Advanced Knowledge Engineering Tools to Support Risk-Informed Decision Making: Final Report (Public Version)

N. Siu, S. Dennis, M. Tobin, P. Appignani, K. Coyne

Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission

G. Young

Office of the Chief Information Officer
U.S. Nuclear Regulatory Commission

S. Raimist

ECM Universe

December, 2016

U.S. Nuclear Regulatory Commission
Washington, DC

ABSTRACT

The Nuclear Regulatory Commission (NRC) has well established processes for the use of risk information, in combination with other engineering insights and analyses, in regulatory decision-making. However, risk-informed applications tend to rely on highly complex analyses that are derived from a wide variety of source documents. In addition, trends affecting the NRC's increased use of risk information include the increasing number of nuclear power plant risk-informed licensing applications, the broader (and more challenging) range of applications of risk assessment methods, the increasing demands on the risk models supporting these applications, and the changing ways that people interact with information systems. Recognizing that significant advances are being made in the information technology community, the NRC has conducted a feasibility study on the application of advanced knowledge engineering tools and techniques to support the improved and expanded use of risk information. This study was undertaken as part of the NRC's Long-Term Research Program, whose main objective is to assess emerging technologies and determine their feasibility for further research or regulatory applications.

This study involved demonstration applications of content analytics software (i.e., IBM Content Analytics Version 2.2) currently available to the NRC staff. The project applications, called "use cases," included the identification and characterization of events involving multiple nuclear power plants ("multi-unit events"), and the characterization of results from recent nuclear power plant probabilistic risk assessment (PRA) studies. These use cases represented two different applications that the staff had already developed an experience base using traditional information search techniques (e.g., ADAMS searches) but were highly resource intensive. The use case results indicate that: a) the content analytics software tested is generally effective and efficient in helping analysts identify target documents of interest, b) subject matter experts must be involved when using the software to develop practical problem-specific tools, c) the software tool developed to support this feasibility analysis appears to be capable (without further development) of supporting current NRC staff activities beyond those explored in the feasibility study, and d) further development of the software could increase its power and usefulness. Therefore, the NRC's current in-house content analytics capabilities appear to be well suited for increasing the efficiency of certain risk applications, provided sufficient time and resources are available to develop use-case specific information. Additionally, further improvements in the content analytics software platform could reduce future time and resource demands, making such tools more accessible to the staff.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of J. Lane and Idaho National Laboratory staff in obtaining the electronic files for most of the Licensee Event Reports addressed in this project. Fruitful discussions with project observers D. Halverson and T. Nakanishi are also appreciated.

TERMINOLOGY

Acronyms

ACRS	Advisory Committee on Reactor Safeguards
ADAMS	Agencywide Document Access and Management System
ANS	American Nuclear Society
ASME	American Society of Mechanical Engineers
ASP	accident sequence precursor
CDF	core damage frequency
EDG	emergency diesel generator
FSAR	Final Safety Analysis Report
FY	fiscal year
IBM	International Business Machines
ICA	IBM Content Analytics
INL	Idaho National Laboratory
IPE	Individual Plant Examination
IPEEE	Individual Plant Examination of External Events
KE	knowledge engineering
LAR	License Amendment Request
LCO	Limiting Condition of Operation
LER	Licensee Event Report
LOCA	Loss of Coolant Accident
LTRP	Long-Term Research Project
MSPI	Mitigating Systems Performance Index
NRC	U.S. Nuclear Regulatory Commission
NRR	Office of Nuclear Reactor Regulation
OCIO	Office of the Chief Information Officer
OCR	optical character recognition
PDF	portable document format
PRA	probabilistic risk assessment
PSA	probabilistic safety assessment
PSAM	Probabilistic Safety Assessment and Management
RES	Office of Nuclear Regulatory Research
RMIEP	Risk Methods Integration and Evaluation Program
RMTF	Risk Management Task Force
ROP	Reactor Oversight Program
SAMA	Severe Accident Mitigation Alternative
SME	subject matter expert
SPAR	Standardized Plant Analysis Risk
SRM	Staff Requirements Memorandum

Other Terms

Term	Usage in this report
Annotator	A software tool used to annotate a document with information used to facilitate searches. Such information is extracted or deduced from the document content.
Content Analytics	A broad class of software tools that use a variety of approaches (e.g., natural language queries, trends analysis, contextual discovery and predictive analytics) to identify patterns and trends across an unstructured database.
Corpus	A selected set of documents which provides the search space for the project use cases.
Crawler	A software tool used to browse data sources and extract content.
Discovery/Exploration	An open-ended use of a database in which responses to a particular search can suggest further searches that provide alternate perspectives on the topic of interest.
Facet	A particular window on/view of/aspect of data in a database. For example, a “multi-unit” facet for a database of nuclear power plant operational events could capture multi-unit aspects (e.g., causes, coupling mechanisms, near misses) of those events. Operationalized in International Business Machines (IBM®) Content Analytics Version 2.2 (ICA 2.2) through the construction and use of keyword lists.
Indexer	A software tool that builds a specialized data structure used to facilitate rapid access to data records.
Keyword	A string of characters whose presence indicates that the document may have information relevant to the facet. Keywords can, but need not be complete words or phrases.
Search	A process involving the identification of specific answers to a particular question.
Structured Data	A collection of information that is grouped into specific fields. Spreadsheets and relational databases are typical examples.
Text Mining	A process of developing (e.g., through the identification of patterns and trends) information from text.
Unstructured Data	A collection of information that is not grouped into specific fields. Although text documents typically include high-level structural elements (e.g., organization by chapters and sections) and often include lower-level structural elements (e.g., tables), the information within these elements is unstructured. Moreover, the structural elements will typically vary from document to document.
Use Case	A specific application of a tool aimed at identifying, in a realistic setting, the positive and negative aspects of the tool and the tool application process.

1. INTRODUCTION

In accordance with its Probabilistic Risk Assessment (PRA) Policy Statement issued in 1995 [1], and as recently discussed by the U.S. Nuclear Regulatory Commission's (NRC's) Risk Management Task Force (RMTF) in NUREG-2150 [2], the NRC continues to improve and expand its use of risk information in its regulatory activities. Although much progress has been made, continuing challenges include addressing the breadth, depth, diversity, and quality requirements of NRC's risk information needs, and the analogous demands on the information needed to develop the PRA models that generate this risk information. The drive to improve and expand risk-informed regulatory capabilities within an environment of constrained resources constitutes a challenge to information systems and solutions available to the staff, and to the staff themselves.

