Contract No:

This document was prepared in conjunction with work accomplished under Contract No. 89303321CEM000080 with the U.S. Department of Energy (DOE) Office of Environmental Management (EM).

Disclaimer:

A U.S. DEPARTMENT OF ENERGY NATIONAL LAB • SAVANNAH RIVER SITE • AIKEN, SC • USA

A Semi-Automated Approach for Curating a Glossary of Key Terms for Open-Source Data Queries

T. L. Danielson

September 2022

SRNL-STI-2022-00462, Revision 0

SRNL.DOE.GOV

DISCLAIMER

Printed in the United States of America

Prepared for
U.S. Department of Energy

SRNL-STI-2022-00462
Revision 0

Keywords: *Natural Language Processing*
Machine Learning
Nuclear Energy

Retention: *Varies*

A Semi-Automated Approach for Curating a Glossary of Key Terms for Open-Source Data Queries

T. L. Danielson

September 2022

Savannah River National Laboratory is operated by Battelle Savannah River Alliance for the U.S. Department of Energy under Contract No. 89303321CEM000080.

Savannah River National Laboratory®

REVIEWS AND APPROVALS

AUTHORS:

____

T. L. Danielson, SRNL, Environmental Sciences & Dosimetry Date

TECHNICAL REVIEW:

____

S. Branney, SRNL, Nonproliferation Policy Support Date

APPROVAL:

____

Marion Cofer, Manager Date
SRNL, Environmental Sciences & Dosimetry

PREFACE OR ACKNOWLEDGEMENTS

This work is funded by the National Nuclear Security Administration's Office of Defense Nuclear Non-Proliferation Research and Development (NA-22).

EXECUTIVE SUMMARY

In FY20, the Savannah River National Laboratory (SRNL) was funded by the National Nuclear Security Administration's Office of Defense Nuclear Non-Proliferation Research and Development (NA-22) to build a machine learning based modeling pipeline that could extract proliferation events of interest from open text-based data sources. As a test case, the research team targeted the identification/fusion of events and indicators that fissile core fabrication would be executed at the Savannah River Site prior to its official announcement in May of 2018. The demonstration prototype proved successful by applying natural language processing and graph theoretical techniques to identify contextual shifts in key words and phrases that acted as indicators that pit production would be carried out at the Savannah River Site up to two years prior to the official announcement.

Having demonstrated proof-of-concept, the team was funded in FY22 to continue development on the prototype modeling pipeline by applying it to a domain of the worldwide landscape of civil nuclear power activities. A key challenge in using open-source data is that the continuously expanding volume of content can produce significant noise from out-of-domain data. Therefore, a key step in the modeling pipeline is the acquisition of sufficient, primarily domain-specific data. A common approach to curate a sufficiently large, domain-specific dataset is to develop a "glossary of key terms" that can be used to formulate queries to data sources, generally with input from subject matter experts. Manually performing this task risks overlooking key terms that have high specificity and can result in missing data or excessive noise from more general uses of a term. Likewise, a broad topical domain such as the worldwide landscape of civil nuclear power might require a team of multiple experts to capture sufficient relevant terminology, reducing the accessibility of applying the pipeline to broad domains of interest. Therefore, to overcome these challenges, the SRNL team has developed a semi-automated approach for curating a glossary of key terms that recursively refines the vocabulary extracted from readily available, domain-specific sources. Here, the approach is outlined and the preliminary datasets that were extracted using the glossary are characterized to demonstrate the effectiveness of the approach.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SRNL	Savannah River National Laboratory
NA-22	National Nuclear Security Administration Office of Defense Nuclear Non-Proliferation Research and Development
SME	Subject Matter Expert

1.0 Introduction

In FY20, the Savannah River National Laboratory (SRNL) was funded by the National Nuclear Security Administration's Office of Defense Nuclear Non-Proliferation Research and Development (NA-22) to build a machine learning based modeling pipeline that could extract proliferation events of interest from open text-based data sources. As a test case, the research team targeted the identification/fusion of events and indicators that fissile core fabrication would be executed at the Savannah River Site prior to its official announcement in May of 2018. The initial development efforts were centered around two datasets: a decahose Twitter database (i.e., a global sampling of one in every ten Tweets) and a broad internet archive (e.g., long-form news articles). The prototype modeling pipeline was built on a foundation of natural language processing, and more specifically, time dependent word embedding models. The analysis of the temporal embedding models included a graph theoretical approach that detects contextual shifts in key words and phrases, indicating an event of interest in previous time windows. The FY20-21 prototype demonstrated success by identifying indicators that fissile core fabrication would occur nearly two years prior to the official announcement (Danielson, 2022a).

In FY22, SRNL was funded by NA-22 to continue development efforts on the modeling pipeline to detect new developments, and/or previously unknown interests, in civil nuclear power activities around the world. Development on this broader topical domain will allow testing of the modeling pipeline's capability to extract meaningful events in a worldwide data environment that incorporates potential location-dependent variations in the availability of open-source data, as well as several examples of civil nuclear reactors that are in different phases of completion (e.g., see Danielson, 2022b). A successful modeling pipeline will be capable of extracting events of interest regarding the planning, construction, operation, and shutdown of reactors worldwide, and ideally demonstrate the capability to allow inference.

