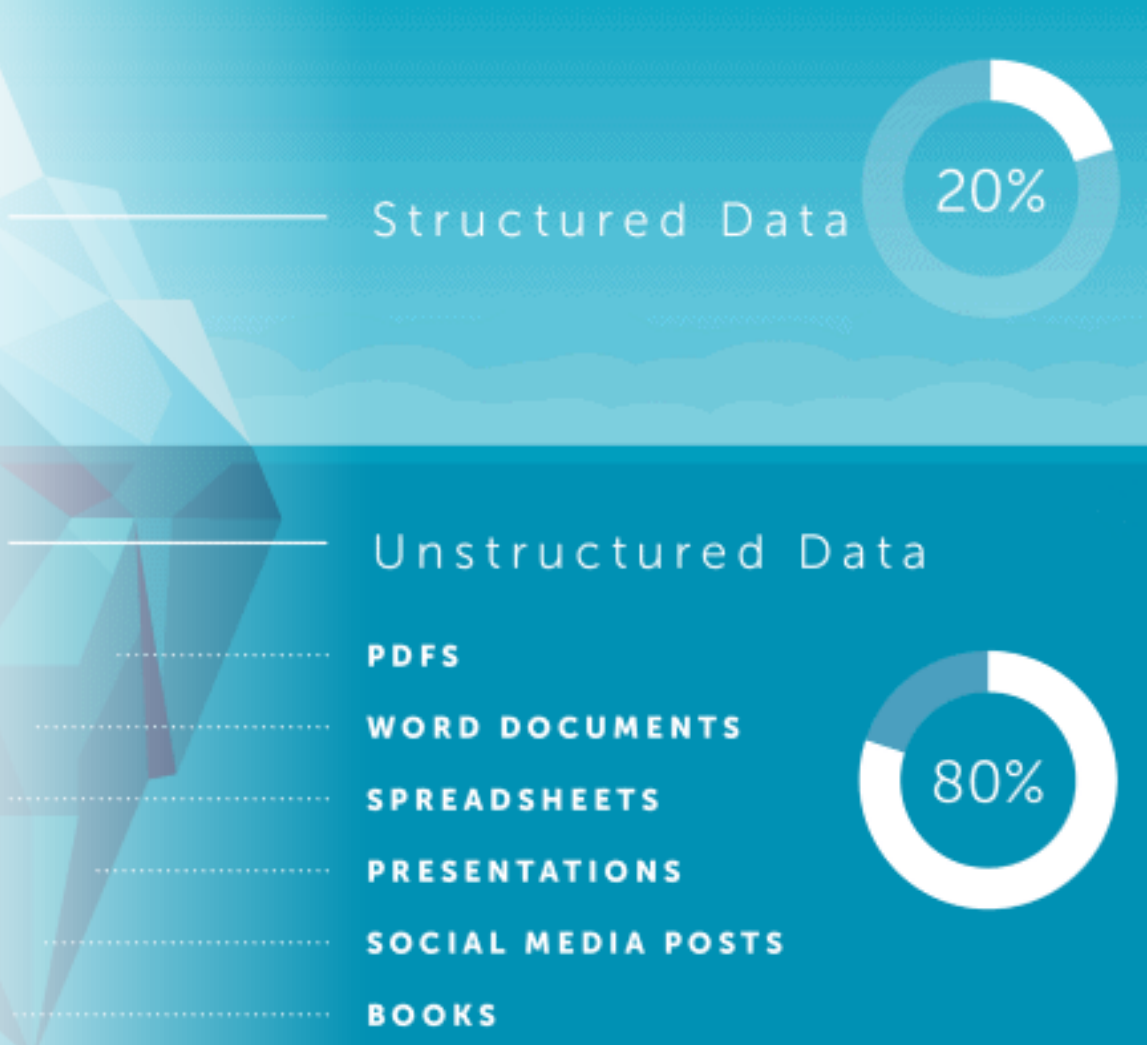
Resource Prediction Using Natural Language Processing

Trey Hathaway
U.S. Nuclear Regulatory Commission
RES/DSA/AAB
August 18, 2021

NRC Data Science and Artificial Intelligence Regulatory Applications Workshops: Current Topics

Natural Language Processing

- Techniques that allow computers to understand the contents of natural language
  - Allows for the extraction of information and insights from documents
  - Collection of techniques:
    - Rule-based, statistical, or neural



Use Cases Goals

Apply Natural Language Processing techniques to NRC data and use cases

Demonstrate Successes

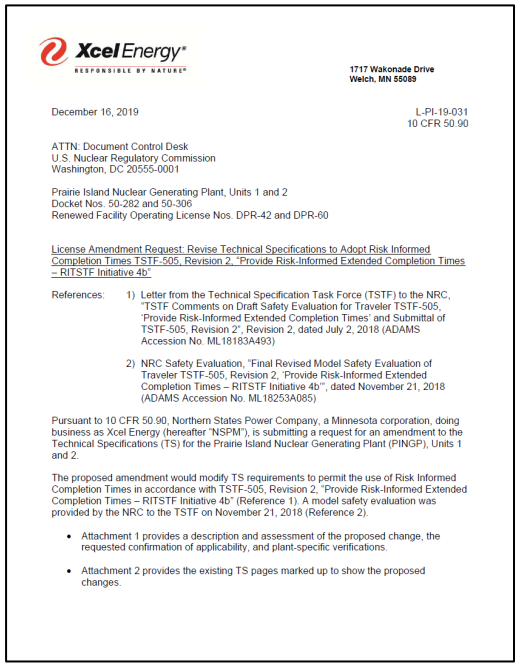
Resource Prediction

- ***Challenge:*** Deviations between resource estimates to complete a licensing review and the actual hours charged
- ***Goal:*** Create tool to assist project managers in formulating resource estimates
  - Leverage historical data
  - Find historically similar reviews
- ***Method:*** Use term frequency-inverse document frequency vectors to represent documents and perform similarity calculations
  - Rank documents based on similarity

