Leveraging AI to Expand Home Electrification and Efficiency Incentive Data Access

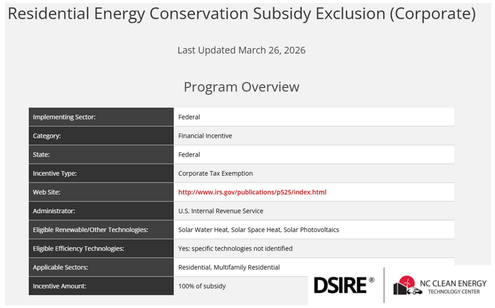
Dawson Chan[1], Sheel Gada[1], Kaleb Lamphiear[1], Giovanni Loia[1], Tommy Yue[1], Nick Montoni[2], Brian Lips[2]
[1]Department of Chemical Engineering – University of Washington; [2]NC Clean Energy Technology Center

Background

Electric utility providers and local governments offer economic incentives for renewable energy systems, energy-efficient equipment, and building improvements. The NC Clean Energy Technology Center (NCCETC) maintains the Database of State Incentives for Renewables and Efficiency (DSIRE), a comprehensive source for policies and incentives supporting the use of renewable energy. Currently, this database is manually populated and updated. Due to the volume of available incentives, it has become unfeasible to continue this method, especially for small utilities with under 30,000 customers.

The aim of our project is to use a language learning model (LLM) to create an automated web-scraping tool that finds and summarizes these incentives, returning an in-depth summary of information formatted to be entered into DSIRE.

a)



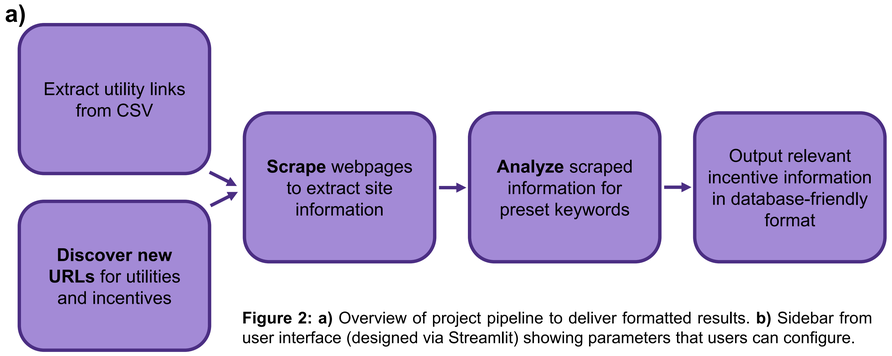
b) **Examples of keywords chosen for scraping**

Incentive Type	Derived Keywords
Rebate Program	Rebate, incentive application …
Grant Program	Grant, funding, eligibility
Solar renewable energy credit program	Solar, efficiency, program …

Figure 1: a) Example overview of an entry in DSIRE. **b)** Types of incentives as categorized within DSIRE alongside the keywords derived for scraping ang analyzing webpages.

Pipeline Overview

a)



b)

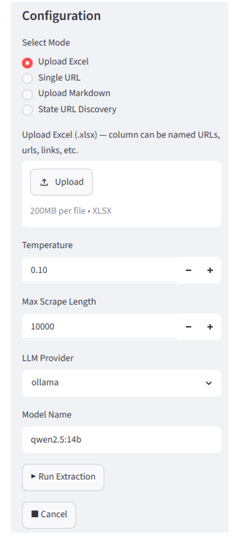


Figure 2: a) Overview of project pipeline to deliver formatted results. **b)** Sidebar from user interface (designed via Streamlit) showing parameters that users can configure.