A potentially limiting step in the modeling pipeline's ability to identify and extract events of interest related to civil nuclear power is the acquisition of sufficient data to train word embedding models. Furthermore, with hundreds of millions of Tweets per day, it is necessary to have a filtration step in the pipeline that reduces the overall noise and captures primarily domain-specific data. A common approach to curate a sufficiently large, domain-specific dataset is to develop a "glossary of key terms" that can be used to formulate queries to data sources, generally with input from subject matter experts (SMEs). Performed manually, SMEs may miss key terms that have high specificity, resulting in missing data or excessive noise from more general uses of the term. Likewise, a broad topical domain such as the worldwide landscape of civil nuclear power might require a team of SMEs to capture sufficient relevant terminology, reducing the accessibility of applying the pipeline to broad domains of interest. Therefore, to overcome these challenges, the SRNL team has developed a semi-automated approach for curating a glossary of key terms that iteratively refines the vocabulary extracted from readily available, domain-specific sources.

In the following sections, the algorithmic approach will be described, and the glossary of key terms will be presented. Additionally, a preliminary characterization of the datasets returned from queries will be outlined to provide an evaluation of the performance of the approach.

2.0 Algorithmic Approach

The algorithmic approach has been designed to be semi-automated, whereby n-gram keyword phrases are extracted from a user-specified text-based data source and post-processed to a significantly reduced set for fine tuning. Here, the online Information Library that is compiled by the World Nuclear Association has been used as a resource for capturing keywords that are relevant in a worldwide context related to civil nuclear power. Rather than extracting all n-grams (e.g., 1-, 2-, 3-, …, grams) and recording the frequency of occurrence, as was done in FY20-21, the current approach only extracts n-gram noun phrases from the source text. Subsequently, the noun phrases are reduced by recursively identifying n-1–grams that are substrings of longer n-grams in the terms list. This approach reduces the number of extracted noun phrases to approximately 5% of the initial set, leaving a much more manageable list of key terms for SME fine tuning.

The algorithm, as implemented in the current work, is outlined by the following steps:

1) Scrape text from the Information Library on world-nuclear.org[1]
    a. All "Country Profiles" pages[2]
    b. "Nuclear Power Reactors" page[3]
    c. "Nuclear Energy and Sustainable Development" page[4]
    d. All "Nuclear Fuel Cycle" pages[5]
    e. All "Safety and Security" pages[6]
    f. All "Current and Future Generation" pages[7]
2) Label and group noun phrases[8] from the text (77,010 noun phrases extracted)
    a. Remove punctuation and/or stop words from each n-gram's tokens
    b. Lemmatize each token in the noun phrases
3) Create a vocabulary of lemmatized[9] 1-grams from all noun phrases captured in Step 2 (22,762 total words)
4) Remove all noun phrases from Step 2 that occur only one time (20,175 n-grams remaining)
5) For all remaining n-grams greater than length 1
    a. Use fuzzy string matching[10] to match n-grams to a single 1-gram that has the highest similarity, creating a mapping of 1-grams to a list of n-grams (e.g., the word "nuclear" is mapped to 3,320 n-grams, such as "nuclear activities", "nuclear assets", "nuclear fission", etc.)
        i. Note: numerical tokens in an n-gram are kept, but ignored, in fuzzy string matching
    b. Eliminate all 1-grams that are mapped to fewer than four n-grams (10,758 total n-gram terms remaining mapped to 1,270 root one grams)
6) Recursively identify all n-1 – grams that are substrings of n-grams in the terms list from Step 5b (e.g., "advanced nuclear reactor" can be simplified to the substring "nuclear reactor")
    a. 1-grams that appear in a 2-gram are removed from the list
    b. Note: The simplification is stopped at 2-grams (2,425 n-grams remaining)
7) SME down selection, fine tuning, and addition of Boolean logic (e.g., "fission" AND "nuclear")

The SME down-selection process involves several different actions, such as adding abbreviations and acronyms, adding alternative spellings as necessary (e.g., "nuclear programme" versus "nuclear program"), and eliminating key phrases that are too broad and/or potentially out of domain. Additionally, key terms that have ambiguity (e.g., "fission", which might refer to nuclear fission or biological fission in the open source) are given additional terms that must be found in the document (i.e., Tweet or article) for the document to be returned in the query. Ultimately, the size of the final list of terms is tailored based on the cost and efficiency of submitting queries in addition to capturing the breadth of the topical domain of interest. Here, the final list contains 501 key terms, listed in Appendix A.

