Final Report: Automated Keyword Intelligence

Objective

To perform automated keyword research and SEO analysis by scraping content from four leading web development service pages using Python. The goal was to extract relevant long-tail keywords using NLP, analyze their search performance using Google Trends (via PyTrends), and generate a structured report for SEO planning.

Target URLs

1. https://radixweb.com/services/web-development
2. https://www.netguru.com/services/web-development
3. https://www.scnsoft.com/web-development
4. https://www.appnovation.com/services/web-development

Tools & Technologies Used

- Web Scraping: Selenium, BeautifulSoup, requests
- HTML Parsing: lxml
- NLP Keyword Extraction: nltk (tokenization, n-grams, stopword removal)
- Keyword Analysis: PyTrends (Google Trends Python API)
- Report Generation: pandas, openpyxl

Scraped Elements per URL

- <title> tag: Identify primary page intent and keywords
- Meta Description: Determine how the page markets itself
- H1-H3 Tags: Structure and hierarchy of SEO keywords
- Body Text (<p>): Extract keyword-rich content
- Blog/Internal Links: Understand topical relevance & linking

Keyword Extraction Strategy

- NLP Preprocessing: Tokenization, lowercase, stopword and punctuation removal
- Phrase Length: Extracted 2 to 4-word combinations (bigrams to quadgrams)

Final Report: Automated Keyword Intelligence

- De-duplication: Filtered out repeated keywords

Keyword Performance Analysis

- Volume: Mean search interest (simulated via Google Trends)
- Difficulty: Simulated using volume-based estimation
- CPC: Estimated based on relative interest
- Type: Head term (2 words or less) or Long-tail keyword (3-4 words)

Generated Report Columns

Keyword | Source | Volume | Difficulty | CPC | Type

Key Insights

- Extracted ~200 unique 2-4 word keyword phrases
- Most high-volume keywords are 2-word head terms
- Long-tail phrases show greater semantic richness
- Frequent phrases: "web development services", "custom web apps", etc.
- Good alignment of meta tags and body content for SEO

Conclusion

This project successfully automates:

1. Scraping SEO-relevant content
2. Extracting multi-word keyword phrases using NLP
3. Analyzing keyword performance with Google Trends
4. Generating a shareable Excel SEO report

The resulting Excel file 'web_dev_keywords.xlsx' is ready for SEO audits or content strategy.

This process can be automated weekly/monthly and integrated with email or Slack for regular updates.