

---

## SMU Keyword Optimizer – Proposal & Proof of Concept

**Prepared by:** Crystal J. Hollis

**Date:** August 31, 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. While internal tagging habits are strong, the [keyword audit](#) revealed over-reliance on generic terms like *Posed*, *Events*, *Groups*, and *Portraits*, which help with internal filtering but limit public discoverability.

Inconsistencies such as mismatched parent-child terms (*Signs* vs. *Signage*), inclusion of technical workflow terms (e.g., *B-roll*, *Final*, *Backdrops*), and use of street addresses instead of venue names further reduce SEO value and create inefficiencies. These issues result in missed opportunities to surface content in PhotoShelter and Google, and contribute to hundreds of staff hours spent annually on keyword reviews and corrections.

What is needed is a smarter keyword tool that balances automation with editorial control. The tool should be low-cost, easy to share, and copy-paste ready for FileMaker workflows, while generating structured IPTC metadata blocks (Headline, Description, Keywords, Persons Shown, Locations, etc.) that align with SEO and accessibility best practices. This approach will streamline QA, improve discoverability, and strengthen SMU's online presence.

---

### 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. The Keyword Optimizer follows SMU's *Photo Metadata Best Practices* [internal document for SEO], as well as guidelines from Google and PhotoShelter on target audience, required IPTC for SEO, keyword and filenames (see Appendix for more details).

### 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>

By combining controlled taxonomy enforcement with real-world SEO signals (Google Autocomplete, Google Trends) and a locally saved LLM, the Keyword Optimizer can suggest SEO-rich variations and enforce consistency.

## Core Features

### 1. Saved Locally

The Keyword Optimizer runs directly on SMU machines, giving staff full control without relying on third-party platforms that may change or disappear. Local deployment protects sensitive content, works even with limited internet access, and avoids recurring licensing costs. Updates can be shared easily across team members, and the tool remains copy-paste ready for FileMaker, Photo Mechanic, and Adobe Bridge workflows.

### 2. Multiple Keyword Sources (Toggle or “Run All”)

The tool draws from both controlled vocabularies and real-world SEO data, ensuring metadata is consistent internally and relevant externally:

- *Datamuse API*: expands coverage with contextually related terms.
- *Google Autocomplete*: reflects live search behavior and trending phrases.
- *Google Gemma LLM (local)*: generates on-brand slugs, titles, and descriptions with minimal staff effort.
- *Google Trends*: highlights seasonal and trending topics to prioritize.
- *WordNet (offline)*: provides safe, academic-quality synonyms without internet access.  
(See Appendix for more details)

### 3. Fail-Safe Architecture

Staff can toggle sources individually or combine them. If one service fails or becomes costly, others still work—broadening keyword diversity, reducing risk, and strengthening discoverability.

### 4. IPTC Metadata Integration

Outputs are formatted into clean, copy-ready fields, including:

- Batch and Optional Keywords
- SEO Variations
- IPTC Titles and Headlines (8–12 words, keyword-rich)
- Description / Alt Text Caption (short, descriptive, accessibility-friendly)

### 5. User Interface (Gradio Web App)

The prototype uses Gradio for a simple, intuitive interface that can run locally or online. Users can toggle keyword sources or “Run All,” with results displayed in clear sections for easy copying.

### 6. Export Options (Future Versions)

While the primary workflow emphasizes quick copy-and-paste, future versions could allow:

- Downloads as TXT or CSV for archiving or sharing.
- Batch mode to process large keyword sets at once, saving hours on high-volume projects.  
These features would help scale from one-off edits to enterprise-level metadata management.

### 7. Dashboard & Analytics (Future Versions)

A planned Streamlit dashboard could provide interactive analysis and reporting:

- Visualize keyword frequency, gaps, and alignment with the SMU taxonomy.
- Filter by event, year, or school and export insights.
- Track trends and estimate SEO impact to align strategy with audience search behavior.  
This would give SMU both oversight and measurable ROI on metadata practices.