In recent years, a variety of advanced knowledge engineering (KE) tools and techniques potentially useful to NRC staff have emerged. These tools and techniques address such technical challenges as the use of naturally-posed ("natural language") questions and answers to explore technical documents, the analysis of document content, and the encoding and application of expert knowledge in the creation and review of systems models. Some of these tools and techniques (e.g., those associated with advanced natural language processing, as popularized by the International Business Machines (IBM®) Watson project [3]) appear to be in a developmental stage. Others (e.g., those associated with analyzing document content) appear to be already in use by government agencies facing information management problems similar to those faced by NRC.

The NRC's Office of Nuclear Regulatory Research (RES) has performed a feasibility study to explore the application of advanced knowledge engineering (KE) tools and techniques to support PRA activities. The study was initiated in Fiscal Year (FY) 2014 under the auspices of the NRC's Long-Term Research Program (LTRP), which is used to investigate topics expected to meet critical mission needs in 5 to 10 years [4]. Due to the availability of resources, and consistent with the scope of the LTRP, the project was conducted as a scoping study aimed at determining if additional agency effort to develop production-level KE tools aimed at supporting risk-informed applications could be worthwhile.

The feasibility study addressed currently available software tools used to extract information from large, unstructured information bases. In particular, the work explored the potential of "content analytics" tools,[1] as represented by an in-house tool IBM Content Analytics Version 2.2 (ICA 2.2) [5].[2]

---

[1] Although there is no standard definition for the term "content analytics," in an information technology context, it can be generally viewed as describing a broad class of software tools that use a variety of approaches (e.g., natural language queries, trends analysis, contextual discovery and predictive analytics) to identify patterns and trends across an unstructured database (e.g., text).

[2] As originally envisioned [6], the project was intended to involve work along two lines of effort: (1) the content analytics work discussed in this report, and (2) an investigation of the potential of ongoing, non-NRC activities, notably the Open PSA initiative (see www.open-psa.org for information on this initiative)

The work involved the performance of three case studies ("use cases"): 1) the identification and characterization of operational events involving multiple reactor units, 2) the determination of current core damage frequency (CDF) estimates developed in licensee PRAs, and 3) a general exploration of a wide set of documents to identify potentially interesting risk-relevant topics for more detailed investigation. The first two use cases employ the ICA 2.2 tool in a traditional search mode to address prototypical tasks in which the analyst seeks to find answers to specific questions. The last use case employs the ICA 2.2 tool in a more general, discovery-oriented mode.[3]

For each use case, this report discusses the specific objectives, the approach taken, and the results obtained. The report then concludes with some general observations on the state of KE tool technology and a number of recommendations for future NRC work. It should be recognized that the information technology environment is changing rapidly, and that tools that became available (on a limited basis) to the staff during the project (notably the NRC's Agency Wide Documents Access and Management System – ADAMS – Enterprise Search tool) are now being deployed within the agency. This report accounts for this dynamic in its recommendations.

N.B.: This report has been created from an internal staff report written in September, 2016. The only changes made involve the removal of non-publicly available information. The essential content and conclusions are therefore current as of September, 2016.

---

and related projects, aimed at improving current PRA modeling and documentation practices [7]. (From an NRC staff reviewer's perspective, the intriguing aspect of such efforts is that they are aimed at developing standardized representations of models. Such standardized representations could, for example, facilitate and even enrich comparisons between models of similar systems.) Due to time limitations and the demands of higher priority agency activities, the second line of effort was not pursued.

[3] In the content analytics literature, the terms "discover" and "explore" are used to indicate a more open-ended use of a database than implied by the term "search," which involves looking for the specific answers to a particular question. Thus, a tool aimed at supporting "discovery" can, in addition to responding to a particular search, provide information suggesting further searches that provide alternate perspectives on the search topic

2. BACKGROUND

2.1 TRENDS IN NRC RISK INFORMATION NEEDS

As discussed in NUREG-2150 [2], the NRC currently uses risk information in all areas of regulatory purview, including materials (e.g., medical sources), waste (both low level and high level), uranium recovery, fuel cycle facilities, interim spent fuel storage, and transportation, as well as reactors (both power and non-power). The extent and formality of usage varies across the program areas, and the area-specific recommendations provided by NUREG-2150 vary accordingly. These recommendations are provided in Appendix A of this report. It can be seen that a number of the recommendations (e.g., regarding the use of PRA insights when defining design basis events) have KE implications (e.g., in helping users identify fleet-wide insights relevant to a proposed design basis event, and in understanding the models and modeling assumptions underlying those insights).

It's also worth mentioning that, especially in the reactor safety area, risk information is being used to support all regulatory functions (see Figure 1) and at all organizational levels [8]. Thus, for example, risk-informed decision making arises in day-to-day staff decisions (e.g., the prioritization of onsite inspection items), in major Commission policy decisions (e.g., regarding the imposition of broad requirements for filtered, containment venting), and situations in-between (e.g., deciding whether to allow a plant to continue operation under degraded conditions). In principle, it seems that advanced KE tools could provide NRC staff with (1) improved access to a much larger information base than the PRA and hard-linked documents (i.e., the information base would likely include documents not identified by the PRA authors but relevant to the technical issues being reviewed), and (2) flexible, expert-informed tools to query this information base. This querying could support both the creation of the PRA model and efforts to use the model results and insights in support of risk-informed decision making. Thus, these tools could support performance of the Analysis and Deliberation steps identified by the RMTF's characterization of the regulatory decision making process (see Figure 2), and likely other steps in that process as well.

Figure 1. NRC's regulatory functions

Figure 2. The regulatory decision making process (from NUREG-2150 [2])

A number of trends relevant to risk information needs have been underway for some time at the NRC. These include:

- the increasing use of currently available PRA models (including Standardized Plant Analysis Risk – SPAR – models [9]) in "routine" applications;
- an increasing number of applications requiring extensions of the PRA models (e.g., allowed levels of reactor pressure vessel embrittlement, understanding the safety significance of consequential steam generator tube ruptures), with an associated burden on decision makers to understand model assumptions and limitations in these extended applications;
- the increasingly demanding use of the PRA results and insights (e.g., to support decisions where the absolute results play a significant role in the decision making process and/or where there are increasing demands on the explanatory power of the PRA in addressing system details);
- the increasing need to review reactor design applications with novel features (e.g., multiple reactor modules, new concepts of operation);
- the difficulty in communicating generic insights derived from increasingly focused and complex application-specific risk studies across organizational boundaries;
- the changing ways that people are interacting with information systems;[4] and
- the changing demographics of NRC staff, which affects the average level of risk-related experience of the staff (many new staff members have not had the chance to develop hands-on experience with practical PRA modeling prior to joining NRC).