Resource Prediction

- **Term Frequency-Inverse Document Frequency (tf-idf)**
  - Weighting factor for words
  - Product term frequency and inverse document frequency
- **Term Frequency (tf)**
  - How frequency a word appears in a document
  - Importance of word
- **Inverse document Frequency (idf)**
  - How frequently a word appears in a collection of documents

Term Frequency-Inverse Document Frequency
(Vector Representation)



Xcel Energy®
RESPONSIBLE BY NATURE®

1717 Wakonade Drive
Welch, MN 55089

December 16, 2019

L-PI-19-031
10 CFR 50.90

ATTN: Document Control Desk
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

Prairie Island Nuclear Generating Plant, Units 1 and 2
Docket Nos. 50-282 and 50-306
Renewed Facility Operating License Nos. DPR-42 and DPR-60

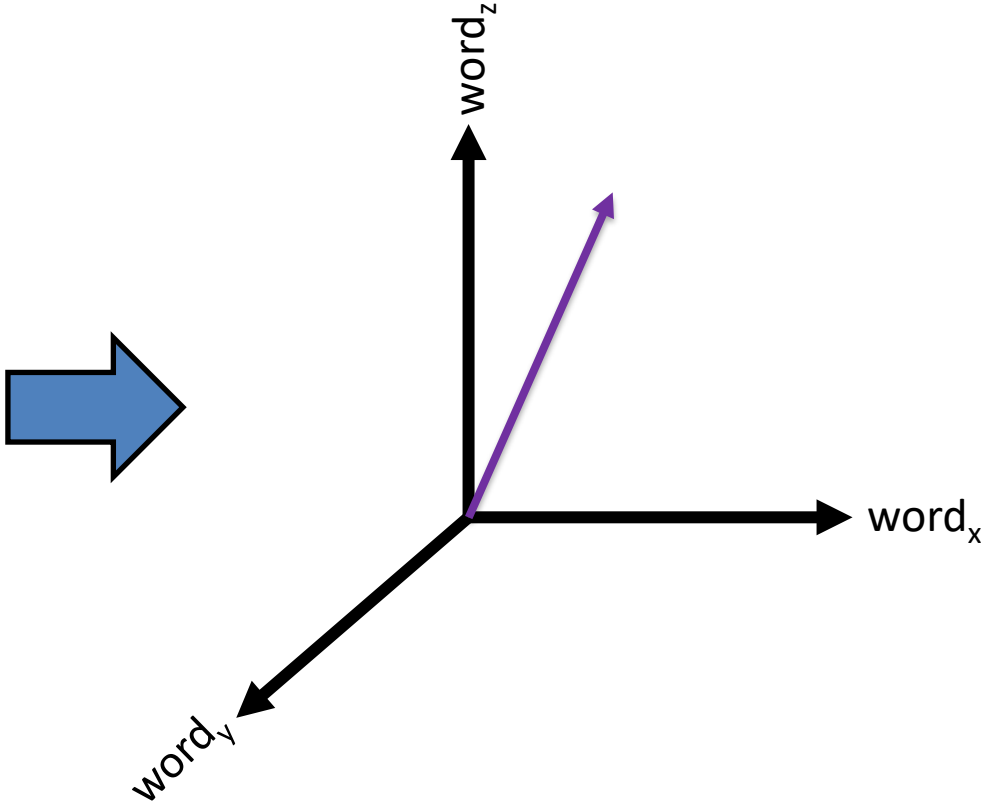
License Amendment Request: Revise Technical Specifications to Adopt Risk Informed Completion Times TSTF-505, Revison 2, "Provide Risk-Informed Extended Completion Times – RITSTF Initiative 4b"

References:
1) Letter from the Technical Specification Task Force (TSTF) to the NRC, "TSTF Comments on Draft Safety Evaluation for Traveler TSTF-505, 'Provide Risk-Informed Extended Completion Times' and Submittal of TSTF-505, Revision 2", Revision 2, dated July 2, 2018 (ADAMS Accession No. ML18183A493)
2) NRC Safety Evaluation, "Final Revised Model Safety Evaluation of Traveler TSTF-505, Revision 2, 'Provide Risk-Informed Extended Completion Times – RITSTF Initiative 4b'", dated November 21, 2018 (ADAMS Accession No. ML18253A085)

Pursuant to 10 CFR 50.90, Northern States Power Company, a Minnesota corporation, doing business as Xcel Energy (hereafter "NSPM"), is submitting a request for an amendment to the Technical Specifications (TS) for the Prairie Island Nuclear Generating Plant (PINGP), Units 1 and 2.