## Example Output

**Job Number:** 26-999

**Neg Number:** 12345D

**Recommended Filename Slugs:**

smu\_football\_vs\_east\_texas\_am\_game  
smu\_mustangs\_football\_ford\_stadium  
smu\_boulevarding\_fan\_experience  
smu\_mustang\_band\_cheerleaders\_spirit  
smu\_football\_coaches\_players\_action

**Headline:** SMU Football vs East Texas A&M – Game Day Highlights at Ford Stadium

**Title:** SMU Mustangs Football vs East Texas A&M

**Description / Alt Text:** Game day at Southern Methodist University featuring SMU Mustangs football versus East Texas A&M. Scenes include boulevarding traditions, spirited fans, Peruna the pony, cheerleaders, the Mustang Band, and on-field action at Gerald J. Ford Stadium in Dallas, Texas.

**Keywords:**

**(Batch + SEO):** SMU Football, SMU Mustangs, East Texas A&M, College Football, NCAA Football, Atlantic Coast Conference, ACC Football, Game Day, Gerald J. Ford Stadium, Dallas Sports, Dallas College Football, Alumni Events, Alumni Weekend, SMU Athletics, SMU Game Day, Welcoming Community, University Football, Texas College Football

**(Optional + SEO):** Boulevarding, Tailgating, Game Action, Peruna Pony, SMU Cheerleaders, Spirit Squad, Pom Squad, Mustang Band, Student Athletes, Football Coaches, Fan Experience, SMU Fans, College Fans, Game Day Signs, Game Day Signage, Pony Ears, Youth Fans, Families, Family Weekend, Alumni Families, Group Photos, Posed Groups, Animals, Mascots, College Traditions, School Spirit

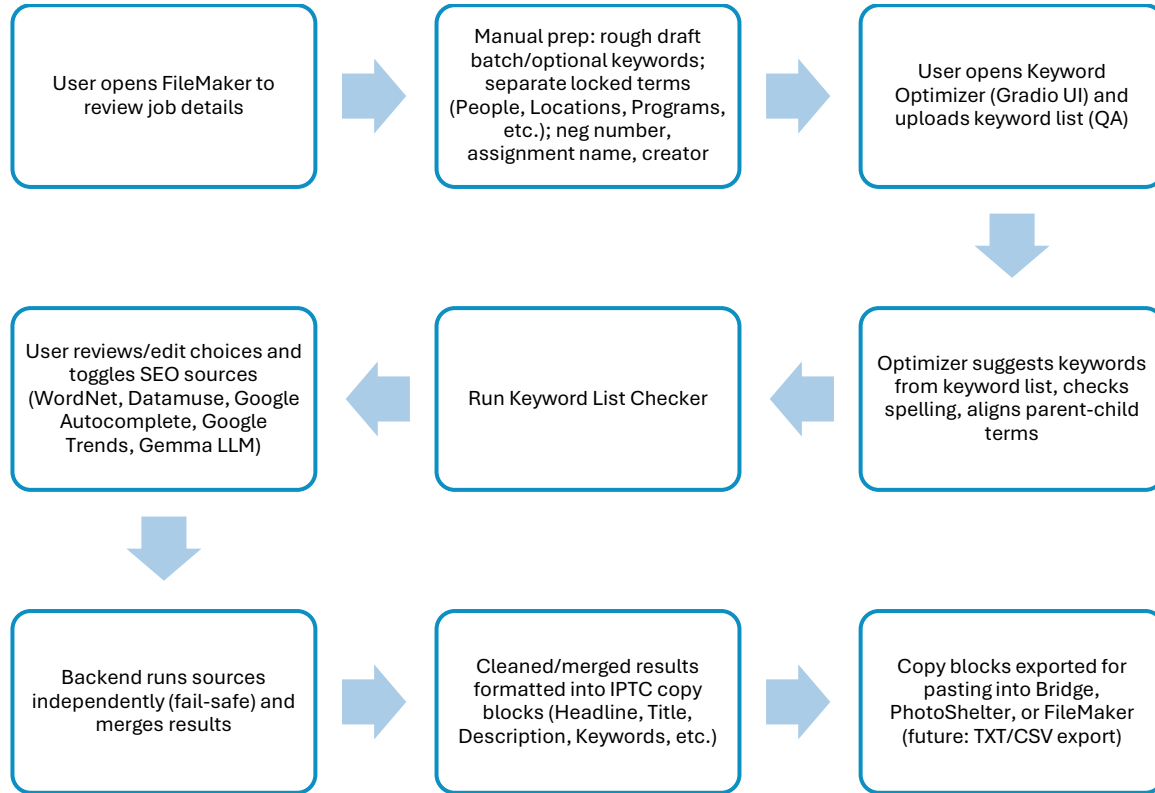
**Locked Terms (skips synonyms and variations):** Southern Methodist University, SMU, Mustangs, Athletics, Football