The pipeline includes different modes based on the structure of the sites and/or stored data:

- Upload Excel — provide a .xlsx with a list of URLs.
- Single URL — attach a link or website that you wish to crawl, scrape, and analyze
- Upload Markdown — If you have pre-scraped information from a website, you can upload the file for it to be analyzed (bypasses the scraping portion)
- State URL Discovery — Choose which search engine, state, search topics, and results per search topic to find relevant URLs

Scraper

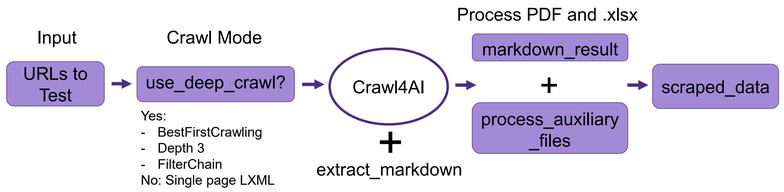


Figure 3: Scraper process flow diagram. The first portion of the pipeline was designed to gather all the relevant information found on a utility's website through a process called scraping.

- Crawling Method = BestFirstCrawlingStrategy - follows linked subpages across the site, prioritizing pages that score highest for incentive-related keywords
- max_depth - crawls up to 3 levels deep and stays within the seed domain
- FilterChain:
  - SEOFilter - helps identify pages with strong SEO characteristics (meta tags, headers, ect.)
  - DomainFilter - which domains to include or exclude

URL Discoverer

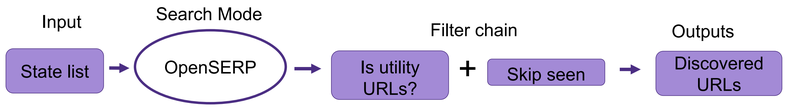
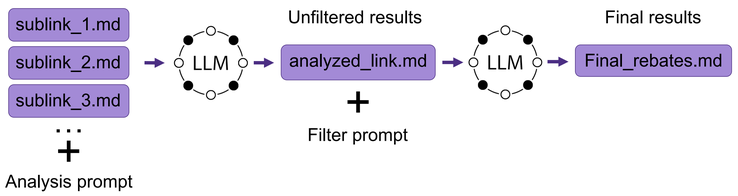


Figure 5: URL Discoverer process flow diagram. As an alternative to inputting known utility URLs, the URL discoverer was created to locate electric utilities that have not been captured in DSIRE by finding their official websites.

- Search topics - Utility ownership types (cooperative, municipal, investor-owned)
- Engine choice - Bing / DuckDuckGo tolerate automation; Google gives best results but triggers CAPTCHA
- Three-pass filter - block aggregators, keep utility domain signals, brand-name fallback on title keywords
- Merge - combines newly discovered URLs into the database with domain-level deduplication, skipping sites already in the database and backing up the old file.

Analyzer

a) Scraped markdown



b)

Model	Context window	Parameter count	Performance
Llama_3.1:8b	128k	8 billion	Good lightweight baseline, but struggles to filter
Gemma_4:31b	256k	31 billion	Conservative analysis
Gpt_5.4-pro	~1.05M	Undisclosed	Best overall model but $$$

Figure 4: a) Analyzer process flow diagram. The scraped website information is fed into a series of LLMs, with the initial analyzer designed to find the program name, URL, concrete rebates, eligibility, and any additional utility information. These results are then fed into another LLM designed to filter and discard any sections deemed irrelevant. **b)** Three different models were tested. A model complexity increased, performance improved, but cost to run increased. Testing more models is recommended.

Recommendations & Limitations

Parameters that should not be modified by user:

- **Search Depth -** Set at 3. Depth = 2 misses information; Depth = 4 is too slow
- **AI Prompts -** Both analyzer prompts have been tuned to minimize errors
- **Temperature –** Range of 0 (most rigid) to 1 (most creative). Set at 0.1 for both functions to keep responses grounded in source material

Parameters that user can modify:

- **LLM -** Large models are more reliable, but may require advanced hardware
- **LLM Time -** Maximum time model can spend per link
- **Truncation Length -** Must change to fit each model's capacity

Further investigation:

- **Paid API or Dedicated PC -** Provide the compute power externally
- **Multi-user deployment -** Deploy via Streamlit to act as accessible website

Results

a)

