts and next steps.
- **Appendices:** Includes raw data, tools and resources, and additional information.

## **1.4 Current Gap**

Currently, many SMEs either rely on manual content creation methods or outsource their content needs, which can be both time-consuming and costly. Existing automated solutions often lack the sophistication required to produce contextually relevant and engaging textual content tailored to specific business needs. This document addresses this gap by presenting a systematic approach to automated keyword selection, offering a practical and scalable solution for SMEs.

## **1.5 Scope and Limitations**

This study focuses primarily on the personal coaching industry as a representative example of an SME. While the methodologies and findings are expected to be broadly applicable across various SME sectors, the specificity of the business type should be considered when extrapolating results. Additionally, the study assumes access to particular tools and resources for data collection, keyword analysis, and AI content generation, which might not be universally available to all SMEs.

## **2. Methodology**

This section provides a detailed account of the research procedures and methodologies applied in this study.

### **2.1 Selection of Business Field and Categorization**

The coaching industry was selected for this study due to its rapid growth and significant reliance on an online presence. In recent years, the demand for personal and professional coaching services has surged, driven by increasing recognition of the value of personal development and continuous learning. This industry heavily depends on digital platforms such as websites, blogs, and social media for client engagement, service promotion, and brand building.

A colleague based in Berlin provided her human potential coaching website (theklasalmon.com) for testing purposes. This allowed for practical insights and validation of the proposed methodology.

Recommendations based on the analysis will be implemented into the colleague's website. The ranking state after the implementation will be compared with the current one to evaluate the effectiveness of the recommendations. The current state of the rankings for the domain theklasalmon.com is discussed further below in Section 2.4 Current State.

### **2.2 Initial Keyword Generation**

In the initial phase of the study, the internal AI-based service ask.IONOS was utilized to generate a comprehensive list of relevant keywords for the coaching

industry. This service leverages advanced language models to provide contextually appropriate keyword suggestions. Both short-tail (e.g., “life coaching”, “personal coach”) and long-tail keywords (e.g., “wellbeing assessment and planning coach”, “confidence and self-esteem coaching”) were considered to cover a broad spectrum of potential search queries. This approach ensures the inclusion of diverse and relevant keywords that can capture varied search intents.

### **2.3 Data Collection**

The finalized list comprised 264 keywords. For each keyword, a Google Search request was made, and the first 50 organic results were scraped, corresponding to the first five Google search result pages. This data collection approach aimed to capture a comprehensive snapshot of the competitive landscape and the prevalent content strategies within the coaching industry. The scraped data included URLs, as well as corresponding indexes, which were essential for subsequent analysis. Overall, the dataset consists of 13,114 rows, each representing a Google Search result.

### **2.4 Current State**

An analysis of the data gathered revealed that the domain *theklasalmon.com* appeared in the search results eight times. The following keywords yielded one or more subpages of this domain: *Human Potential Coach Berlin*, *certified human potential coach*, *Human Potential Coaching in Berlin*, *Certified Human Potential Coach in Germany*, *Human potential coaching*, *Human Potential Coaching in English and German*, *thekla salmon*.

Several issues have been identified with these keywords:

- Homogeneity: The keywords are, to a significant extent, very similar to each other, which might limit their effectiveness in capturing a diverse audience.
- Search Volume: The keyword *Thekla Salmon* is likely to have little search volume, thus limiting its potential impact on attracting new visitors.

### **2.5 Role of AI in Keyword Selection and Content Generation**

Large Language Models (LLMs) play a critical role in this research, particularly in the processes of keyword/keyword cluster selection and content generation. LLMs, trained on vast amounts of publicly available data from the internet, provide valuable insights and generate high-quality recommendations. These models excel in understanding context, semantic relationships, and content relevance, making them indispensable tools for refining keyword strategies and generating engaging textual content. Their ability to analyze and synthesize information at scale ensures that the content produced is not only competitive but also aligned with the target audience’s expectations and needs.

### 3. Data Preparation and Dataset Description

Comprehensive data preparation involved several steps to ensure the accuracy and relevance of the dataset:

#### 3.1 Domain Extraction and Analysis

For more effective aggregation and analysis, domains were extracted from the URLs obtained. This step aimed to facilitate a domain-level analysis that would offer a more aggregated view of the competitive landscape within the coaching industry. This process resulted in the identification of 6,080 unique domains. Notably, domains such as facebook.com and m.facebook.com, and alike, were counted as separate domains.

