**SMU Keyword Optimizer – Proposal & Proof of Concept**

**Prepared by:** Crystal J. Hollis  
**Date:** September 1, 2025

**Problem Statement**

The current process for applying metadata keywords to SMU’s photo and video assets is manual, inconsistent, and not optimized for SEO practices. While existing practices ensure coverage of SMU-affiliated terms, they don’t always align with best practices for search engine visibility and discoverability.

According to the Photo Metadata Best Practices document and [PhotoShelter](https://support.photoshelter.com/hc/en-us/articles/360047332274-On-Site-SEO-Images-and-IPTC-Metadata) , fields like Description (Alt Text), Title, Keywords, Persons Shown, and Locations are crucial for SEO. However, our workflows over rely on generic terms, miss opportunities for keyword variation,

At the same time, the keyword list itself is massive and constantly evolving; with the addition of SEO keywords it might be challenging to find the right terms quickly.

What is needed is a smarter keyword tool that balances automation and quality control:

* Low-to-no-cost, sharable, accessible, and copy-paste ready for FileMaker workflows.
* Combines multiple libraries (WordNet, Datamuse, SymSpell) for synonyms and variations with real-world SEO signals (Google Autocomplete, Google Trends) for what is presently used by users on search engines.
* Incorporates a locally saved, free LLM to optimize SEO suggestions (keywords, slugs, descriptions, headlines, titles).
* Provides structured IPTC copy blocks (Headline, Description, Keywords, etc.) while leaving final QA decisions to post-production editors.

This approach augments human judgement with smarter keyword suggestions, consistent formatting, and improved SEO performance.

**Proposed Solution**

The **Keyword Optimizer** is a tool designed to:

* Suggest SEO-friendly synonyms and variants using external sources (Google Trends, Google Autocomplete, Datamuse, WordNet).
* Enforce SMU’s approved taxonomy by validating against the master keyword list.
* Generate structured IPTC metadata blocks (Job Number, Neg Number, Title, Headline, Description/Alt Text, Keywords, Persons Shown, Locations, Creator, Credit) using a locally run, open-source, lightweight large language model.
* Provide easy-to-copy output for use in Adobe Bridge, Photo Mechanic, FileMaker, and PhotoShelter.

This system reduces guesswork, improves consistency, and positions SMU’s media assets for stronger SEO performance.

**Proof of Concept**

A working prototype has been developed using Gradio. It demonstrates how users can:

1. Enter job information and keywords.
2. Automatically receive structured metadata output blocks aligned with IPTC and SEO best practices.
3. Preserve keywords that do not need synonyms or variations (referred to as “locked terms”) while preparing optional SEO suggestions for editors.

Prototype Demo Link: <https://huggingface.co/spaces/crystaljhollis/smu-keyword-optimizer>

**Core Features**

1. Multiple Keyword Sources (Toggle or “Run All”):

* Datamuse API – Contextually related word suggestions.
* Google Autocomplete – Real-world search queries for SEO value.
* Google Gemma LLM – Custom prompts by feeding it context.
* Google Trends – Identify current keyword spikes.
* WordNet (offline) – Safe, local synonym generator.

Notes on the Sources:

* Datamuse API
  + A word-finding query engine for developers, to find words that match a given set of constraints and likelihood in a given context. <https://www.datamuse.com/api/>
* Google Autocomplete:
  + Google’s Query Suggest API is an unofficial endpoint: <https://growthrocks.com/glossary/googles-query-suggest-api>
    - Note, great for low volume (a few dozen requests a day); for our uses. However Google can change or depreciate it anytime.
    - Uses:
      * <https://medium.com/@datajournal/scrape-google-people-also-ask-cefb9489c647>
      * <https://dev.to/dmitryzub/scrape-google-autocomplete-suggestion-with-python-3cgb>
  + SerpAPI handles all scraping/API but have usage limits and recurring costs. Great fallback method: <https://serpapi.com/google-autocomplete-api>
  + Another fallback method: Google Programmable Search API. Official but suggestions are not as rich as Autocomplete: <https://developers.google.com/custom-search/v1/overview>
* Google gemma 3-270m open model:
  + <https://huggingface.co/google/gemma-3-270m>
* Google Trends
  + <https://developers.google.com/search/blog/2025/07/trends-api>
  + <https://trends.google.com/trends/?geo=US>
  + Uses:
    - <https://medium.com/@datajournal/how-to-scrape-google-trends-5033c610c2d5>
* WordNet with NLTK
  + Wordnet is a lexical database that can be used to find meanings of words, synonyms, antonyms, and more: <https://pythonprogramming.net/wordnet-nltk-tutorial/>

2. Fail-Safe Architecture

* Each source runs independently with its own error handling.
* If one source fails, results from the others are still returned.
* “Run All” mode merges outputs and removes duplicates.

3. IPTC Metadata Integration

* Formats output into:
  + Primary and Optional Keywords
  + SEO Variations
  + IPTC Title (8–12 words, keyword-rich)
  + Alt Text Caption (short, descriptive, accessibility-friendly)
* Copy-paste friendly for Adobe Bridge, PhotoShelter, or FileMaker.

4. Possible Export Options (We primarily would prefer copy-paste but should have an option for exports too)

* Download results as TXT or CSV.
* Batch mode (upload a CSV of primary keywords for bulk generation).

5. User Interface (Gradio Web App)

* MVP is using Gradio. Can be deployed online or saved and ran locally as an executable file.
* Intuitive UI with source toggle or “Run All” option.
* Output displayed in clean sections for quick copying.