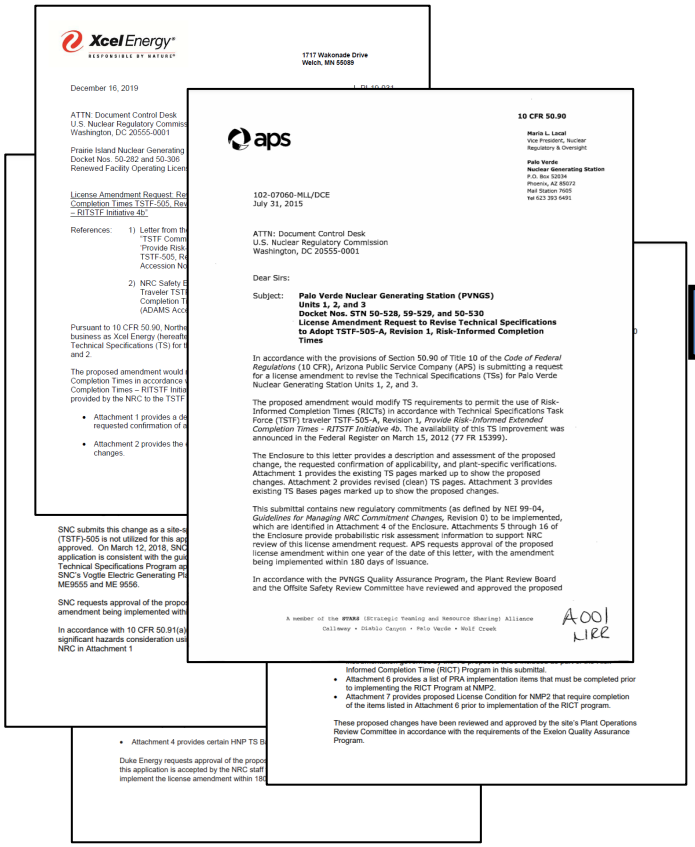
The proposed amendment would modify TS requirements to permit the use of Risk Informed Completion Times in accordance with TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times – RITSTF Initiative 4b" (Reference 1). A model safety evaluation was provided by the NRC to the TSTF on November 21, 2018 (Reference 2).

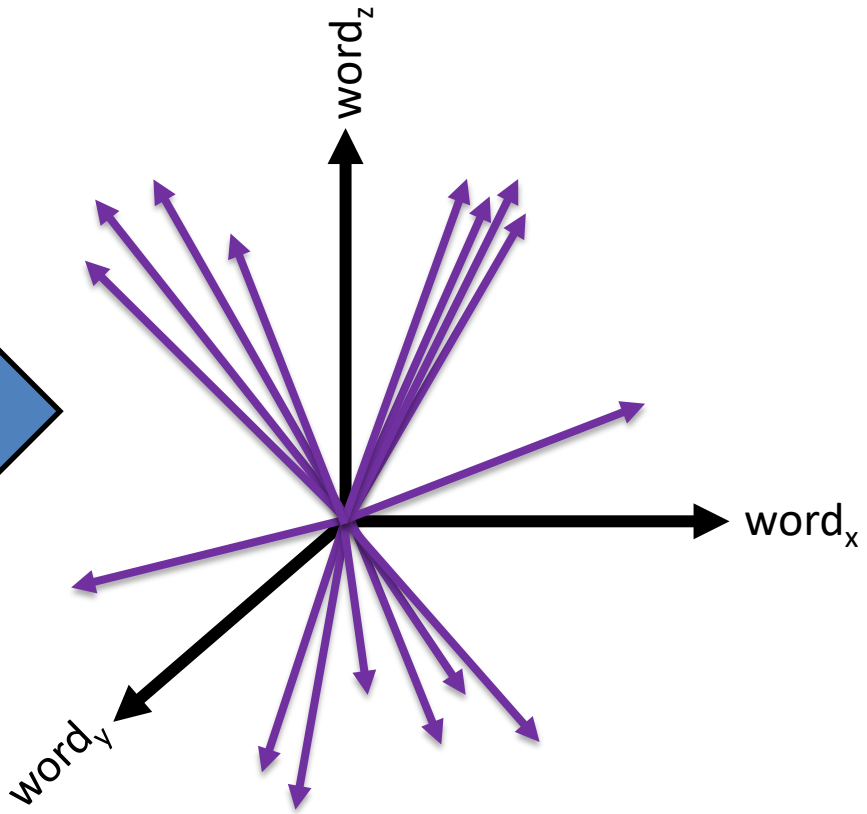
- Attachment 1 provides a description and assessment of the proposed change, the requested confirmation of applicability, and plant-specific verifications.
- Attachment 2 provides the existing TS pages marked up to show the proposed changes.



- Represent a document as a vector
  - The vector reflects the word usage in the document
  - The vector will have 1000's of dimensions

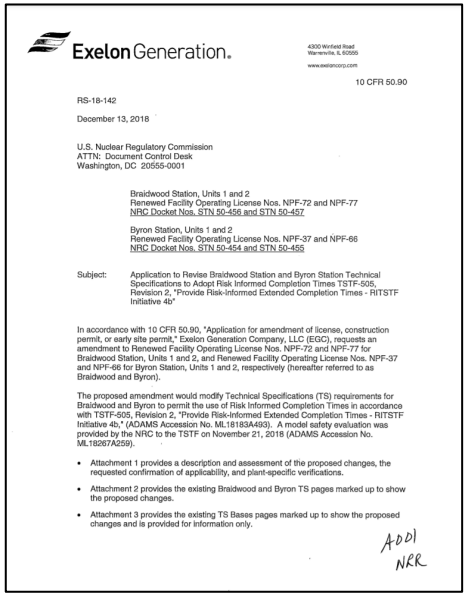
Term Frequency-Inverse Document Frequency
(Vector Space Corpus)





- Represent the collection of documents as vectors
  - Create a vocabulary of all words used in the collection

Term Frequency-Inverse Document Frequency
(Similarity Calculations)



Exelon Generation.