**Persons Shown (Locked):** President Jay C. Hartzell, Provost Rachel Mersey Davis, Kenekukwu K.C. Mmeje, David B. Miller

**Role Variants:** University President, Provost, Vice President of Student Affairs, Trustees, Donors

**Location (Locked):** Dallas Hall Lawn, Bishop Boulevard, Gerald J. Ford Stadium, Dallas, Texas

## Diagram



## Roadmap: Timeline and Estimated Budget (Time Allocation)

### Budget & Time Allocation for Magna Media Arts LLC Staff:

- This project is being developed within the existing freelance support structure at **\$20/hr**, with hours drawn from current allocations so there is no additional cost to SMU.
- Crystal Hollis (Post-Production Specialist, 15 hrs/week):** As part of her regular freelance support, Crystal is dedicating about 5–7 of her 15 weekly hours to Keyword Optimizer development, with the remainder focused on production deliverables. This work applies Crystal's newly earned Python Developer skills to SMU's needs.
- Sean Hollis (Assistant, 15 hrs/week):** Sean, a Dallas College student pursuing a Bachelor of Applied Technology in Software Engineering, is dedicating 0–5 of his 15 weekly hours to development, testing, and data entry, with the remainder focused on production deliverables.
- By treating the Keyword Optimizer as a pilot R&D effort, SMU gains innovation without added cost, while reducing manual keywording over time and freeing staff hours for higher-priority work.

### Estimated Timeline (September 2025 – May 2026)

- Phase 1 (September–October 2025):** Continue to manually improve and standardize metadata for SEO. At the same time, finalize the prototype UI in Gradio and validate against SMU's keyword list.
- Phase 2 (November 2025–January 2026):** Integrate SEO keyword sources (Google Autocomplete, Trends, Datamuse, WordNet) alongside the locally run Gemma LLM. Add keyword scoring (difficulty/volume) to help editors make informed choices.
- Phase 3 (February–May 2026):** Develop export options (TXT/CSV), batch processing, and explore direct metadata automation with PhotoShelter and Adobe Bridge.

---

## Conclusion

The Keyword Optimizer represents a practical, low-cost way for SMU to modernize its metadata workflows while building toward long-term SEO gains. By balancing automation with editorial control, the tool improves consistency, surfaces richer keywords that match real-world search behavior, and reduces the hours currently spent on manual cleanup. Developed entirely within SMU's existing freelance support structure, it delivers innovation without additional cost and creates a foundation that can scale with future needs. Most importantly, it positions SMU's photo and video assets to be more visible, accessible, and impactful—amplifying the reach of the university's brand while freeing staff time for higher-value creative and strategic work.

---

## Appendix

### Audience & SEO Guidance - Context

#### Target Audience for Photos / Videos

([Photoshelter for Photographers](#)):

- What would you like people to search for in Google (or other search engines) to land on your website's homepage?
- Be sure to consider your local market primarily.
- Who are your competitors in this space? If you run a Google search for the keywords you want to use, how are those competitors addressing their target audience by the way they write?
- What is the budget of your ideal client? Be sure to do your research and consider target demographics.
- What other sites/forums/modes of social media are your clients spending their time on?
- Define ideal audience in 1-2 sentences.