#### 3.2 Business Description Acquisition

Each domain was subsequently queried using ask.IONOS to obtain a business description. This process leverages ask.IONOS’s access to a large language model (LLM) database to provide accurate business descriptions. Notably, this method does not rely solely on the domain name but utilizes contextual data from the LLM’s training corpus. One crucial criterion was that ask.IONOS was instructed not to guess the business description; if it didn’t know the answer, it was instructed to respond with “I do not know.” This precaution ensured the accuracy and reliability of the business descriptions provided.

As a result, this method yielded 3,901 business descriptions. Examples of the descriptions are presented in the table below.

Table 1: Examples of business descriptions

| Domain                       | Business Description  |
|------------------------------|---|
| coachtrainingalliance.com    | Institute offering certified coach training programs                  |
| thewholistichhealthcoach.com | Australian site offering holistic health coaching                     |
| kimmariecoaching.com         | Provides life coaching and personal development services by Kim Marie |
| dr-bock-coaching-akademie.de | Coaching academy offering training and certification in coaching      |

#### 3.3 Keyword Clustering

The next critical step involved clustering the collected keywords based on semantic meaning using ask.IONOS. This process identified nine distinct clusters, each representing a different aspect or topic within the coaching industry.

##### Potential Process Flaws:

- **Homogeneity in Clusters:** If the clusters are too homogeneous, they might not capture the diversity of search intents within the industry.

- **Overlapping Keywords:** Keywords might belong to multiple clusters, leading to overlaps that could complicate subsequent analyses.
- **Insufficient Granularity:** Too few clusters might not provide sufficient granularity, whereas too many clusters might dilute the significance of each group.

### 3.4 Domain Clustering

Clustering techniques were applied to aggregate domains based on the businesses and institutions they represent.

**3.4.1 Business Description Clustering** Initially, domains were clustered based on business descriptions provided by ask.IONOS. However, this method did not yield satisfactory results due to the brevity and lack of unique tokens in the descriptions. Moreover, around 30% of the domains in the dataset did not have a Business Description.

**3.4.2 Website Textual Content Clustering** A secondary approach involved clustering domains using the textual content from the websites. Websites were scraped, and relevant textual content was extracted. This method faced challenges as many websites lacked comprehensive, well-defined content, leading to insufficient clustering results.

**3.4.3 Similarity Score-Based Clustering** A final method utilized the website theklasalmon.com (the colleague’s initial website) as a base for defining the Similarity Index. Cosine similarity was employed to measure the similarity of textual content from other websites. This approach calculated the distance of the embeddings represented as vectors, enabling the identification of competitors and assessing the similarity of their businesses to the colleague’s business. This method, combined with business descriptions and thorough data cleaning techniques, resulted in six identifiable clusters. Although not perfect, this method provided a robust foundation for further analysis.

While implementing clustering based on the textual content, thorough data preparation was conducted, excluding textual content from subpages such as /privacy, /contact, etc., as well as their German equivalents.

This score was divided into six clusters or indexes, where ‘1’ represents the best fit and ‘6’ the least fit.

### 3.5 Dataset Description

The dataset will be provided as a separate file attached to this document.

**3.5.1 Legend** The table below provides descriptions for the columns used in the tables throughout the document.

**Table: Column Descriptions**

| Column Name         | Description  |
|---------------------|--|
| Keyword             | The specific search term entered into Google   |
| Keyword Cluster     | Represents the semantic group to which each keyword belongs  |
| Keyword Type        | Indicates whether the keyword is categorized as a long-tail or short-tail  |
| Domain              | Denotes the domain name retrieved from a Google search result  |
| Is Relevant         | A binary indication of the <b>relevance</b> (in-depth explanation of the term below) of the domain                         |
| Language            | The language in which the website’s content is written, determined by scraping the website                                 |
| Business Desc       | A descriptive summary of the business or institution behind the domain, generated with ask.IONOS                           |
| Similarity Index    | A measure of the similarity between the website and the base website (theklasalmon.com)                                    |
| URL                 | The specific Google search result URL  |
| URL Index           | The position of the URL in the Google search results   |
| Nr Keywords         | Number of keywords for which a domain is listed under the Google Search results  |
| Nr Keyword Clusters | Number of keyword clusters for which a domain is listed under the Google Search results                                    |
| Nr Appearances      | Overall number of appearances of a domain in Google Search results based on the keywords analyzed                          |
| Keyword Strategy    | Definition of a keyword strategy based on comparisons of <i>Nr Keywords</i> and <i>Nr Keyword Clusters</i> to the averages |