4300 Winfield Road
Warrenville, IL 60555
www.exeloncorp.com

10 CFR 50.90

RS-18-142

December 13, 2018

U.S. Nuclear Regulatory Commission
ATTN: Document Control Desk
Washington, DC 20555-0001

Braidwood Station, Units 1 and 2
Renewed Facility Operating License Nos. NPF-72 and NPF-77
NRC Docket Nos. STN 50-456 and STN 50-457

Byron Station, Units 1 and 2
Renewed Facility Operating License Nos. NPF-37 and NPF-66
NRC Docket Nos. STN 50-454 and STN 50-455

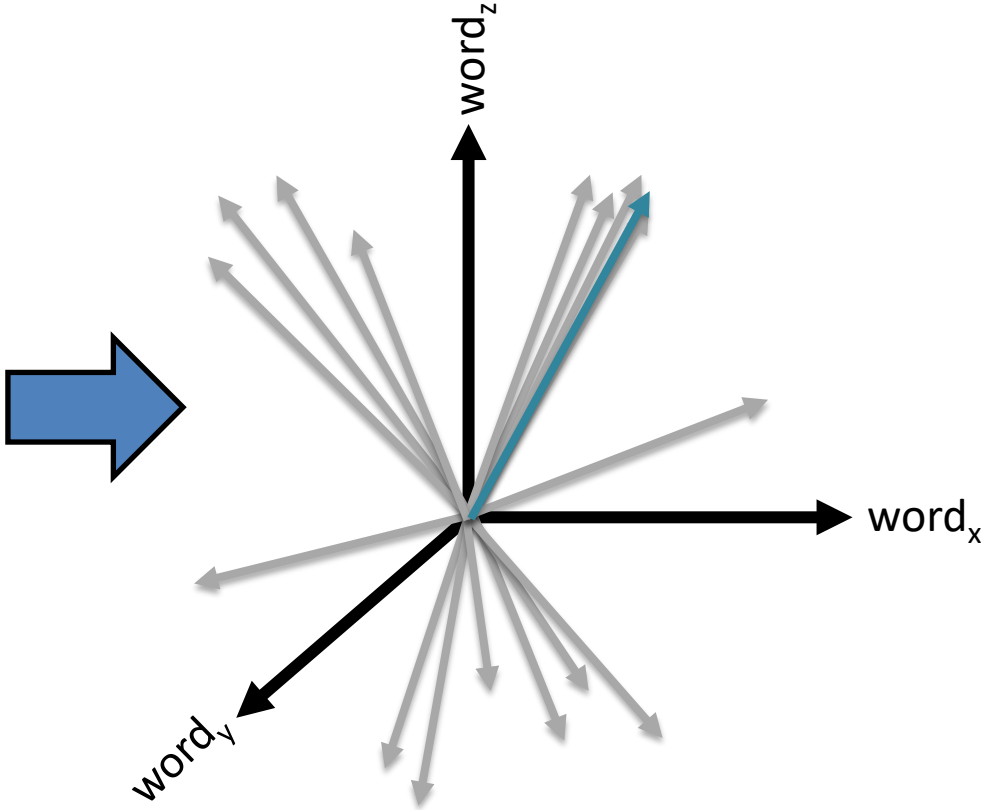
Subject: Application to Revise Braidwood Station and Byron Station Technical Specifications to Adopt Risk Informed Completion Times TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times - RITSTF Initiative 4b"

In accordance with 10 CFR 50.90, "Application for amendment of license, construction permit, or early site permit," Exelon Generation Company, LLC (EGC), requests an amendment to Renewed Facility Operating License Nos. NPF-72 and NPF-77 for Braidwood Station, Units 1 and 2, and Renewed Facility Operating License Nos. NPF-37 and NPF-66 for Byron Station, Units 1 and 2, respectively (hereafter referred to as Braidwood and Byron).

The proposed amendment would modify Technical Specifications (TS) requirements for Braidwood and Byron to permit the use of Risk Informed Completion Times in accordance with TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times - RITSTF Initiative 4b," (ADAMS Accession No. ML18183A493). A model safety evaluation was provided by the NRC to the TSTF on November 21, 2018 (ADAMS Accession No. ML18267A259).

- Attachment 1 provides a description and assessment of the proposed changes, the requested confirmation of applicability, and plant-specific verifications.
- Attachment 2 provides the existing Braidwood and Byron TS pages marked up to show the proposed changes.
- Attachment 3 provides the existing TS Bases pages marked up to show the proposed changes and is provided for information only.

