Final Report: Keyword Extraction and SEO Analysis

Objective

To extract, analyze, and categorize high-quality SEO keywords from the blog post:

URL: https://www.talentica.com/blogs/custom-saas-development/

Goal: Identify and group actionable short-tail and long-tail keyword phrases for content planning, SEO optimization, and market insights.

Tools & Technologies Used

Selenium - Render and scrape dynamic content from web page

BeautifulSoup - Parse and extract HTML elements

RAKE + NLTK - AI-based keyword extraction from body content

PyTrends - Retrieve average Google Trends search volume

Pandas - Data processing and structuring into Excel

OpenPyXL - Export styled .xlsx file

Data Extracted from the Web Page

Title - Identifying page topic and keyword usage

Meta Description - SEO meta content

H1-H3 Tags - Content structure and topic hierarchy

Body Content - Extract long-tail keyword phrases

Blog Tags - Semantic categorization

Internal Links - Page architecture and SEO interlinking

NLP-Based Keyword Extraction

- Used RAKE algorithm to extract 2-4 word phrases

- Filtered only relevant long-tail and mid-tail terms

- Removed duplicates, punctuation, and stopwords

Keyword Analysis

Final Report: Keyword Extraction and SEO Analysis

When Google Trends had no data, we:

- Simulated search volume using keyword frequency on the page

- Simulated CPC and SEO difficulty for prioritization

Insights

- The page uses many long-tail phrases, ideal for niche targeting.

- Most valuable terms relate to:

  custom software

  SaaS architecture

  startup product development

- Several keyword phrases lacked trend data indicating low competition opportunities.

Conclusion

This task successfully delivered a scalable keyword research pipeline that combines:

- Dynamic web scraping

- NLP-based phrase extraction

- SEO trend approximation

- Automated report creation

This framework can now be reused across multiple blog posts or URLs for competitive analysis, SEO strategy, and content optimization.

Recommendations (Next Steps)

- Apply to more blog pages in the same domain

- Integrate real CPC/volume APIs (e.g., Google Ads, SEMrush)

- Push reports to Google Sheets or send via email weekly

- Use results to optimize existing content or write new posts