**3.5.2 Explanation of the Term ‘Relevance’** Relevance, particularly in the context of Google Search results, is a complex and multifaceted concept. It is not a definitive measure and can vary significantly depending on the individual performing the search. The relativity of relevance is best illustrated through the following points:

**Non-Definitive Nature:**

Relevance in Google Search results is inherently subjective and situational. What might be considered highly relevant for one person may not hold the same importance for another. Google’s algorithms rank pages based on numerous factors that statistically correlate with user satisfaction, but this relevance is a general approximation rather than an absolute measure.

**Subjectivity of Relevance:**

The relevance of search results is relative to the searcher’s intent, context, and expectations. This can be demonstrated through the following example.

### **Illustrative Example: Keyword “Personal Development Coach”**

Consider a user searching for the term “personal development coach.” The effectiveness of the search results largely depends on how well the returned domains match the searcher’s intent.

- **Relevant Results:** Domains such as canadacoachacademy.com, which offers “Coaching certification programs in Canada,” and qualitymind-global.com, which focuses on “Mindfulness and mental wellness programs,” are likely to be deemed highly relevant. These sites closely align with the search intent by providing direct and valuable content related to personal development coaching.
- **Irrelevant Results:** Conversely, a domain like indeed.com, although it may contain some related information, does not exclusively concentrate on personal development coaching. Instead, it serves as a global job search engine, thus only remotely delivering content similar to the keyword provided. This misalignment renders indeed.com less relevant for this specific search query.

Based on this logic, all domains in the database were classified accordingly. Domains were assessed and labeled as either relevant or not relevant depending on how well their content matched the keywords’ semantic and contextual expectations. This structured classification enables a more precise evaluation of keyword performance, ensuring that the selected keywords resonate well with diverse user intents within the coaching industry.

By acknowledging the subjective nature of relevance and employing a methodical approach to evaluate it, this analysis aims to discern and prioritize keywords that offer the highest potential for meeting the varied needs of searchers. This ensures that the keywords/keyword clusters identified are not only competitive but also meaningful.

For further analysis, 4,710 domains were identified as relevant.

## **4. Keyword and Content Analysis**

This section provides insights into analysis framework of each research question, as well as answers.

The analysis framework section outlines the methodologies and procedures applied to identify relevant domains per keyword and prioritize keywords for further content generation. This structured approach ensures that the keywords selected are not only relevant but also competitive within the given industry. The ultimate goal is to identify high-performing keywords that can augment the digital presence of SMEs in the coaching industry.

### **4.1 Relevant Domains per Keyword**

To identify the most pertinent keywords for further analysis, the following steps were undertaken:



For each keyword, the relevance share of the Google search results was calculated. This value represents the proportion of search results that were marked as relevant. An excerpt of the table is presented below:

Table 3: Keywords and the Share of Relevant Results

| Keyword                                 | Share Relevant Results | Total Results |
|---|------------------------|---------------|
| Holistic health coaching services       | 0.94                   | 50            |
| Life coach in Berlin                    | 0.92                   | 50            |
| Performance coaching services in Berlin | 0.92                   | 50            |
| Personal development coaching Berlin    | 0.92                   | 50            |
| ...                                     | ...                    | ...           |

This filtering process helped identify 53 high-performing keywords, or best 20%. Examples of the best keywords based on the analysis include: “Holistic health coaching services”, “Life coach in Berlin”, “Performance coaching services in Berlin”, “Personal development coaching Berlin”. These 53 keywords will be addressed as **Subset 1**.