ADDI
NRR



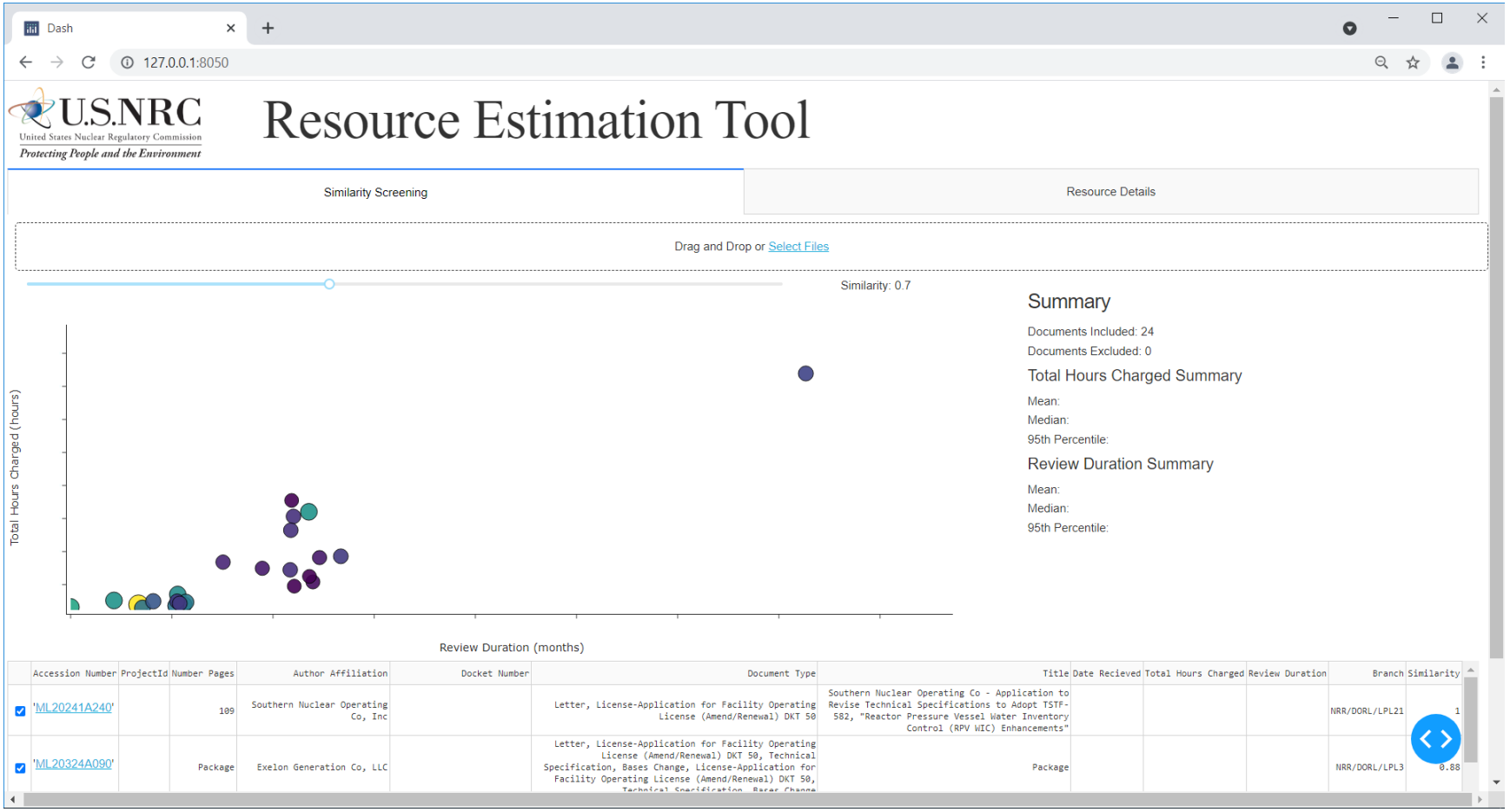
- A new document is converted to a vector based on the vocabulary of the collection of documents
  - The similarity (angle between vectors) is calculated as the dot product between vectors
  - Documents ranked by similarity score