([Google Search Central](#)):

- Write clear, original, helpful, people-first content
- Don't chase every synonym- use natural phrasing. Google can match related queries without exact keywords
- Anticipate searcher language i.e. expert vs. beginner variants like "charcuterie" vs "cheese board"

#### Required IPTC fields ([Photoshelter for Photographers](#)):

- **Image Credits for search results:** IPTC Creator, Credit, and Copyright Fields
  - Note: Creator is "Creator/Photographer" in Photo Mechanic, "Creator" in Adobe Bridge, and "Author" in PhotoShelter
  - **This will not be generated in the optimizer but may be fed to Gemma LLM. Editors must use photo mechanic/adobe bridge templates**
- **Headline:** Brief synopsis of the caption. PhotoShelter uses IPTC Headline as the page's HTML title and sitemap.
- **Title:** shorthand reference, what you would call the file instead of the file name. For our purposes it will be the assignment/event name
  - Note: Google recommends page titles of 70 characters or less.
- **Description:** Describe contents of photo (who, what, when, where, why). PhotoShelter uses this as alt text (accessibility) and indexed by site search.
- **Alt Text:** Separate field on IPTC, but for our purposes it will be copied from Description.
- **Sublocation:** Address of a place if off-campus, building name if on-campus.
  - Note: search engines indexes City, State, and Country but not sublocation. Make sure to include location in the description.
- **Credit:** person or organization to receive credit. i.e. Southern Methodist University / Guy Rogers III
  - **Note:** Credit is "Credit" in Photo Mechanic, "Credit Line" in Adobe Bridge, and "Credit/Provider" on PhotoShelter

- **Copyright URL:** link referencing copyright ownership. Can be <https://www.photoshelter.com/support/license> or SMU's own published copyright URL.
  - Note: Copyright URL is "Copyright URL" in Photo Mechanic, "Copyright Info URL" in Adobe Bridge, and "Web Statement URL" in PhotoShelter.
  - **This will not be generated in the optimizer. Editors must use photo mechanic/adobe bridge templates**
- **Licensor URL:** link to a page on how to acquire a license. For our purposes it could be a contact info to Photo/Video dept.
  - Note: This field is usually a URL field under "Licensor" in Photo Mechanic and Adobe Bridge. PhotoShelter refers to this as "Licensor URL".
  - **This will not be generated in the optimizer but may be fed to Gemma LLM. Editors must use photo mechanic/adobe bridge templates**

#### Keyword guidelines from [Photoshelter for Photographers](#):

- Separate keywords by a comma and a space.
- Don't treat keywords like hashtags on social media (i.e. #photooftheday2020, photooftheday, #houstonwedding, photography2020 are not useful keywords).
- Make sure keywords are specific to the content of the image. Instead of simply using "dog," try "Border Collie Dog."
- Supplement a keyword like "Grand Canyon" with a specific location or landmark that the picture includes (e.g. "Maricopa Point" or "Grand Canyon Maricopa Point").
- Quality over quantity: Include at least 5 keywords per image. Don't add keywords that are irrelevant and avoid using too many keywords that are generalized. Google and other search engine algorithms consider this practice "keyword stuffing" and they tend to penalize sites for this. Using a keyword like "best photos 2020" will not help a user find your particular image. Search engines tend to reward websites with the most relevant keywords that relate to the text content in the rest of the website.
- What will someone search for to find the photo in question? Using a keyword "photo" will not help as much as "Corporate Headshot." If a potential client wants to find a Corporate Headshot photographer in your area, what are the most important phrases they will type in? You can even try searching in Google and you will see similar searches.
- Be discerning about what you actually use as keywords from these suggestions.
- If your keywords are as specific and relevant as you can get them, there are still other factors involved in search engine rankings.

#### Filenames

- It is recommended by [PhotoShelter](#) to use helpful names instead of generic names. i.e. red-wine-vineyard.jpg instead of img001.jpg. For our purposes we should use descriptors in the file names after the catalog number and sequence number.
- It is worth noting that [Google](#) recommends using hyphens (-) instead of joining words together or using underscores (\_) to separate words in URLs. Not only does it help users and search engines better identify concepts but historically underscores are commonly used for denoting concepts that should be kept together i.e. various programming languages use it to name functions such as "format\_date".
- Currently, PhotoShelter's URL **does not** use the file name; it uses the gallery name and internal IDs (i.e. <https://photos.smu.edu/gallery-image/Athletic-Forum-Lunch-with-EliManning/G0000zitb5WfAmzY/I0000YKK1GBC85uw/C0000d8v0qGF8fRA> )
- But it is something to be aware of should there be changes or migration to a different system that uses filenames in the URL.

## 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 deprecate 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/>
-

---

## 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>
- 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>