---

[1] https://www.world-nuclear.org/information-library.aspx
[2] https://www.world-nuclear.org/information-library/country-profiles.aspx
[3] https://www.world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-power-reactors/nuclear-power-reactors.aspx
[4] https://www.world-nuclear.org/information-library/energy-and-the-environment/nuclear-energy-and-sustainable-development.aspx
[5] https://www.world-nuclear.org/information-library/nuclear-fuel-cycle.aspx
[6] https://www.world-nuclear.org/information-library/safety-and-security.aspx
[7] https://www.world-nuclear.org/information-library/current-and-future-generation.aspx
[8] Noun phrases are grouped using the Python package spaCy
[9] Lemmatization is performed using NLTK's WordNetLemmatizer
[10] Fuzzy string matching is based on the token sort ratio as implemented in the Python package fuzzywuzzy

3.0 Preliminary Characterization of Datasets

Two data sources have been queried using the glossary of key terms that is shown in Appendix A: a decahose Twitter database and a broad internet archive (e.g., news articles, blogs, and licensed content) sourced from newsapi.ai. To evaluate the datasets, preliminary exploratory data analysis was performed prior to pushing the data through the FY20-21 prototype modeling pipeline so that the glossary of key terms could be modified if need be. Because the team already had access to a Twitter database that spanned from August of 2014 through April of 2018, this timeline was maintained for a preliminary data pull. In the case of the internet archive dataset, a query parameter was allowed for returning only English language articles. This option was used which circumvented the need to translate query terms and the returned documents that are returned to/from several languages. Contrastingly, no language filtration was performed on the Twitter dataset, though approximately twenty percent of the Tweets returned were non-English (they were captured by non-translated keywords because of the use of English "hashtags" that also appear in the glossary of terms). Given the shorter length of Tweets, the preliminary exploratory data analysis was performed by applying open-source translation tools[11] to non-English Tweets. Preliminary evaluations of translated Tweets indicate that the translations are sufficient for the word2vec-type word embeddings that will be trained.

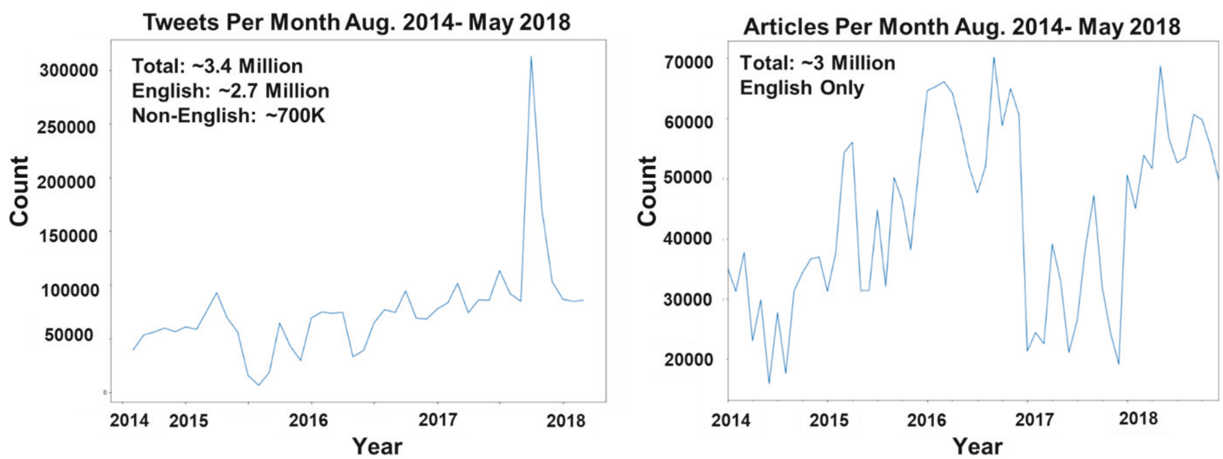

Figure 3-1. Tweets per month and articles per month from the two data sources.

The number of articles per month from the Twitter and internet archive databases is shown in Figure 3-1. Note that in late 2017, a large spike occurs in the volume of Tweets that were returned from the query. This appears to be most significantly attributed to the key term "uranium", where a significant multi-national event occurred related to uranium during that time. This example illustrates the "viral" nature of how headline events spread through social media data sources. While potentially relevant, these "viral" topics within the dataset were shown to cause a significant contextual shift across large portions of the key terms of interest in the FY20-21 efforts and therefore, identifying them in the early phases of the research can help to explore the impact on the modeling pipeline in the downstream efforts. The top 50 most frequently occurring key terms from the glossary that have returned Tweets/articles from the data queries are shown in Figure 3-2. Notably, in comparing the two lists, 25 out of 50 terms are the same. In general, the 25 terms that are only in the top 50 most frequently occurring for the Twitter dataset appear to have higher specificity than those from the internet archive dataset. This might be due to the more succinct nature of Tweets compared to longer-form articles.

[11] The Python package deep_translator was used for translation by employing the GoogleTranslator.