Community Office for Resource Efficiency (CORE)	Grants & Funding Programs	Grant	$10,000 to $50,000 per project	Non-profit organizations and government agencies in Aspen area	Contact (aspencore.org)(https://www.aspencore.org/grants) or call (970) 925-9775	
Aiken Electric Cooperative (AEC)	H2O Select PLUS2 - Water Heater	Rebate	Up to $400 rebate on H2O Select PLUS2 water heater purchases, with $50 more for Energy Star	Residential customers only, must be an AEC member	Program information and application available at https://www.aikenco-op.org/water-heater-	Residential
Aiken Electric Cooperative (AEC)	Smart Thermostat	Free Equipment	Ecobee Smart Thermostat, free to AEC members who sign up for Beat the Peak program	Residential and commercial customers only, must have smart thermostat installed	Program information available at https://www.aikenco-op.org/ecobee-smart-thermostat	Residential, Commercial
Aiken Electric Cooperative (AEC)	Community Solar	Grant	$0.05/kWh credit for residential customers participating in community solar program	Residential customers only, must be an AEC member	Program information available at https://www.aikenco-op.org/community-solar	Residential
California Alternative Rates for Energy (CARE) / Family Electric Rate Assistance	Discounts on energy bills for income qualified households	Rebate	CARE: 20-35% monthly discount on electric bills. FERA: 18% monthly discount on	Must be income-qualified. Must bea utility customer or Community	Program information available at https://www.cpuc.ca.gov/industries-and-topics/electrical-	Residential
Central Alabama Electric Cooperative	Heat Pump Rebates	Rebate	$350/ton, $250/ton, $350 per ton, $400 per ton to the homeowner,	Standard Homes: Dual Fuel systems and mini-split	For more information, contact us at 1-800-545-5735 ext. 2118	Residential, Commercial

b)

Program Name:California Alternative Rates for Energy (CARE)/ Family Electric Rate Assistance (FERA) Programs

#Program URL: https://www.cpuc.ca.gov/industries-and-topics/electrical-energy/electric-costs/care-fera-program

Program Details

- **Concrete Rebate Amounts:**
- CARE: 20-35% monthly discount on electric bills
- FERA: 18% monthly discount on electric bills

Eligibility

- **Eligibility Requirements:**
- Must be income-qualified
- Must be a utility customer or Community Choice Aggregator (CCA) customer
- Applies to electric bills in California

Utility Information

- **Utility Company Name:** California Public Utilities Commission (CPUC)
- **Utility Company Size:** Statewide

--- SOURCE: cpuc_ca_gov_2520a79d.md ---

Figure 6: Example IncentivAI output. After each scrape, final incentives are returned in both Excel **(a)** and markdown format **(b)**. Exact categories included may vary by incentive, but general fields include Program Name, Program Type, Financial Details, Eligibility, Application Process, and Sector.

To quantify our pipeline's effectiveness, we used a subset of scraped markdown files from the city of Anaheim, CA (selected for sebpage complexity and organization. 61 of the 65 definitive rebates were accurately pulled for a **success rate of 94%**. 17 loosely associated rebates which did not fit the specifications of the prompt were missed.

Acknowledgments

This work was supported by Dr. Dave Beck and Chino Ruttanasupagid.
Special thanks to our stakeholders at the NC Clean Energy Technology Center

References

1. Klesel, M., & Wittmann, H. F. (2025). Retrieval-Augmented Generation (RAG). Business & Information Systems Engineering. https://doi.org/10.1007/s12599-025-00945-3
2. Wang, S., Khramtsova, E., Zhuang, S., & Zuccon, G. (2024). FeB4RAG: Evaluating Federated Search in the Context of Retrieval Augmented Generation. Proceedings of SIGIR 2024. doi:10.1145/3626772.3657853
3. Blázquez-Ochando, M., Prieto-Gutiérrez, J. J., & Ovalle-Perandones, M. A. (2025). Prompt engineering for bibliographic web-scraping. Scientometrics, 130(7), 3433–3453. https://doi.org/10.1007/s11192-025-05372-5