Based on the data, for this particular case, four distinct keyword clusters were identified, where keywords exhibited the greatest share of relevant domains per keyword. These clusters provide a structured categorization of the keywords, facilitating focused content generation efforts. The identified clusters and their corresponding relevance shares are presented in the table below:

Table 4: Keyword Clusters and the Shares of Relevant Domains

| Keyword Cluster                    | Share    | Nr Keywords | Weighted Value |
|------------------------------------|----------|-------------|----------------|
| Mindfulness and Meditation         | 0.224490 | 49          | 11.000010      |
| Career and Personal Development    | 0.255814 | 43          | 11.000002      |
| Life Purpose and Work-Life Balance | 0.375000 | 24          | 9.000000       |
| Strategy and Vision                | 0.250000 | 36          | 9.000000       |
| Communication and Empowerment      | 0.250000 | 28          | 7.000000       |
| Personal Growth                    | 0.307692 | 13          | 3.999996       |
| Wellness and Self-Care             | 0.090909 | 22          | 1.999998       |
| Goal Setting and Focus             | 0.000000 | 39          | 0.000000       |
| Human Potential and Development    | 0.000000 | 11          | 0.000000       |

This table encapsulates the performance of the keyword clusters, categorized by their relevance share. Clusters such as “Mindfulness and Meditation”, “Career and Personal Development”, “Life Purpose and Work-Life Balance”, and “Strategy and Vision”, and their corresponding keywords emerged as the most relevant categories, offering the highest share of keywords suitable for in-depth analysis and subsequent content generation. These clusters indicate the primary areas of interest and search intent within the coaching industry, providing valuable insights for SME content strategy development.

#### 4.2 Heterogeneity of the Domain Structure per Keyword

The heterogeneity of the domain structure provides insights into keyword effectiveness and the competitive landscape. Calculating the share of unique domains for each keyword indicates a more diverse and competitive search landscape. Keywords with high shares of unique domains are more diverse and competitive, while those with lower shares suggest less diversity and are dominated by fewer domains. Based on this analysis, a group of keywords were identified as clear winners.

Table 5: Keywords with Highest Shares of Unique Domains

| Keyword                         | Total Entries | Nr Domains | Share Unique Domains |
|---------------------------------|---------------|------------|----------------------|
| Accountability                  | 50            | 50         | 1.00                 |
| Growth and development coaching | 50            | 50         | 1.00                 |
| Personal Effectiveness Coach    | 50            | 50         | 1.00                 |
| Holistic Wellness Coach         | 50            | 50         | 1.00                 |
| ...                             | ...           | ...        | ...                  |
| Mental Health                   | 50            | 35         | 0.70                 |
| Balance                         | 50            | 32         | 0.64                 |
| Meditation                      | 50            | 32         | 0.64                 |
| Coach                           | 50            | 30         | 0.60                 |

This table demonstrates keywords such as “Accountability”, “Growth and development coaching”, and “Personal Effectiveness Coach” having high shares of unique domains in their search results indicating significant diversity and competitiveness. Conversely, keywords like “Coach”, “Mental Health”, and “Meditation” show dominance by fewer domains, suggesting less competitive landscapes. The top 20% of these keywords were selected for further analysis to determine their most relevant results. These keywords will be addresses as **Subset 2**.

### 4.3 Defining the Intersection of Subset 1 and Subset 2

By intersecting Subset 1 (high relevance share keywords) and Subset 2 (high heterogeneity keywords), a refined list of highly effective keywords was generated. These keywords balance the need for diversity in search results and relevance to the business context. This dual approach ensures that the keywords not only cover a wide range of search domains but also align closely with the industry’s most pertinent search terms.

The final list consists of 20 keywords deemed most suitable based on the analysis. These keywords are listed along with their respective keyword clusters, which further categorize them according to thematic focus areas.

Table 6: Final List of Keywords and Corresponding Keyword Clusters

| Keyword Cluster                    | Keyword   |
|------------------------------------|---|
| Career and Personal Development    | Online personal development coach                 |
| Career and Personal Development    | Relationship coach Berlin                         |
| Career and Personal Development    | Relationship Coaching Berlin                      |
| Career and Personal Development    | Coaching for young professionals                  |
| Communication and Empowerment      | Stress Management Coaching                        |
| Communication and Empowerment      | Empowerment Coaching                              |
| Communication and Empowerment      | Personal Effectiveness Coach                      |
| Life Purpose and Work-Life Balance | German-speaking life coach                        |
| Life Purpose and Work-Life Balance | Life and Performance Coaching Services            |
| Mindfulness and Meditation         | Health and lifestyle coach for busy professionals |
| Mindfulness and Meditation         | Mindful Eating Coach                              |
| Mindfulness and Meditation         | Mindful Living Practices Coach                    |
| Mindfulness and Meditation         | Holistic lifestyle coaching                       |
| Mindfulness and Meditation         | Healthy Habits and Nutrition Coach                |
| Mindfulness and Meditation         | Mindfulness Practices Coach Berlin                |
| Personal Growth                    | Professional growth coaching in Berlin            |
| Strategy and Vision                | Growth and Development Coaching                   |
| Strategy and Vision                | Motivation and Accountability Coaching            |
| Strategy and Vision                | Empowering Mindset Coach Berlin                   |