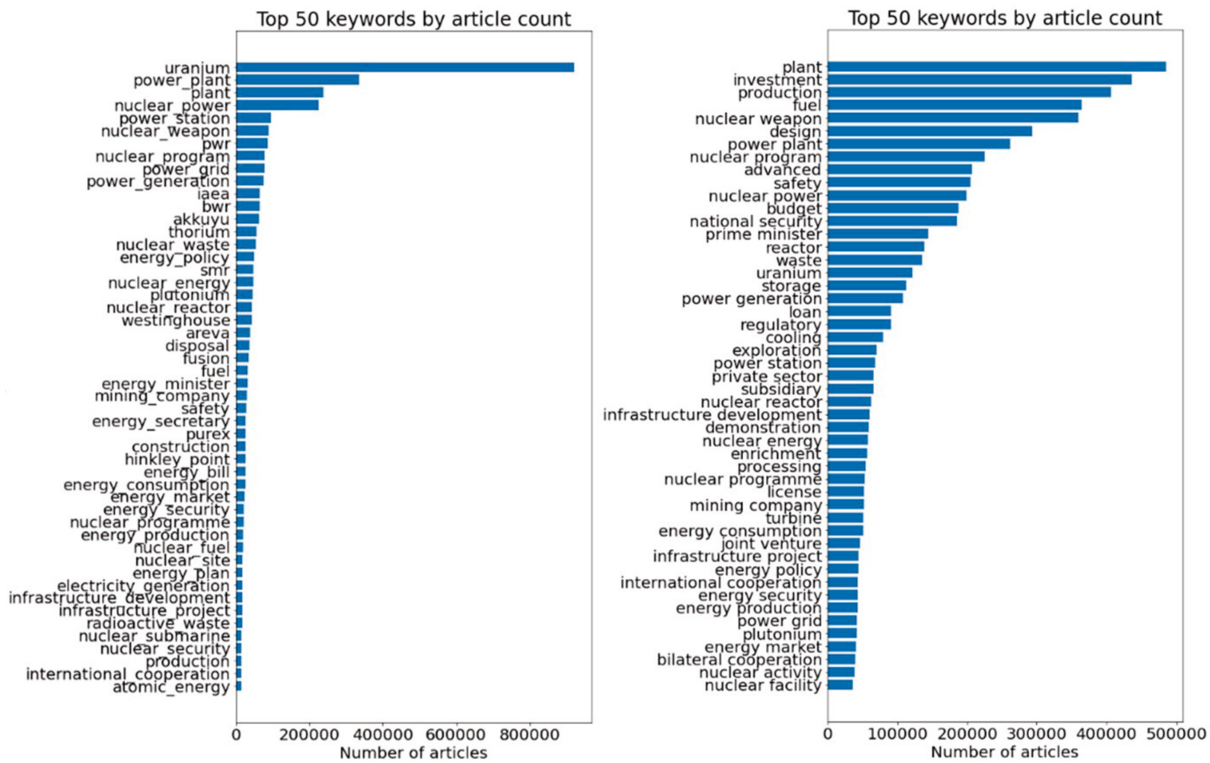

Figure 3-2. Total number of Tweets (left) and articles (right) for the top 50 keywords returned by the data queries.

Word embedding models were trained on the individual corpuses to evaluate the contextual representation of the key terms that could be obtained from the queries. Table 3-1 provides the top ten most similar (i.e., based on cosine similarity) words mapped to a subset of the key terms that are related to worldwide civil nuclear power activities, including contractors, reactors in various stages of development, and a reactor type. The lists of most similar terms show a realistic contextual representation, indicating that there is likely sufficient data for the event extraction modeling pipeline that will leverage the use of time dependent word embedding models and metrics, such as the cosine similarity over time for identifying contextual shifts. Preliminary time dependent word embedding models using 30-day, growing windows (i.e., Window 1 = 30 days of data, Window 2 = 60 days of data, etc.) were trained on the Twitter dataset to explore the top 50 most similar words over time for a selected entity, Rosatom, shown in Figure 3-3. Notably, several entities that will serve as strong case studies appear in the evolving list, indicating that events across time, and throughout the world, can be identified using the prototype modeling pipeline. Finally, Figure 3-4 shows a principal component analysis plot of word vectors representing the glossary of key terms in an embedding model that was trained on the full internet archive dataset. Notably, key terms that belong to similar sub-domains (e.g., broad nuclear reactor terms, cooperative/diplomatic terms, and energy infrastructure terms) are grouped closely together, indicating the embedding model has a strong contextual representation of civil nuclear power. Note that these are only examples of possible sub-domains that could potentially be broken into more descriptive and/or complete sub-domains as development continues.

Table 3-1. Top 10 most similar words from word embedding models trained individually on the full internet archive and Twitter datasets for select key terms. Red terms indicate that the term appears in both lists.

	rosatom		akkuyu		paks		vver		el dabaa	
	NewsAPI	*Twitter*	*NewsAPI*	*Twitter*	*NewsAPI*	*Twitter*	*NewsAPI*	*Twitter*	*NewsAPI*	*Twitter*
1	atomstroyexport	fennovoima	el-dabaa	rosatoms	belene	hungary	vver	generation_iii	kureimat	dabaa
2	technopromexport	hanhikivi	mersin	mersin	akkuyu	hungarian	vvers	mwe	zayt	shabab
3	rusnano	pyhjoki	belene	dabaa	n-plant	belarus	klt	rivne	el_nino	egypt
4	rostec	rusatom	atravyets	turkey	bohunice	belarusian	pressurized-water	vessel	gabal	beni
5	energoatom	energoatom	paks	sinop	temelin	armenia	wwer	npp	balah	zamaleksc
6	rosenergoatom	fortum	visaginas	belarusian	kozloduy	dabaa	vver-type	bohunice	gouna	sisi
7	tvel	subsidiary	khmelnytsky	pyhjoki	astravyets	mochovce	russian-designed	oskarshamn	el	sissi
8	likhachev	tenex	kozloduy	kozloduy	mochovce	dukovany	ritm	rostov	foum	cairo
9	turboatom	krienko	dabaa	tomari	opole	budapest	lwrs	atucha	talkha	burullus
10	tekhnopromexport	rosatoms	hanhikivi	ruppur	visaginas	atucha	cpr	cernavoda	arish	suef