The March 11, 2011 Great East Japan earthquake and subsequent tsunamis, core meltdowns at the Fukushima Dai-ichi plant, and strong safety challenges at other Japanese nuclear power plants, have highlighted a number of additional issues relevant to the performance and use of PRA. The resolution of these issues may further affect the agency's risk information needs. These issues include [11]:

---

[4] For example, Ref. 10 provides an overview of changes in reading habits (e.g., skimming versus deep reading) spurred by digital devices and media and associated effects on comprehension.

- the scope of current PRAs for operating U.S. plants (many of which, according to a recent U.S. Government Accountability Office report [12], have not updated their treatment of external events since the Individual Plant Examination of External Events – IPEEE – program in the 1990s [13]);
- the risk from events involving multiple units at a site and even multiple sites;
- the assessment and treatment of uncertainty for extreme natural events (and, more generally, low-probability high-consequence events); and
- the appropriate balancing of deterministic and probabilistic information in regulatory decision making.

Regarding the last point, following the Fukushima Dai-ichi accident, there have been calls to reduce the emphasis on, or even to entirely abandon the explicit use of risk information. Such calls include recommendations to perform “worst-case analyses” and to develop mitigation strategies that are independent of accident cause. Even within the PRA community, there have been proposals to increase the emphasis on “conditional analyses” and “resilience” [14,15]. Recognizing the importance of challenging assumptions, these proposals deserve a thorough and open debate that is beyond the scope of this paper. The outcome of such a debate, some of which is expected to occur as the agency addresses the recommendations of NUREG-2150, could very well affect the agency’s risk information needs.

2.2 KNOWLEDGE ENGINEERING CHALLENGES

2.2.1 General Challenge

A tremendous amount of information potentially useful to the development and use of risk models is currently being generated. For example, for the Fukushima Dai-ichi accident alone, there are currently over 20 official Japanese investigation and lessons-learned reports and numerous official reports from other countries and international bodies, all of which provide voluminous but useful information from a variety of viewpoints. In addition to event reports, such varied sources as inspection reports, PRA model results (e.g., the output of the agency’s SPAR models [9]), and research and development efforts (including activities outside as well as within the nuclear industry) are important. Further, as indicated in the previous section, the volume and variety of this information is likely to grow. The fundamental KE challenge is how to enable users to efficiently access and use this information.

This challenge is clearly not unique to the risk arena – significant development efforts are underway in the commercial information technology sector and a number of products are already available, as discussed later in this report. It is of interest to determine if these efforts and products are sufficiently mature to efficiently accommodate the typical characteristics of risk problems, including:

- A systems viewpoint. A risk-informed decision generally needs to consider the performance of the system as a whole, and should not focus exclusively on one aspect of the problem. In some cases, analysts and decision makers may need to cope with situations where potentially important aspects are poorly understood.
- Diverse and implicit sources of information. The basis for a PRA model may reside in a wide range of sources (e.g., licensing basis information, operating experience, licensee submittals) that may or may not be explicitly referenced in the PRA model's documentation. Understanding of this basis can be key to appropriate use of the model's results and insights.
- Involvement of multiple disciplines. Dealing with the system as a whole typically requires input from a wide range of technical disciplines. These disciplines have, in addition to their unique bodies of knowledge, their own technical cultures which affect how they create and consume information, and how they view and deal with uncertainties.
- Problem complexity. A risk problem may require consideration of a large number of disparate scenarios. For example, both scenarios triggered by low-probability/high-consequence natural disasters that overwhelm facility defenses and scenarios involving chains of more likely but also more independent events could be important to a facility's risk profile.
- Treatment of rare events. Risk assessments and risk-informed decision making often deal with rare, beyond design basis events. In some situations, analysts and decision makers need to deal with novel designs and even design principles. In situations where direct experiential data are sparse, modeling (including modeling assumptions) plays a fundamental role and it is critical that modeling details be adequately understood.
- Addressing details. Risk-significant scenarios can arise from unique, plant-specific design and operational features that lead to subtle dependencies between potential failure events. Changes in relatively small details (e.g., the routing of a particular set of electrical cables) can impact a risk study's results and insights.
- Involving a broad user base. Within the NRC, risk information is being used by decision makers with a broad range of technical backgrounds and exposure to risk concepts. With the ever-increasing scope of risk-informed decision making applications, the breadth of this user base is also likely to grow.

These characteristics call for a KE approach that enables a wide range of users to draw inferences across a very wide, yet very technical set of information. Further, as a practical matter, since the implementation of such an approach would likely require substantial involvement by a wide range of subject matter experts (e.g., to provide word/phrase associations and search heuristics), an additional challenge involves the efficient use of such experts.

2.2.2 An Example

To illustrate one type of risk-related challenge that might be addressed by improved KE tools, consider the flooding of the Blayais nuclear power plant (a four-unit site) in December 1999. As discussed by Vial, Rebour, and Perrin [16], that event involved a storm that caused a loss of offsite power to Units 2 and 4, followed by a flooding-induced loss of Unit 1 essential service water (Train A) and of the low-head safety injection and containment spray system pumps for Units 1 and 2, as well as flooding of a number of areas of Units 1 and 2. The beyond-design basis flooding involved the overtopping of a protective dyke from the combined effects of storm surge and wind-driven waves. According to Ref. 16, weaknesses in the site's flooding protection included:

- Lack of consideration of the extreme meteorological conditions experienced in the plant design;
- Loss of site accessibility due to area flooding (affecting arrival of additional equipment and staff);
- Lack of consideration of the simultaneous impact on multiple units;
- Problems in promptly detecting flooding of key plant rooms;
- Problems in managing the release of water from flooded plant areas.

These weaknesses are similar (and in some cases, identical) to those highlighted by the Fukushima Dai-ichi accident, some 11 years after the Blayais event.