Resource Prediction

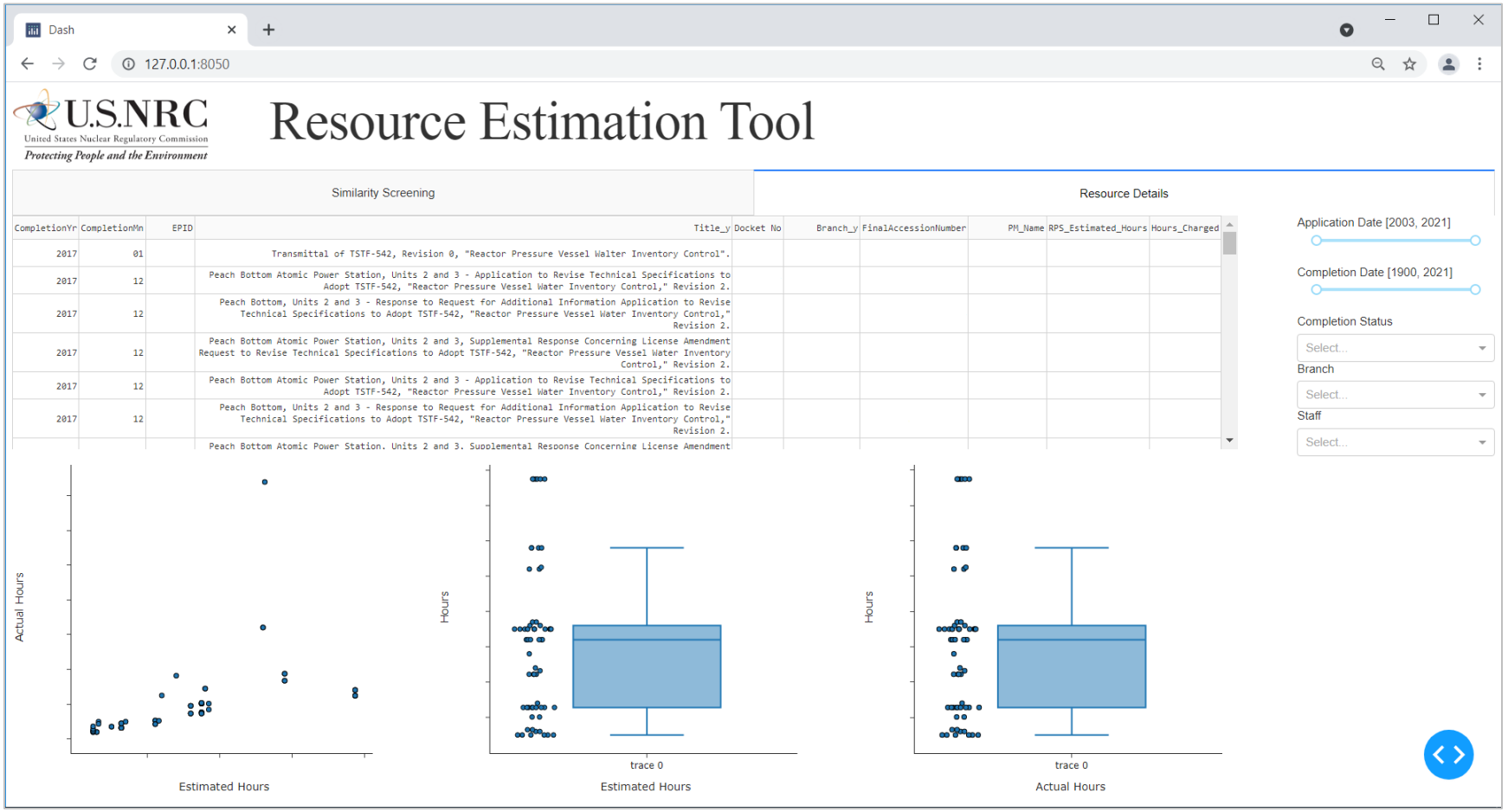
Approach

- Acquire historical licensing actions and resource requirements
- Extract text data from pdf files
- Clean data
- Create tf-idf matrix
- Create User Interface
  - Extracts text data
  - Performs similarity calculations

Resource Estimation Tool



Resource Estimation Tool



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Current Status and Follow-on Work

- Preliminary acceptance testing complete
  - Historical data provides reasonable estimates of required resources and review durations
- NRR/EMBARK and NRR/DORL coordinating to finalize visualizations
- Develop and deploy final User Interface
- Potential Follow-on Work:
  - Search capabilities
  - Predict Branch assignments
  - Predict Standard Review Plan
  - Predict which Regulatory Guide(s) was used for the licensing action

Regulatory Named Entity Recognition

- ***Challenge:*** Title 10 of the Code of Federal Regulations (CFR), and other regulatory documents, reference sections of 10 CFR
  - Revisions to 10 CFR could impact other sections
- ***Goal:*** Create a tool to find and extract 10 CFR references from documents