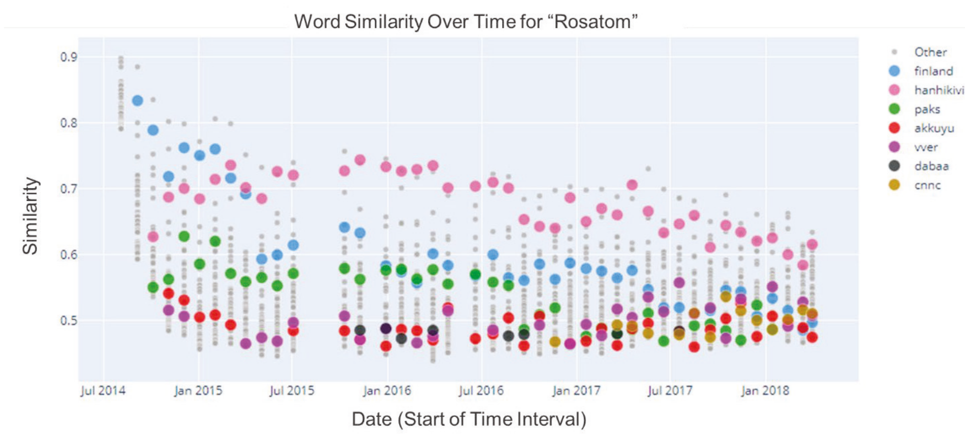

Figure 3-3. Top 50 most similar words over time to "Rosatom". Colored markers indicate a potential entity of interest for the current study.

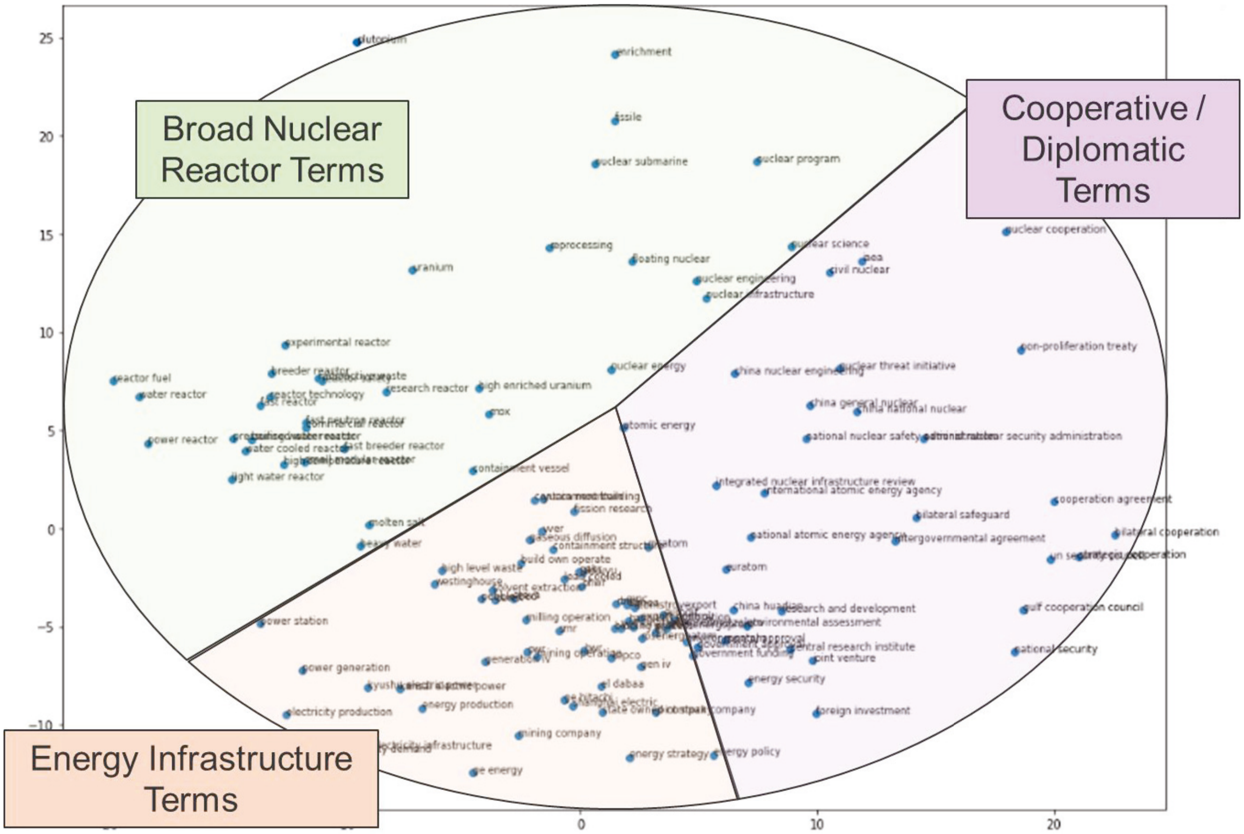

Figure 3-4. Principal component analysis plot of word vectors from the glossary of key terms, demonstrating that key terms from similar sub-domains are in close proximity.