The Blayais event is now widely acknowledged as an important indicator of the potential importance of external flooding [17]. However, a review of the conference programs for the Probabilistic Safety Assessment and Management (PSAM) and Probabilistic Safety Assessment (PSA) conferences held after Blayais and prior to Fukushima shows considerable interest in the treatment of internal flooding (especially in the 2008-2011 time period), but little activity regarding external flooding. (One of the few exceptions is a 2011 paper by Ferrante, et al. on dam failure frequencies [18].) Note also that the 2009 version of the American Society of Mechanical Engineers/American Nuclear Society (ASME/ANS) PRA standard [19] included a requirement that, consistent with pre-Blayais NRC guidance on the treatment of external floods in the Individual Plant Examination of External Events program [20], allows the screening of a non-seismic external event if the design basis for that event meets deterministic criteria provided in the NRC's 1975 Standard Review Plan [21]. The ASME/ANS standard also provided the following text prior to its requirements for external flooding PRA: "These [external flooding PRA] approaches, based on a combination of using of the recurrence intervals for the design-basis floods and analyzing the effectiveness of mitigation measure to prevent core damage, have usually shown that the contribution to CDF [core damage frequency] is insignificant." The 2013 version of the ASME/ANS standard [22] has some revisions intended to address the potential for premature screening of external events (including external floods).

Despite the Fukushima-spurred attention to external flooding, discussions at an international workshop on multi-unit risk revealed that as late as November, 2014, prominent members of the PRA community were unaware of the Blayais event. It can be seen that, from both PRA-model building and risk-informed decision support perspectives, the relevant KE challenge is how to better ensure that important lessons from key events (which may not be widely recognized as "key" at the time of their occurrence) are brought into risk assessment and risk management activities without requiring the occurrence of an accident or even a severe condition (e.g., the flooding of the Fort Calhoun site in 2011 [23]).

2.3 THE PROMISE OF NEW TECHNOLOGIES

The NRC, as with any organization that deals with large volumes of information, has a number of information technology systems and associated activities aimed at: (1) electronically capturing information important for the agency's decision making efforts, and (2) making the resulting information base accessible to the staff. In addition to the NRC's official recordkeeping system (ADAMS), staff can access information through a variety of tools, including the agency's website and staff-created Sharepoint sites. Users can employ a variety of search tools[5] and other aids (e.g., hyperlinks, file structures, citations and reference lists, document tables of content and indices) to find relevant files (e.g., text documents, spreadsheets, databases, images, computer codes and models) and specific pieces of information in these files.

Within the information technology industry, advances continue in improving the access to and use of information. One of the most widely publicized activities was highlighted on January 14, 2011, when an IBM-developed computer system called "Watson" defeated two human experts on the television quiz show "Jeopardy!." The central problem in Jeopardy! is for a player to, when presented with a clue in the form of an answer to an unspecified question:

1. activate a buzzer (which announces the player knows the question) before the other players; and
2. state the correct question.

Additional problems for players include the selection of the next clue to reveal (this involves a search strategy aimed at identifying high-value clues), and deciding on the size of the wager to risk in the game's final round. To win the game, a player must be quick as well as accurate. Moreover, since incorrect responses are penalized, the player must be able to assess his/her/its confidence in his/her/its candidate response.

The technical challenges posed by Jeopardy! are large and numerous. They include the breadth of potential topics; the volume, form, and trustworthiness of potentially relevant information; and the complexities (e.g., ambiguity, context-dependence, implicitness, and non-uniqueness) of natural language. (Table 1 provides examples of these complexities.) The ability of computer technology to meet these challenges was demonstrated by Watson's success. However, the Watson project, which was large and sustained (the project started in 2005 and involved a core

[5] The standard search tool in Microsoft Windows (which enables the identification of documents whose titles contain the desired keywords and the Microsoft Office documents that contain those keywords) and the file indexing capability provided by Adobe Acrobat (which enables the identification of relevant Portable Document Format – PDF – documents and even the particular keyword instances within those documents) are useful aids to users looking for documents on a local or network drive. Tools enabling searches for documents within ADAMS include ADAMS P8, Web-Based ADAMS, and the recently released ADAMS Enterprise Search. The performance of a number of these tools is discussed in Section 4 of this report.

team of about 20 researchers [3]), was aimed at a very specific problem with clearly established constraints. In a 2011 post-Jeopardy! workshop held for government agencies, IBM indicated that the core content analytics technology used in Watson was available in more practical, ready-to-use applications.

Both the private and public sectors have begun to use content analytics software to provide visibility into the amount of content that is being created, the nature of that content and how it is used. Based on responses to an NRC sources sought notice [24] it appears that there are several companies with content analytics capabilities and products. Following discussions with staff and contractors from the NRC's Office of the Chief Information Officer (OCIO), who were using an IBM product (ICA 2.2) to develop a content analytics tool to support in-house analyses of inspection reports by Office of Nuclear Reactor Regulation (NRR) staff, it was decided to focus project activities (in the project's first line of effort) on the evaluation of content analytics technology, as represented by ICA 2.2.

Table 1. Examples of Natural Language Challenges Arising in Jeopardy!

Jeopardy! Clue	"Its largest airport is named for a World War II hero; its second largest, for a World War II battle."
Clue Category	US Cities
Technical Challenge	Example
Ambiguity	"US" (vs. "us" vs. "U.S.")
Context-dependence	"US" (interpretation depends on the following "cities")
Implicitness	- "is named" is missing from the second clause - Indirect indication of two airports plus the fact that this clues arises in a television show (with an expectation of public knowledge) implies a large, famous city
Non-uniqueness	Clue could have been written in many ways (e.g., "named after" instead of "named for," "ace" instead of "hero").

2.4 CONTENT ANALYTICS AND ICA 2.2

Referring to Webster's, the term "content analysis" is defined as the "detailed study and analysis of the manifest and latent content of various types of communication … in order to ascertain their meaning and probable effect" and the term "analytics" is defined as "the science of analysis" [25]. In the information technology world, where increasing amounts of resources are being spent to make better use of large (and ever-increasing) amounts of unstructured information, "content analytics" methods and tools are being used to, among other things, help users improve their searches and enhance their "discovery" activities (i.e., activities to develop insights through exploration of the available information).