| Keyword Cluster     | Keyword                      |
|---------------------|------------------------------|
| Strategy and Vision | Values-Based Coaching Berlin |

This table indicates that keywords have been effectively grouped according to their respective thematic clusters, supporting targeted content generation efforts. The final list entails highly relevant and diverse keywords critical for enhancing the online presence and competitive advantage of SMEs within the coaching industry.

Together, these combined efforts of sorting by relevance, evaluating domain heterogeneity, and selecting the top keywords form a rigorous analysis framework. This framework ensures that the chosen keywords are well-aligned with the industry’s needs and competitive landscape, providing maximum engagement potential for SMEs.

#### 4.4 Defining Top Competitors

Identifying the top competitors is yet another critical step in understanding the competitive landscape within the coaching industry. The analysis is conducted by examining the frequency with which domains appear across various high-relevance keywords. This section will present two tables: one offering an overview of domain frequencies and the other providing a more detailed examination of the top competitors.

##### Importance of Identifying Top Competitors

Understanding top competitors and their online strategies is crucial for several reasons:

- **Strategic Insights:** Identifying which domains frequently appear in top search results provides insights into successful SEO and content strategies. It highlights the approaches that allow these domains to maintain high visibility.
- **Benchmarking:** By examining the business descriptions and comparing them against the study’s base website, businesses can benchmark their own strategies against top performers. This enables the identification of gaps, potential improvements, and innovative practices.
- **Targeted Keyword Analysis:** This detailed analysis allows for a granular look into how competitors achieve their rankings, including the specific keywords and clusters they dominate. Understanding this can inform more precise and effective keyword strategies for SMEs aiming to enhance their digital presence.
- **Competitive Positioning:** In a crowded digital marketplace, knowing the key players and their strengths helps in better positioning one’s own offerings. This strategic positioning can lead to more effective marketing campaigns, higher user engagement, and improved conversion rates.

### Overview of Relevant Domains by Frequency

The first table offers an aggregated view of all relevant domains sorted by the frequency of their appearances across high-relevance keywords. This frequency metric indicates how often a domain appears in the top search results, reflecting its prominence and competitive positioning in Google Search.

Table 7: Overview of Relevant Domains by Frequency Measured

| Domain                 | Frequency |
|------------------------|-----------|
| coachfoundation.com    | 83        |
| positivepsychology.com | 58        |
| luisazhou.com          | 49        |
| katarina-stoltz.com    | 44        |
| ...                    | ...       |

### Granular Look into Competitors' Achievements

The second table delves deeper into the aggregated data, not only showing domain frequency but also providing the associated keyword, keyword cluster name, business description, and the distance to the base website specified (theklasalmon.com). This comprehensive view allows for a nuanced understanding of how competitors promote their businesses online, specifically in Google Search results.

This detailed table can be found in the supplementary file provided with this document.

By combining a high-level overview with detailed insights, the analysis offers a comprehensive understanding of the competitive dynamics within the coaching industry. This dual-level approach ensures that businesses can adopt both broad strategies and detailed tactics to enhance their online competitiveness and achieve greater visibility.

## 4.5 Analyzing Keywords by Top Competitors

The analysis of keywords by top competitors aims to understand whether well-ranked domains focus on semantically similar keywords within specific clusters or adopt a broader mix across different keyword clusters. This section explores the patterns and strategies of top competitors' keyword usage, analyzing the pros and cons of both approaches, and underscores the importance of these insights.

**4.5.1 Competitor Definition and its Characteristics** In the context of this study, "competitors" are defined as the domains that frequently appear in the top search results for high-relevance keywords related to the coaching industry. These domains are identified based on their presence in Google Search results for keywords within specific semantic clusters and are further evaluated for their relevance and competitiveness.