4.0 Conclusions and Future Work

The semi-automated approach that was described in Section 2.0 has offered an improved technique over the approach in FY20-21, requiring significantly less upfront manual fine tuning of the glossary of key terms. Likewise, this general approach can be transferred to any domain of interest, provided the user can identify a text-based data source that broadly captures the relevant language (i.e., noun phrases). During the development stages of the modeling pipeline, the team is using a retrospective dataset. However, in a real deployment of the proposed modeling pipeline, where data may be streaming or periodically updated, this approach could also be applied for dynamically updating query terms as new information is obtained (e.g., if a previously unknown event occurs) and more data becomes necessary.

The exploratory data analysis presented in Section 3.0 has demonstrated that the glossary of key terms has proven effective at obtaining a domain-specific corpus of documents and Tweets related to civil nuclear power around the world. At this time, no additional key terms will be added, however additional data, spanning from May of 2018 to August of 2022, will be obtained using the glossary, thereby providing the team with a more up-to-date database for testing and development of the modeling pipeline. In the next stages of the work, the datasets will be fed through the prototype modeling pipeline that was developed in FY20-21 so that events of possible interest can be extracted.

5.0 References

Danielson, 2022a. T. L. Danielson, B. Mayer, N. Muralidhar, J. Miller, H. Dogan, N. Self, P. Butler, F. Liu. *Machine Learning Modeling Pipeline for Extracting Nuclear Proliferation Events of Interest from Open Data Sources (U)*. SRNL-STI-2022-00036, Rev. 0. January, 2022.

Danielson, 2022b. T. L. Danielson and E. D. LaBone. *Hierarchical Conceptual Model Event and Activity Domains for Forecasting State-Sponsored Civil Nuclear Power Activities*. SRNL-STI-2022-00120, Rev. 0. March, 2022.

Appendix A. Glossary of Key Terms

The following terms were used to formulate queries to the data sources. The base term is listed in the column "Query N-Gram". The queries are keyword based and therefore if the term appears in the article/Tweet, the article/Tweet is added to the dataset. Some terms, such as "enrichment", require higher specificity to ensure noise is not introduced into the dataset (e.g., educational enrichment versus uranium enrichment). Therefore, in some cases, Boolean logic is used to enhance the specificity of the query term. For example, the query for "enrichment" would also require "nuclear" or "uranium" to appear in the article/Tweet in order to return the article/Tweet.

Query N-Gram	AND Terms	OR Terms
advanced	nuclear	
aes92	nuclear	
akkuyu		
angarsk	nuclear	
ap1000	nuclear	
apr1400	nulcear	
areva		
athabasca basin	nuclear	
atmea		
atomic energy		
atomic radiation		
atomic research		
atomic reactor		
atommash		
atomenergoprom		
atomredmetzoloto		
atomspetstrans		
atomstroyexport		
atomtechexport		
atrium	nuclear	
beryllium	nuclear	
bharat heavy electricals		
bidding process	nuclear	
bilateral cooperation		
bilateral safeguard		
bismuth		
bn800	nuclear	
bochvar national research institute		
bohunice	nuclear	
boiling water reactor		
borosilicate glass		
breeder reactor		
brest	nuclear	

brussels convention		
budget	nuclear	
build own operate		
bwr		
cameco		
candu		
cap1400	nuclear	
cea	nuclear	
kyzylkum	uranium	
central research institute	nuclear	
centrifuge plant		
chemical combine	nuclear	
china general nuclear		
china huadian		
china national nuclear		
china nuclear engineering		
cigar lake	uranium	
civil nuclear		
cnnc international		
cnnc overseas		
cogeneration plant		
commercial reactor		
construction company	nuclear	
construction contract	nuclear	
construction permit	nuclear	
construction project	nuclear	
consulting service	nuclear	
containment building	nuclear	
containment structure	nuclear	
containment vessel	nuclear	
contaminated material	nuclear	
control rod		
conversion facility	nuclear	
coolant	nuclear	reactor
cooling	nuclear	
cooperation agreement	nuclear	
cpr1000	nuclear	
daya bay	nuclear	
dco application	nuclear	
decommissioning	nuclear	
demonstration	nuclear	
desalination plant		
design	nuclear	