ICA 2.2 is one of a number of commercial products that provide content analytics capabilities. As discussed by Zhu et al. [5], ICA 2.2 consists of a number of major software components, including:

- "crawlers" which go through the documents in a pre-selected set (called a "corpus" by IBM) and extract document content;
- document processors which convert the unstructured text data generated by the crawlers into structured data using rules provided by text analytic "annotators"[6] (including standard annotators to do such things as identify the document language, perform a linguistics analysis, and identify text patterns using user-supplied rules, as well as any additional custom annotators);
- an indexer which prepares an optimized index of the processed document content (called a "text analytics collection," or "collection" for short) suitable for high-speed text mining and analysis; and
- a text miner application which provides the user interface enabling an analyst to search the corpus.

ICA 2.2 is a general product which can be customized to address the needs of specific problems. This customization process requires: (i) software engineers to configure the tool (e.g., to control how a crawler uses system resources and when it should be run) and develop desired annotators, and (ii) Subject Matter Experts (SMEs) to work with the software engineers to collaboratively define the search problem of interest and ensure efficient tool development.[7]

From an SME (and other end-user) perspective, most of the work performed by the software engineers is "behind the scenes." For example, the SME generally does not construct or perform a detailed review of the annotators produced by the software engineers. Rather, the SME uses a customized text mining application, also produced by the software engineers, which provides a number of tools supporting user searches and discovery. The principal tools are "facets," different windows on the corpus data, and their associated searches. Other tools can filter search results and support the development of statistics (e.g., matching document counts, frequencies of and trends in search phrase occurrences, and correlations of pairs of search phrase occurrences) and the visual identification of relationships between facets.

Figure 3 shows a screen shot of the final version of the ICA 2.2 tool as customized for the LTRP project. Figure 4 provides labels for a number of key elements in Figure 3. In both figures, the facets and subfacets available to the project (organized in a hierarchical fashion) appear on the left-hand side of the figure, and the screen displays the results of a search for the subfacet

---

[6] An "annotator" is a software tool used to annotate a document with information used to facilitate searches. Such information is extracted or deduced from the document content.

[7] Although not used in this project, ICA training is available for an SME desiring to work directly with the annotators. A "sand box" environment can be set up for the SME to fine tune his/her solutions prior to handing them off to the software engineer to deploy into production. Ideally this would shorten the development cycle while increasing the accuracy of the search algorithms.

“Multi-Unit” under the facet “Facility by Name.” (This subfacet contains the names of all plants on multi-unit sites, including Fitzpatrick and Hope Creek.)

Regarding the central construct of facets, Figure 5 is a screen shot showing some of the keywords[8] underlying the sub-facet “Multi-Unit Failure Phrases” (under the facet “Multi-Unit Events”).[9] As further discussed in Section 3, the keywords were developed by the software engineers, in coordination with the SMEs.

---

[8] A “keyword” is a string of characters (a text “token”) whose presence indicates that the document may have information relevant to the facet. Keywords need not be actually be words or phrases.

[9] Note that the figure shows that 159 documents in the corpus (which contains 333,512 documents) match the search criteria specified in the query window, i.e., the full set of keywords in the sub-facet “Multi-Unit Failure Phrases.” The frequency statistics shown in the figure indicate the number of documents containing the relevant keywords. Thus, for example, the phrase “shutdown of both units” appears in 26 of the 159 documents. The correlation statistics are not meaningful in this case since the search only addresses one facet. They are of more interest when considering searches involving pairs of facets.