### **Characteristics of Competitors:**

- **Prominence in Search Results:** Competitors frequently appear in top positions for keywords, indicating strong SEO practices and authoritative content.
- **Relevant Content:** Competitors provide content that is closely aligned with the search intent of the high-relevance keywords, ensuring high user engagement and satisfaction.
- **Coverage Across Clusters:** Competitors may either concentrate on semantically similar keywords within a specific cluster or adopt a broader mix across different clusters, reflecting diverse keyword strategies.

Based on the characteristics and statistical analysis of the data for further analysis, only domains with more than four occurrences in the results were considered. Overall, 279 domains were analyzed.

**4.5.2 Importance of Analyzing Competitors' Keyword Strategies** Understanding whether top competitors concentrate on semantically similar keywords or use a broader mix is crucial for several reasons:

- **Strategic Benchmarking:** By analyzing competitors' keyword strategies, SMEs can benchmark their SEO strategies against industry leaders. This helps identify best practices and areas for improvement
- **Informed Decision-Making:** Insights into the effectiveness of different keyword strategies enable SMEs to make informed decisions about their own keyword focus. This can lead to more targeted and effective content marketing efforts
- **Competitive Advantage:** Knowing the strengths and weaknesses of top competitors' keyword strategies allows SMEs to exploit gaps in the market. By adopting a balanced and well-informed approach, businesses can enhance their competitive positioning
- **Resource Optimization:** Understanding the pros and cons of high concentration vs. broader mix strategies helps in optimizing resource allocation. Businesses can prioritize efforts that align with their objectives, whether it's building deep authority or broadening reach

### **High Keyword Concentration vs. Broader Mix:**

High keyword concentration within specific clusters is associated with the following advantages: Authority Building, Relevance and Depth, SEO Efficiency. However, it also has disadvantages: Limited Reach, Over-Specialization.

A broader mix of keywords across different clusters offers advantages such as Diverse Audience Reach, Flexibility, and Adaptability, Enhanced Opportunities, but it may lead to disadvantages such as Diluted Authority and Complex SEO Management.

**4.5.3 Case Study Examples** To categorize the domains, following methods were taken into consideration: *Arbitrary Multiplier*, *Standard Deviation*, and *Percentile Method*. For this study, the percentile method was chosen to determine the thresholds for both unique keywords and keyword clusters. Specifically, the 70th percentile was selected as the threshold to ensure that identified domains significantly differ from the average in terms of the number of unique keywords and keyword clusters. All domains inbetween the threshold will be assigned to a group called “moderate”.

The following table provides an overview of three groups and their corresponding share of domains

Table 8: Distribution of domains based on the keyword strategy

| Keyword Strategy           | Nr Domains |
|----------------------------|------------|
| High Keyword Concentration | 34         |
| Broader Mix                | 56         |
| Moderate                   | 189        |

Overall, the analysis shows that the majority (almost 70%) of websites analyzed have no particular keyword strategy, which is based on the keyword clusters.

**Possible Explanations for the Disribution:**

- Lack of Keyword Strategy on the keyword cluster level: Many domains may not consider keyword clustering based on semantic meaning and do not strategize at this level
- Clustering Effectiveness: The clustering method used may not accurately reflect the true semantic relationships between keywords, suggesting a need for adjustments in methodology

**High Concentration Example:**

Table 9: Domains implementing ‘high keyword concentration’ strategy

| Domain                     | Nr Keywords | Nr Keyword Clusters | Frequency |
|----------------------------|-------------|---------------------|-----------|
| morefromlifecoaching.com   | 15          | 4                   | 16        |
| weljii.com                 | 14          | 4                   | 14        |
| thinkcoachingacademy.co.za | 12          | 4                   | 13        |
| healthcoachinstitute.com   | 12          | 3                   | 12        |

**Broader Mix Example:**

Table 10: Domains with ‘broader mix’ keyword strategy

| Domain                      | Nr Keywords | Nr Keyword Clusters | Frequency |
|-----------------------------|-------------|---------------------|-----------|
| coach-julia-kuehling.com    | 7           | 4                   | 10        |
| lifecoachhub.com            | 9           | 5                   | 9         |
| lifecoachtraining.co        | 9           | 5                   | 9         |
| thecoachingtoolscompany.com | 8           | 5                   | 8         |