development agreement	nuclear	
development company	nuclear	
director general	nuclear	
disposal	nuclear	waste
dongfang electric		
dry cask	nuclear	
el dabaa		
electricity demand		
electricity generation		
electricity grid		
electricity infrastructure		
electricity production		
elliot lake	nuclear	uranium
emergency planning	nuclear	
energy bill		
energy commission		
energy consumption		
energy cooperation		
energy demand		
energy export		
energy information administration		
energy market		
energy minister		
energy plan		
energy policy		
energy production		
energy regulator		
energy regulatory commission		
energy research		
energy secretary		
energy security		
energy strategy		
energy trading		
energy trade		
energy technology		
energy union		
engineering service	nuclear	
enrichment	nuclear	uranium
environmental approval	nuclear	
environmental assessment	nuclear	
environmental contamination	nuclear	
environmental impact	nuclear	
environmental management	nuclear	

environmental protection	nuclear	
epr	nuclear	
euratom		
european sustainable nuclear industrial initiative		
eurodif		
experimental reactor		
exploration	uranium	
export control	nuclear	
export market		
fast breeder reactor		
fast neutron reactor		
fast reactor		
feasibility study	nuclear	
feed-in tariff		
fergana valley		
fhr		
fissile	nuclear	
fission product		
fission research		
flibe salt		
floating nuclear		
fluoride salt	nuclear	
fnr	nuclear	
foreign investment	nuclear	
forte energy		
framework agreement	nuclear	
framatome		
fuel	nuclear	
fusion	nuclear	
gansu	nuclear	
gas import		
gas cooled	reactor	
gaseous diffusion		
ge energy	nuclear	
ge hitachi	nuclear	
gen iv	nuclear	
gen4 module	nuclear	
general atomics	nuclear	
general electric	nuclear	
generation ii	nuclear	
generation iii	nuclear	
generation iv	nuclear	
george besse	nuclear	

government approval	nuclear	
government funding	nuclear	
graphite	nuclear	
grid capacity		
grid connection		
grid infrastructure		
grid operator		
grid stability		
grs	nuclear	
gulf cooperation council		
gurvan saihan		
hangzhou	nuclear	
hanhikivi		
harbin electric		
health hazard	nuclear	
health impact	nuclear	
health risk	nuclear	
heat exchanger		
heathgate resource		
heavy water		
heu fuel		
high temperature reactor		
high enriched uranium		
high level waste		
highly enriched uranium		
high temperature reactor		
hinkley point		
holding company	nuclear	
hualong		
hydrogen production		
iaea		
ignalina	nuclear	
industrial development		
inertial confinement		
infrastructure development		
infrastructure project		
inpro	nuclear	
in situ leach		
integrated nuclear infrastructure review		
integrated regulatory review service		
inter rao		
intergovernmental agreement		
intermediate level waste		

international agreement		
international atomic energy agency		
international control		
international cooperation		
international energy agency		
investment	nuclear	
irradiation facility		
irrs	nuclear	
isl	uranium	
isotope production		
isotope separation		
izhorskiye zavody		
jiangxi province		
joint convention	nuclear	
joint ore reserves convention		
joint stock company	nuclear	
jsc	nuclear	
joint venture	nuclear	
jorc	nuclear	
kansai electric power		
kara balta	nuclear	
kepco		
khan resource		
klt40	reactor	
korea hydro		
kota	nuclear	
kozloduy	nuclear	
kurchatov institute		
kyushu electric power		
la hague	nuclear	
lead cooled		
licence	nuclear	uranium
license	nuclear	uranium
life extension	nuclear	
lifetime extension	nuclear	
light water reactor		
ling ao		
lithium beryllium fluoride		
loan	nuclear	
loviisa		
lwr		
magnesium diuranate		
magnox		

mayak	nuclear	
medical isotope		
milling operation		
mining company		
mining operation		
mixed oxide fuel		
molten chloride		
molten lithium		
molten salt		
mox	nuclear	
msr	nuclear	
national agency	nuclear	
national atomic energy agency		
national energy administration		
national energy policy		
national nuclear centre		
national nuclear safety administration		
national nuclear security administration		
national policy	nuclear	
national security	nuclear	
naval fuel		
naval reactor		
nci	nuclear	
neutron irradiation		
neutron moderation		
next generation nuclear		
ninh thuan	nuclear	
nitride fuel		
nnc	nuclear	
non-proliferation treaty		
nuclear accident		
nuclear activity		
nuclear component		
nuclear control		
nuclear cooperation		
nuclear damage		
nuclear development		
nuclear device		
nuclear disarmament		
nuclear electric		
nuclear electricity		
nuclear energy		
nuclear engineering		

nuclear equipment		
nuclear facility		
nuclear fuel		
nuclear future		
nuclear generation		
nuclear industrial safety agency		
nuclear industry		
nuclear infrastructure		
nuclear installation		
nuclear insurance		
nuclear liability		
nuclear material		
nuclear operator		
nuclear policy		
nuclear power		
nuclear program		
nuclear programme		
nuclear project		
nuclear proliferation		
nuclear reactor		
nuclear regulation		
nuclear regulator		
nuclear regulatory		
nuclear research		
nuclear risk		
nuclear science		
nuclear security		
nuclear site		
nuclear submarine		
nuclear supplier		
nuclear technology		
nuclear threat initiative		
nuclear trade		
nuclear unit		
nuclear utility		
nuclear waste		
nuclear weapon		
nukem technologies		
occupational exposure	nuclear	
occupational health	nuclear	
okbm afrikantov		
olkiluoto		
olympic dam	nuclear	