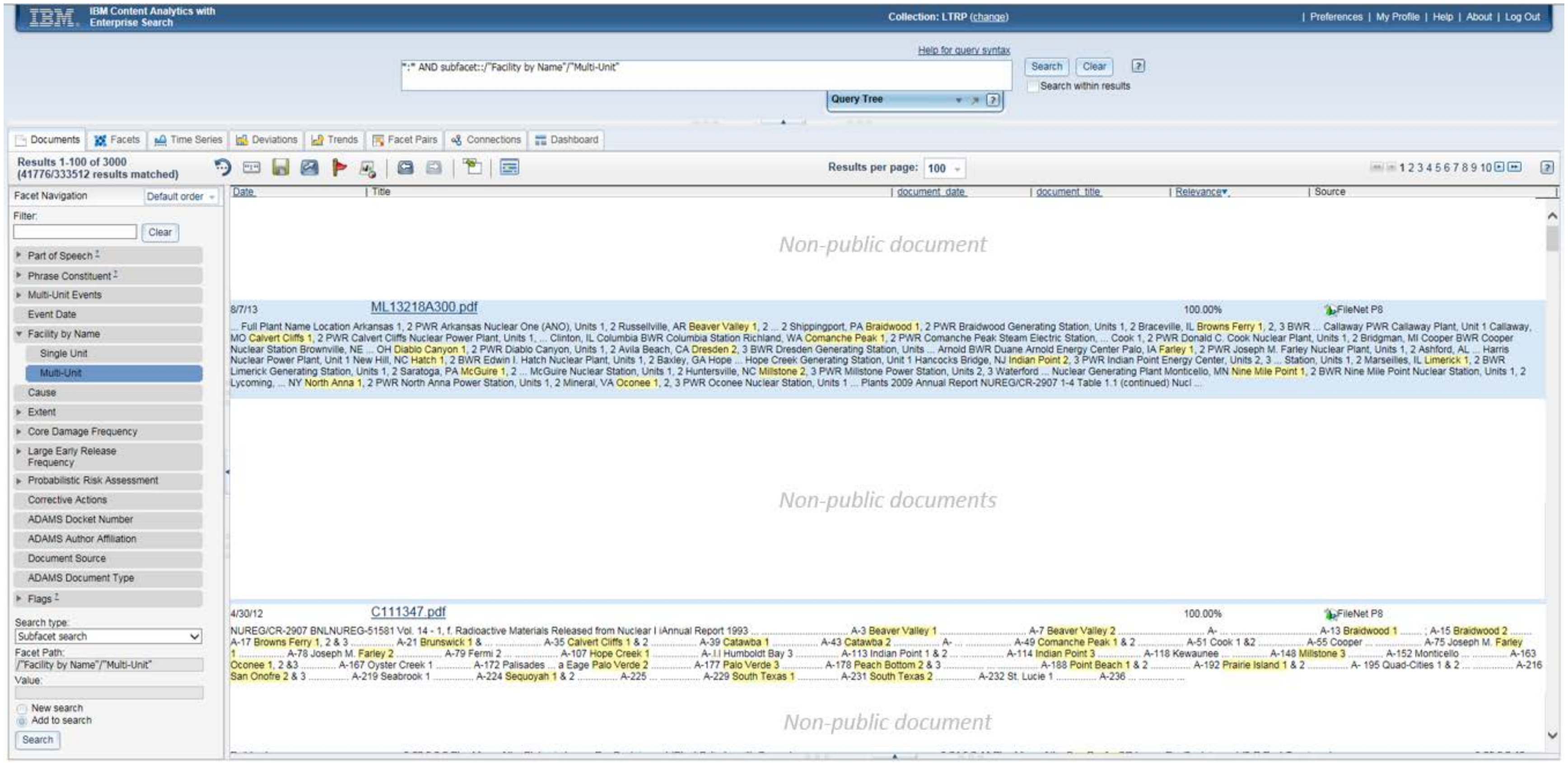

Figure 3. Sample ICA 2.2 window screen shot

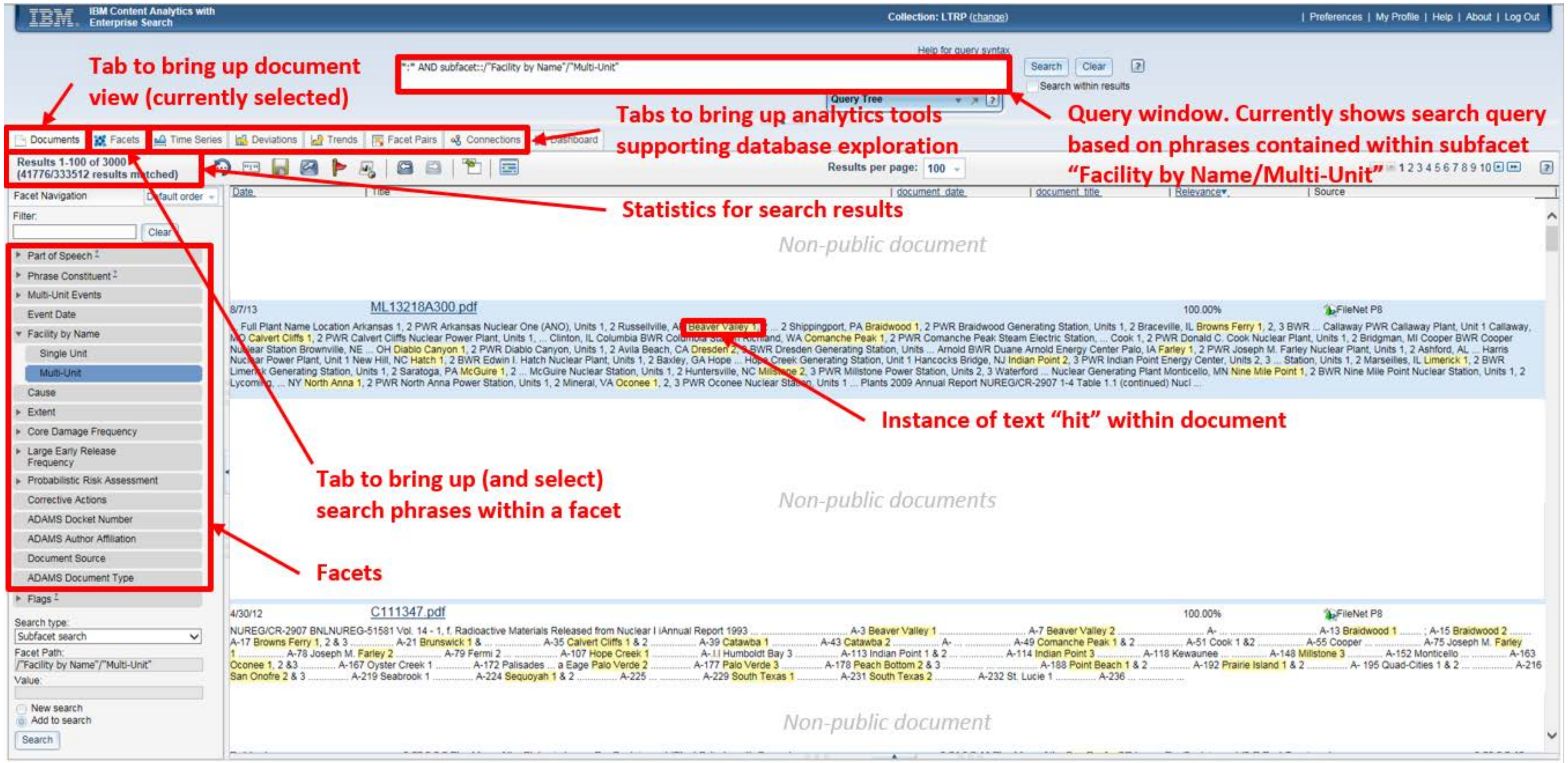

Figure 4. Sample ICA 2.2 window screen shot (labeled)