- ***Method:*** Use Named Entity Recognition (NER) to label text as regulations and extract that text

Named Entity Recognition

(2 CARDINAL) On or before July 26, 1990 DATE , each holder of an operating license for a production or utilization facility in effect on July 27, 1990 DATE , shall submit information in the form of a report as described in 10 CARDINAL CFR 50.75 of this part, indicating how reasonable assurance will be provided that funds will be available to decommission the facility. [21 CARDINAL FR 355, Jan. 19, 1956 DATE , as amended at 35 FR 19660 DATE , Dec. 29, 1970 DATE ; 38 CARDINAL FR 3956, Feb. 9, 1973 DATE ; 45 CARDINAL FR 55408, Aug. 19, 1980 DATE ; 49 CARDINAL FR 35752 WORK_OF_ART , Sept. 12, 1984 DATE ; 53 CARDINAL FR 24049 DATE , June 27, 1988 DATE ; 69 CARDINAL FR 4448 WORK_OF_ART , Jan. 30, 2004 DATE ; 72 CARDINAL FR 49490 CARDINAL , Aug. 28, 2007] DATE 4 Emergency planning zones (EPZs) are discussed in NUREG–0396 DATE , EPA ORG 520/1–78–016, Planning Basis for the Development of State and Local Government Radiological Emergency Response Plans NORP in Support of Light-Water Nuclear Power Plants, December 1978 DATE .

SpaCy Default Entities

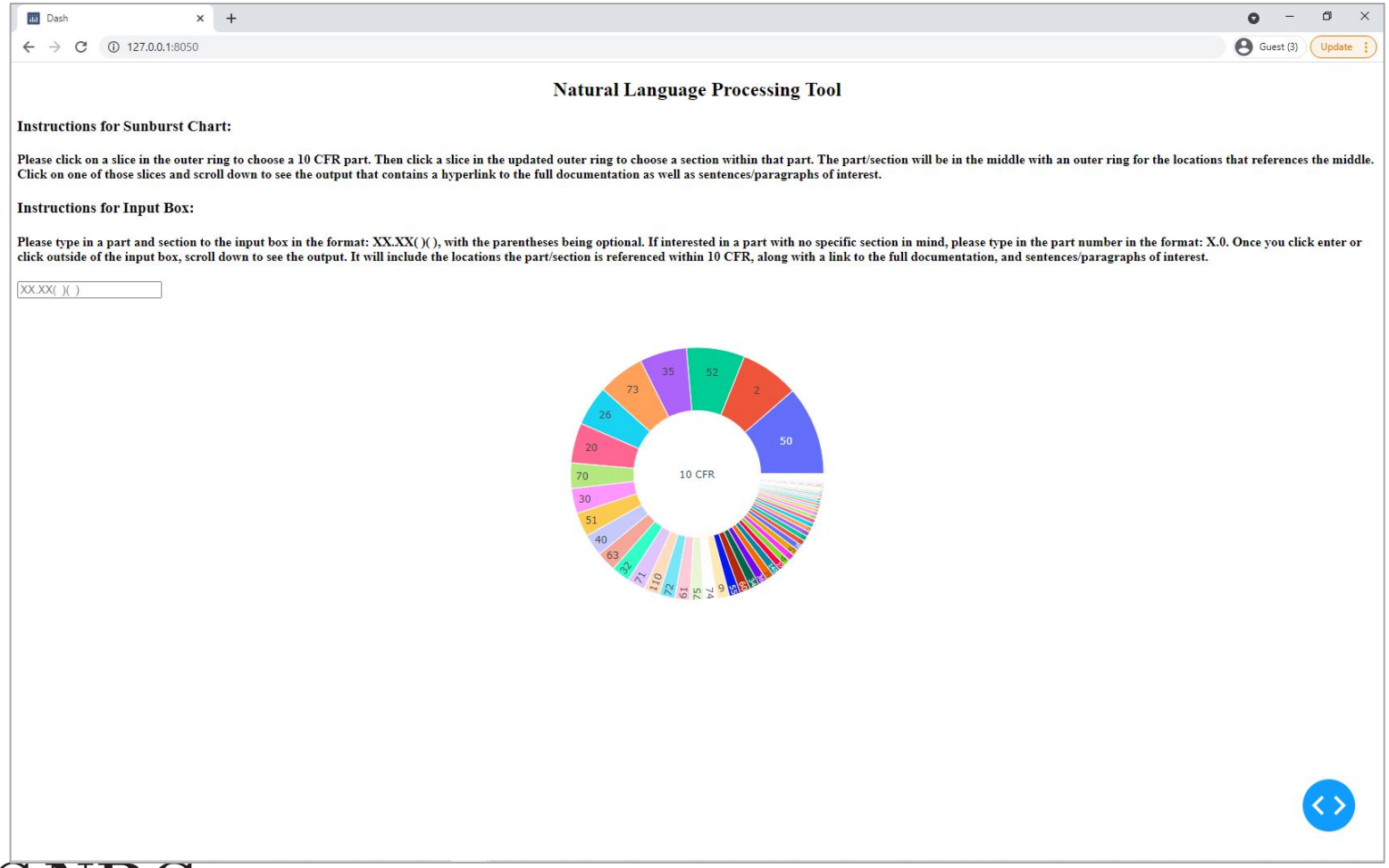
Addition of NRC Specific Language Patterns