oman	nuclear	
open pit mining		
operating company	nuclear	
ore reserve		
oxide fuel		
paks	nuclear	
parent company	nuclear	
passive shutdown	nuclear	
peak load	nuclear	
pebble bed		
phosphate deposit		
phosphate production		
phwr		
pilot scale	nuclear	
placer deposit	nuclear	
plant	nuclear	
plutonium		
podolsk	nuclear	
powder river basin	nuclear	
power demand		
power generation		
power grid		
power market		
power plant		
power production		
power reactor		
power station		
pressurised water reactor		
prime minister	nuclear	
prism unit	nuclear	
private sector	nuclear	
processing	nuclear	
production	nuclear	
project management	nuclear	
proryv	nuclear	
public opposition	nuclear	
purex		
pwr		
qinshan	nuclear	
radiation damage		
radiation exposure		
radiation hazard		
radiation protection		

radiation safety		
radioactive material		
radioactive release		
radioactive waste		
rao ues		
ratch china		
rbmk		
research and development	nuclear	
rde	nuclear	
reactor building		
reactor component		
reactor construction		
reactor containment		
reactor core		
reactor fuel		
reactor module		
reactor power		
reactor safety		
reactor system		
reactor technology		
reactor unit		
reactor vendor		
reactor vessel		
reactor wall		
refuelling outage	nuclear	
regulatory	nuclear	
remote handling	nuclear	
renewal application	nuclear	
repository	nuclear	
reprocessing	nuclear	
research centre	nuclear	
research facility	nuclear	
research infrastructure	nuclear	
research institute	nuclear	
research laboratory	nuclear	
research program	nuclear	
research reactor		
rte	nuclear	
rosatom		
rusatom		
rosenergoatom		
ryst kuil		
safeugard	nuclear	

safety	nuclear	
sek	nuclear	
shanghai electric		
site preparation	nuclear	
site selection	nuclear	
solvenske elektrarne		
small modular reactor		
smart reactor		
smr		
sodium diuranate		
solid waste	nuclear	
solvent extraction	nuclear	
south urals	nuclear	
spent fuel	nuclear	
spent nuclear fuel		
stable salt reactor		
state nuclear power technology corporation		
state owned company	nuclear	
steam generator		
steam supply system	nuclear	
storage	nuclear	
strategic cooperation	nuclear	
strategic investor	nuclear	
strategic relationship	nuclear	
subsidiary	nuclear	
sulfuric acid	nuclear	
technical cooperation	nuclear	
technical feasibility	nuclear	
technology development	nuclear	
technology transfer	nuclear	
test assembly	nuclear	
test reactor	nuclear	
thorium		
tractebel	nuclear	
transport cask	nuclear	
triso fuel		
turbine	nuclear	
tvel		
type 30b cylinder	nuclear	
type b package	nuclear	
un general assembly	nuclear	
un security council	nuclear	
underground mine		

uniper	nuclear	
united nations	nuclear	
uranium		
uranyl carbonate		
urenco		
urex process		
vienna convention	nuclear	
vver		
vvertoi		
waste directive	nuclear	
waste facility	nuclear	
waste management	nuclear	
waste policy	nuclear	
waste product	nuclear	
waste stream	nuclear	
waste treatment	nuclear	
water moderator	nuclear	
water reactor		
water cooled reactor		
westinghouse		
xenotime		
yucca mountain		
zirconium cladding		

Distribution:

M. J. Barnes, mark.barnes@srnl.doe.gov
J. S. Bollinger, james02.bollinger@srnl.doe.gov
N. J. Bridges, Nicholas.Bridges@srnl.doe.gov
P. Butler, pabutler@vt.edu
G. R. Cefus, gregory.cefus@srnl.doe.gov
T. L. Danielson, Thomas.Danielson@srnl.doe.gov
C. C. Herman, connie.herman@srnl.doe.gov
C. M. Gregory, clint.gregory@srnl.doe.gov
R. B. James, Ralph.James@srnl.doe.gov
R. D. Jeffcoat, ron.jeffcoat@srnl.doe.gov
B. Mayer, bmayer@cs.vt.edu
P. R. Nuessle, patterson.nuessle@srnl.doe.gov
N. Self, nwself@vt.edu
M. Taylor, marc.taylor@srnl.doe.gov
T. P. Taylor, tammy.taylor@srnl.doe.gov
S. J. Branney, sean.branney@srnl.doe.gov
M. Cofer, marion.cofer@srnl.doe.gov
L. T. Brown, lindsay.brown@srnl.doe.gov
B. Lee, brady.lee@srnl.doe.gov
M. Duff, martine.duff@srnl.doe.gov
W. Kuhne, wendy.kuhne@srnl.doe.gov
L. Erikson, luke.erikson@nnsa.doe.gov
C. T. Sloan, craig.sloan@nnsa.doe.gov
Records Administration (EDWS)