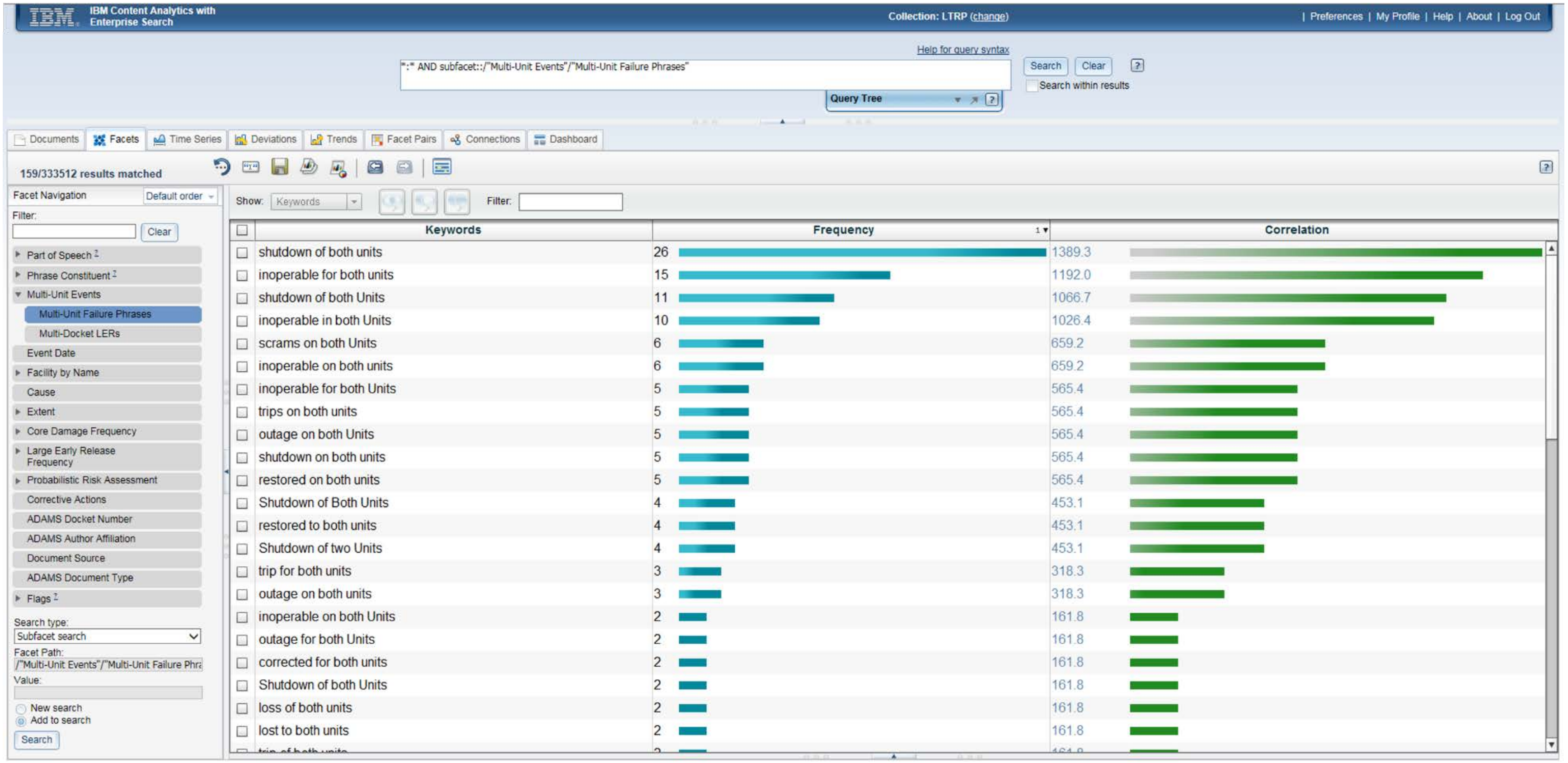

Figure 5. Subfacet "Multi-Unit Failure Phrases" keywords

3. OVERALL PROJECT DESCRIPTION

3.1 OBJECTIVE AND SCOPE

The overall objective of the project was to determine if additional agency effort to develop production-level KE tools aimed at supporting risk-informed applications could be worthwhile.

In keeping with the concept of an LTRP scoping study, the following scope limitations were employed.

- The evaluation was limited to the consideration of content analytics tools.
- The evaluation was performed using ICA 2.2. This tool is judged to be representative of the broad set of content analytics tools that are commercially available.
- The evaluation considered three applications ("use cases") described in the following section.
- The corpus, which provided the search space for ICA 2.2, was limited to the document types shown in Table 2. This corpus, which was finalized in late 2015, includes over 330,000 documents, represents a combination of selected documents in the ADAMS Main Library (which currently contains around 2 million documents, of which roughly half are publicly available) and a number of other documents.

Table 2. Project Corpus Contents

Description	Notes
Publicly available documents from NRC's ADAMS Main Library	Includes NRC staff (NUREG) and contractor (NUREG/CR) reports, staff papers to the Commission (SECY papers) and Commission Staff Requirements Memoranda (SRMs), License Amendment Requests, New Reactor Design Control Documents.
Final Safety Analysis Reports (FSARs)	Provide terminology and design-related information useful for event analysis
SPAR model documentation	Provides design-related information useful for event analysis (e.g., the size of the system involved), PRA results that can be compared with licensee/applicant results
Immediate Notifications	Documents notifying the NRC of events submitted per 10 CFR 50.72
Licensee Event Reports (LERs)	Documents notifying the NRC of events submitted per 10 CFR 50.73
Inspection reports	Staff reports from the NRC's Reactor Oversight Process (1999-present)
IPEs	Licensee submittals in response to Generic Letter 88-20
IPEEEs	Licensee submittals in response to Generic Letter 88-20, Supplement 4
Advisory Committee of Reactor Safeguards (ACRS) letter reports	1985-present
ACRS Meeting Transcripts	1999-present (subcommittee as well as full committee)

3.2 GENERAL APPROACH

3.2.1 Use Cases

The project employed a case-study approach in which ICA 2.2 was applied to two prototypical search problems ("use cases") faced by the staff, plus a research-oriented content analytics exploration of the corpus (see Table 3). The ability of ICA 2.2 to effectively and efficiently meet staff needs was assessed and compared with the capabilities of tools currently available to the staff.

The first use case involved the identification and characterization of operational events involving multiple nuclear power plants (henceforth called "multi-unit events"). This use case was selected because the identification of risk insights for multi-unit sites, particularly following the Fukushima accident, has been of an area of research focus; however, these types of events can be difficult to identify using current document search tools. Being event-oriented, this use case appeared to be similar in nature to IBM demonstrations of ICA 2.2 technology (e.g., an analysis of medical device failure data collected by the U.S. Food and Drug Administration's MedWatch Program and an analysis of product defect and recall data collected by the National Highway Traffic Safety Administration). For the same reason, it also seemed to be similar in nature to the previously mentioned in-house NRR project aimed at characterizing inspection findings and how these findings are used by staff. Thus, it was expected that ICA 2.2 would be capable in this application.

The second use case involved the characterization of current results from licensee PRA studies. This use case involved a search for documents containing current results and for key information within these documents. This use case was selected because information relating to the results from licensee developed PRAs is a useful benchmark for the current status of the nuclear fleet. While this information can be often found through a series of focused key word searches in ADAMS, it is an extremely manpower intensive and inefficient search process. Because this search differed in character from the first use case, the extent to which the ICA 2.2 tool could help the staff address the search problem, and the amount of effort required to provide this help, was less clear.