(2 CARDINAL) On or before July 26, 1990 DATE , each holder of an operating license for a production or utilization facility in effect on July 27, 1990 DATE , shall submit information in the form of a report as described in 10 CFR 50.75 REG of this part, indicating how reasonable assurance will be provided that funds will be available to decommission the facility. [21 FR 355 FR , Jan. 19, 1956 DATE , as amended at 35 FR 19660 FR , Dec. 29, 1970 DATE ; 38 FR 3956 FR , Feb. 9, 1973 DATE ; 45 FR 55408 FR , Aug. 19, 1980 DATE ; 49 FR 35752 FR , Sept. 12, 1984 DATE ; 53 FR 24049 FR , June 27, 1988 DATE ; 69 FR 4448 FR , Jan. 30, 2004 DATE ; 72 FR 49490 FR , Aug. 28, 2007] DATE 4 Emergency planning zones (EPZs) are discussed in NUREG–0396 NUREG , EPA ORG 520/1–78–016, Planning Basis for the Development of State and Local Government Radiological Emergency Response Plans NORP in Support of Light-Water Nuclear Power Plants, December 1978 DATE .

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

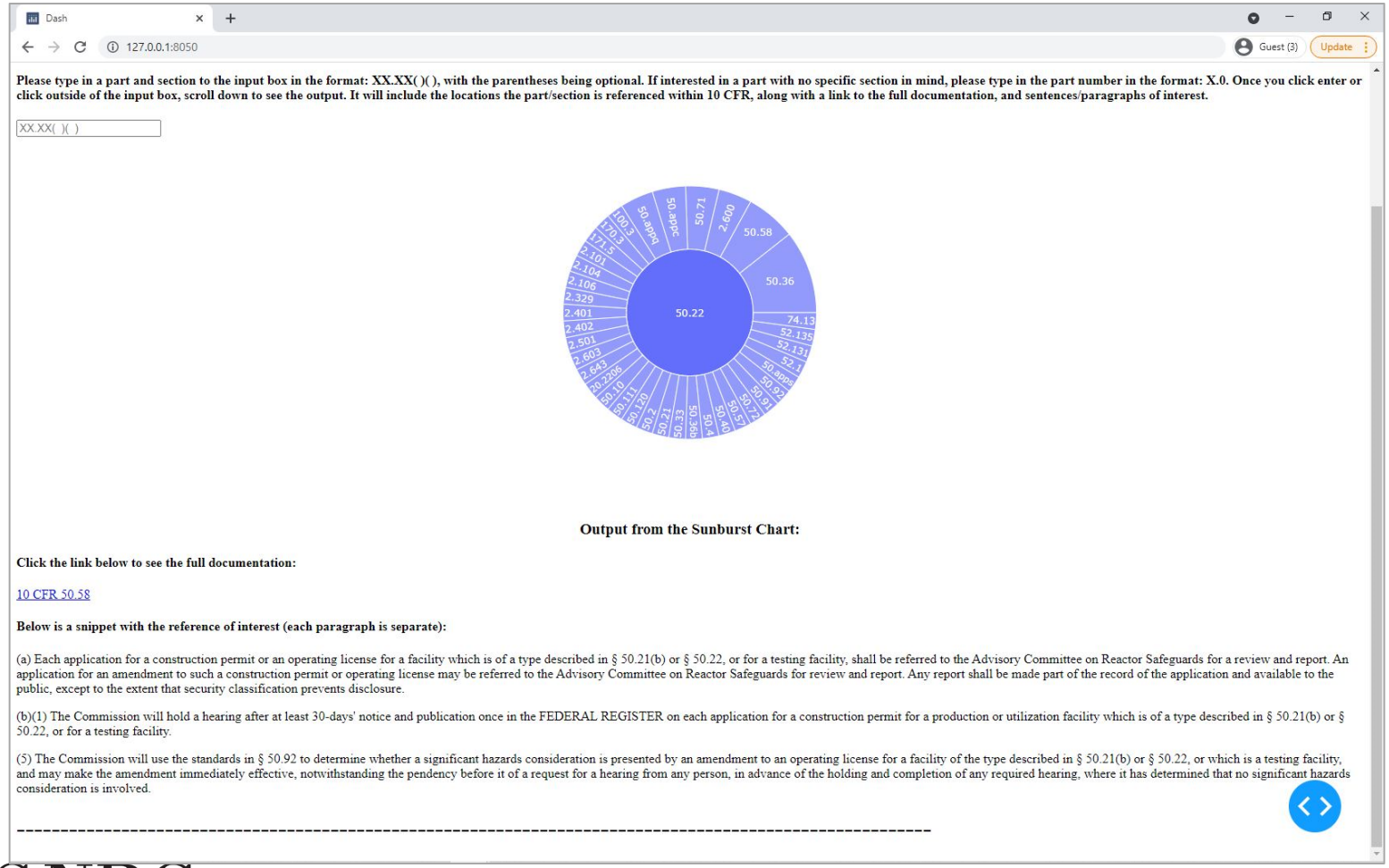
- ***Used Python package Spacy***

10 CFR Reference Identification Tool



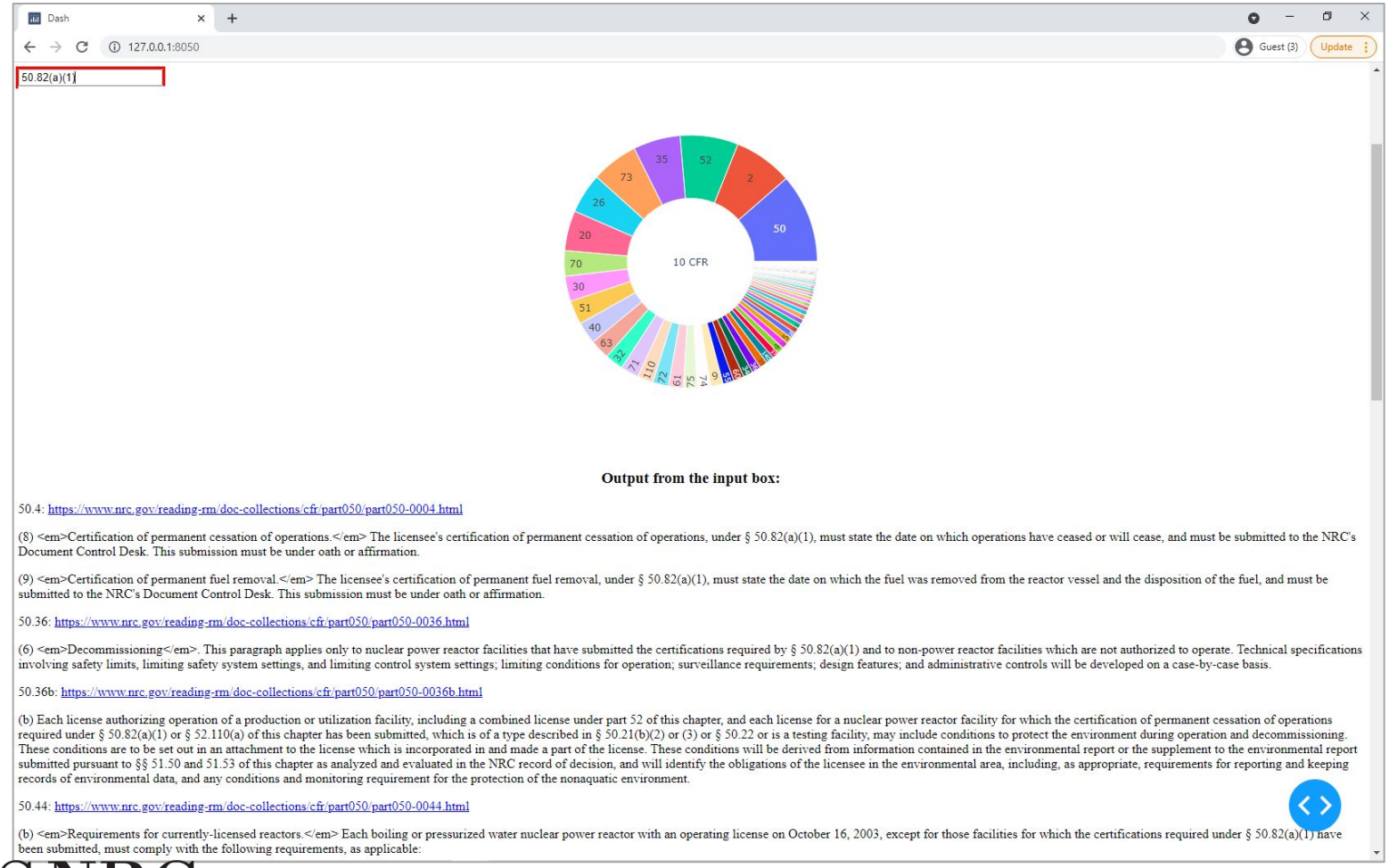
U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

10 CFR Reference Identification Tool



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

10 CFR Reference Identification Tool



Conclusions

- Natural Language Processing is a powerful tool to leverage unstructured data in historical documents
- Deploying these tools would increase efficiency of staff by reducing time required for manual searches
  - Staff can leverage historical data in informing decisions