6. Dashboard & Analytics (Streamlit Integration)

* Provides an interactive dashboard for keyword analysis, profiling, and taxonomy alignment.
* Visualizes keyword frequency, missing metadata, and coverage against the official SMU taxonomy.
* Allows users to filter by event, year, or school and export insights for reporting.
* Built in Streamlit, making it easy to run locally or host on the web as part of the broader toolset.

**Example Output**

**Job Number:** 26-999

**Neg Number:** 12345D

**Recommended Filename Slugs:** owen\_arts\_center\_opening\_ceremony\_ribboncutting, owen\_arts\_center\_audience

**Headline:** Owen Arts Center Opening Ceremony and Ribbon Cutting at SMU

**Title:** Owen Arts Center Opening Ceremony

**Description / Alt Text:** Program footage from the Owen Arts Center opening ceremony at Southern Methodist University, featuring speeches and a ribbon cutting to mark the facility’s opening for Meadows School of the Arts classes in Dallas, Texas.

**Keywords (Batch + SEO):** Owen Arts Center, Meadows School of the Arts, Southern Methodist University, SMU, opening ceremony, ribbon cutting, dedication ceremony, donor event, Dallas Texas

**Persons Shown:** President R. Gerald Turner, Brad E. Cheves, Samuel S. Holland

**Locked Terms:** Owen Arts Center, Meadows School of the Arts, Southern Methodist University, Dallas, Texas

**Current State of Keywords**

(Include charts and CSV summaries here – already drafted in earlier section.)

**Roadmap**

* **Phase 1 (Now):** Prototype UI, structured metadata outputs, validation against SMU keyword list.
* **Phase 2 (Next):** Integration of SEO keyword sources (Google Trends, Autocomplete, Datamuse, WordNet). Add keyword scoring (difficulty/volume).
* **Phase 3 (Later):** Export to CSV/IPTC templates, direct metadata injection into files, optional automation in PhotoShelter/Bridge.

**Benefits to SMU**

* **Consistency:** Enforces taxonomy across all metadata workflows.
* **Discoverability:** Surfaces richer, more diverse keywords to aid internal search.
* **SEO Value:** Aligns metadata with Google Search Central and PhotoShelter best practices.
* **Efficiency:** Reduces manual editing time, allowing staff to focus on creative and strategic tasks.

**Supporting Appendices**

* **Keyword Frequency (CSV)** – distribution of all terms.
* **Unused Taxonomy (CSV)** – missed opportunities for richer indexing.
* **Rogue Keywords (CSV)** – off-list terms that reduce consistency.

References

Data Journal. (September 30, 2024). How to Scrape Google Trends with Python. Retrieved September 1, 2025, from <https://medium.com/@datajournal/how-to-scrape-google-trends-5033c610c2d5>

Data Journal. (November 4, 2024). How to Scrape Google’s “People Also Ask” Using Python. Retrieved September 1, 2025, from <https://medium.com/@datajournal/scrape-google-people-also-ask-cefb9489c647>

Datamuse API. Retrieved September 1, 2025, from <https://www.datamuse.com/api/>

Google. Gemma-3-270m. Hugging Face. Retrieved September 1, 2025, from <https://huggingface.co/google/gemma-3-270m>

Google Programmable Search Engine. Custom Search JSON API. Retrieved September 1, 2025, from <https://developers.google.com/custom-search/v1/overview>

Google Search Central. (February 5, 2025). Search Essentials. Retrieved August 24, 2025, from <https://developers.google.com/search/docs/essentials>

Google Search Central. (July 24, 2025). Introducing the Google Trends API (alpha): a new way to access Search Trends data. Retrieved September 1, 2025*, from* [*https://developers.google.com/search/blog/2025/07/trends-api*](https://developers.google.com/search/blog/2025/07/trends-api)

Google Search Central. URL structure best practices for Google Search. Retrieved August 26, 2025, from <https://developers.google.com/search/docs/crawling-indexing/url-structure>

GrowthRocks. Google’s Query Suggest API. Retrieved September 1, 2025, from <https://growthrocks.com/glossary/googles-query-suggest-api/>

PhotoShelter Support. (April 2025). On-site SEO: Images and IPTC metadata. Retrieved August 24, 2025, from <https://support.photoshelter.com/hc/en-us/articles/360047332274-On-Site-SEO-Images-and-IPTC-Metadata>

Python Programming. Wordnet with NLTK. Retrieved September 1, 2025, from <https://pythonprogramming.net/wordnet-nltk-tutorial/>

NextLVLProgramming. (December 2, 2024). How can I use Python for SEO? [Video]. YouTube. Retrieved August 26, 2025, from <https://www.youtube.com/watch?v=tAuCsqJUvbA>

Rover. (January 13, 2025). How to create an SEO keyword generator using Python and Tkinter. [Video]. YouTube. Retrieved August 26, 2025, from <https://www.youtube.com/watch?v=kMoYvIOtyA8>

Rover. (January 13, 2025). How to create an SEO keyword generator using Python and Tkinter. IT Solutions. Retrieved August 26, 2025, from <https://rover-tech.blogspot.com/2025/01/how-to-create-seo-keyword-generator.html>

SerpApi. Google Autocomplete API. Retrieved September 1, 2025, from <https://serpapi.com/google-autocomplete-api>

SMU. Metadata Best Practices for SEO Optimization Using PhotoShelter for Brands [Internal Document]

Zub, Dmitriy. (July 9, 2021). Scrape Google Autocomplete Suggestion with Python. Retrieved September 1, 2025, from <https://dev.to/dmitryzub/scrape-google-autocomplete-suggestion-with-python-3cgb>