**Moderate Example:**

Table 11: Domains with ‘moderate mix’ keyword strategy

| Domain                      | Nr Keywords | Nr Keyword Clusters | Frequency |
|-----------------------------|-------------|---------------------|-----------|
| humanpotentialinstitute.com | 6           | 1                   | 21        |
| elianagoldsteincoaching.com | 2           | 1                   | 10        |
| theklasalmon.com            | 7           | 2                   | 8         |
| da-coaching.de              | 7           | 1                   | 8         |

## 5. Results and Discussion

The analysis identified high-performing keywords and keyword clusters, revealing significant insights into the competitive landscape within the coaching industry.

The steps taken for identifying these keywords and keyword clusters can be implemented as an automated flow into the AI-based textual content generation process.

Information provided in the Analysis section can and should be used by SMEs while creating textual content for their websites.

### 5.1 Implementation of the Results

The identified keywords and keyword clusters can serve as a basis for reworking content on theklasalmon.com. The keywords should be divided into primary keywords (those most important and relevant for the business field) and secondary keywords (those that complement the primary ones).

All aspects considered, it is hereafter recommended to go with *High Keyword Concentration*. Thereafter, max. of two Keyword Clusters should be considered.

Based on the keyword clusters, a set of further long-tail keywords can be generated if needed using an LLM. The keyword clusters will also serve as a source-of-truth each time the business owner wants to create new content, such as a subpage with new information or a blog post.



Reworking the website’s content should be done in an interactive session with an LLM, providing it with a guide to SEO content, as well as the keyword and keyword clusters, and the current subpage textual content.

After a period of time, the performance of the website will be measured once more to evaluate the effectiveness of the implemented strategies.

## 5.2 Limitations and Future Research

**5.2.1 Limitations of the Study** Although the dataset generated for the analysis contained over 250 keywords based on a specific business field, the list could be larger.

Another limitation is that the textual content of the websites scraped was not sufficient in terms of both quantity and quality for the clustering to yield optimal results. This reflects the inherently messy nature of the Internet, which presents challenges for such analyses.

**5.2.2 Future Research** Future research should consider expanding the scope to include multiple SME sectors to validate and generalize the findings, subsequently optimizing the automated method of keyword and keyword cluster analysis. Additionally, investigating the long-term impact of diversified keyword and keyword cluster strategies on SEO performance can provide deeper insights.

### Blog Content Analysis

Comprehensive analysis of blog metrics, including comments, shares, likes, and other engagement indicators, can highlight the type of content that resonates most with users. Evaluating the frequency of blog posts and their correlation with user engagement and SEO performance can guide content planning and publishing strategies.

### Social Media Channels Analysis

Analyzing the social media channels utilized by top competitors can offer insights into the most effective platforms for specific keywords and keyword clusters. Studying engagement metrics on social media (likes, shares, comments) can help identify the types of content that generate the most interest and interaction from the target audience.

## 6. Conclusion and Recommendations

### 6.1 Final Thoughts

Although textual content is only one of the factors defining a website’s SEO performance, it is nevertheless very important and time-consuming. The task of textual content generation cannot, therefore, be underestimated.

The latest advancements in LLMs provide SMEs with an invaluable source of information regarding their business field and their competitors. LLMs can also

be seen as perfect companions in producing business-accurate, customized, and customer-appealing textual content.

Automated keyword and keyword cluster definitions can support SMEs in the arduous task of SEO optimization, a task for which they are often unprepared.

## 6.2 Recommendations

# 7 Appendices

## 7.1 Raw Data

The dataset can be found in `results_2024-08-12_googlesearchstudy.xlsx`.

### Worksheet Overview and Description:

| Worksheet       | Description   |
|-----------------|---|
| main            | Main table with Google Search results                           |
| keywords        | Table with keywords and the corresponding keyword clusters      |
| keywords_chosen | Table with keywords and keyword clusters identified as best fit |
| competitors     | Table showcasing competitors and their keyword strategies       |

## 7.2 Additional Information

Initial IONOS Technical Talks Erfurt: Watch on YouTube

Link to the IONOS Blog: Vertrauen und Authentizität: Die Rolle von LLMs in der Content-Erstellung