The first two use cases involve searches of the corpus for answers to specific questions. The third use case involved the use of ICA 2.2 in a more open-ended, discovery-oriented mode, in which the user explored the corpus for potentially useful insights. The aids provided by ICA 2.2 to support such explorations, distinguish it (and similar tools) from other search tools available to the staff.

Table 3. Project Use Cases

ID	Description	Notes
1	Search for multi-unit events	Supports characterization of past events involving multiple units at a site. This characterization could identify events that may need to be addressed in a site-wide PRA model.
2	Characterization of current licensee PRA results	Supports decision maker understanding of current risk levels and contributors. This activity addresses a common question raised by managers and external stakeholders.
E	Exploration of corpus	Uses ICA 2.2 in a discovery/exploration mode. This use case supports the project's evaluation of the tool when used in a non-direct search mode.

3.2.2 Technical Approach

In broad terms, Use Cases 1 and 2 involved a team of SMEs and software engineers performing four steps:

1. Specify search problem
2. Develop customized, problem-specific search tool using ICA 2.2
3. Test and refine problem-specific search tool
4. Demonstrate final problem-specific search tool and compare with alternate approaches

Use Case E involved a single SME exercising the customized ICA 2.2 tool developed for the first two uses cases "as-is" (i.e., without further modification).

3.2.2.1 Step 1 – Specify Search Problem

The first step involved discussions between the SMEs and software engineers to, as precisely as possible, define the problem of interest. More specifically, it involved the development of a specification of the search objectives and search space (i.e., the corpus, as discussed in Section 3.1). The initial specification of the search objectives involved the proposal of a broadly worded statement of the use case (see Table 3) and subsequent team discussions. These discussions helped the SMEs gain an appreciation of the features and capabilities of ICA 2.2 and the software engineers better understand the purpose of the use case, relevant reactor systems terminology, and characteristic text strings ("tokens") that could be useful in developing the search tool.

Over the course of the project, as discussed in Section 7, it became clear that the SMEs' initial understanding of the ICA 2.2 tool capabilities were overly optimistic. For both Use Cases 1 and 2, the initial problem specifications were revised to ensure that the ultimate project purpose (technology evaluation) could be performed within the project's available resource constraints.

3.2.2.2 Step 2 – Develop Customized Search Tool

The second step involved the development of custom annotators for the ICA 2.2 tool based on problem-specific facets. From an end-user's perspective, the key activity was the development of custom facets (with associated subfacets) based on sets of phrases (called "keywords") which, when matched, would indicate a successful search instance (a "hit") related to the facet topic. The facet development was done by the software engineers, based on input (e.g., direct suggestions; examples of hits including highlighted, relevant text passages) from the SMEs.

Figure 5, which provides an example of keywords useful for Use Case 1, shows that the keywords include multiple variants (involving capitalization, prepositions, and counters) on basic phrases (e.g., "shutdown of both units"). These variants (all of which can be found in the corpus) are intuitive to human readers but must be provided explicitly to the computer to enable machine-based recognition.

It can be seen that even for simple concepts, keyword entry can be laborious. Perhaps more importantly from an overall technology evaluation perspective, ICA 2.2 does not appear to have natural language tools (e.g., pre-developed sets of phrases or queries) to aid the construction of facets addressing more complex notions (e.g., what exactly was the cause of an accident).[10] A large custom software programming effort would be needed to develop such facets, should a more automatic solution be desired.

3.2.2.3 Step 3 – Test and Refine Customized Search Tool

The third step involved the application of the customized ICA 2.2 tool by the SMEs to the search problem and the identification of potential problems (principally the failure to identify known sources of information in the corpus and the identification of an excessive number of undesired search results – "false positives"). Following discussions with the software engineers, the latter developed refinements to address issues judged to be important for the purpose of a technology evaluation. As discussed in Sections 3 and 4, the refinements ranged from changes in search strategy, through the development of new facets, to modifications of keywords in a given facet. In some cases, it was determined that the corpus did not contain key documents, and the corpus was updated.

Steps 1-3 were performed iteratively as the project team gained experience with the applications. The process required coordination and understanding between the SMEs and software engineers.

3.2.2.4 Step 4 – Demonstrate Customized Search Tool and Compare with Alternate Approaches

The fourth step was performed when the team agreed the customized tool had been developed to the point where it could support a fair assessment. Use Cases 1 and 2 also used search tools currently available to the general NRC staff, as well as ICA 2.2.

---

[10] Note that cause identification is a subjective process, being dependent on the perspective of an analyst, as well as the needs of the problem supported by the cause analysis.

4. USE CASE 1 – MULTI-UNIT SEARCH

As argued by Fleming in 2005 [26] and illustrated by the March, 2011 reactor accidents at the Fukushima Dai-ichi nuclear power plant, events involving multiple reactor units at a single site can be important contributors to site risk. There are numerous technical challenges in assessing these contributions. NRC/RES is currently engaged in a full-scope, Level 3 PRA study intended to address all relevant site radiological sources (including the spent fuel pool and dry cask storage), internal and external initiating event hazards, and modes of operation for a two-unit, Westinghouse four-loop pressurized water reactor station with a large, dry containment [27,28]. The technical approach for addressing multi-unit (and, more generally, multisource) events is described in broad terms in the project's Technical Analysis Approach Plan [29]. To inform the modeling of such events, it is of interest to review past operational events to provide an indication of the likelihood and impact of these events, and of their salient features.

However, such a review, although straightforward in principle, can be extremely labor-intensive. Aids such as LERSearch (https://lersearch.inl.gov/LERSearchCriteria.aspx), ADAMS-related search tools (ADAMS P8, Web-Based ADAMS, and ADAMS Enterprise Search), and general-purpose search aids (e.g., indices for pdf files created using programs such as Adobe Acrobat) are helpful but: a) are not tailored to address the multi-unit problem, and b) do not necessarily provide access to a number of documents that might be useful.[11]

The question addressed by this use case was whether the use of ICA 2.2 can help analysts in identifying and characterizing interesting events. As compared with other tools (and direct manual search), can it reduce the effort required? Can it readily find interesting events not found by conventional means?

4.1 USE CASE 1 OBJECTIVE AND SCOPE

The specific objective of this use case was to evaluate the effectiveness and efficiency of ICA 2.2 in helping users identify and characterize past U.S. operational events involving multiple reactors. To limit staff and contractor resource requirements, and in keeping with the exploratory nature of LTRP projects, the following scope limitations were employed.

- The project corpus was limited to the document types shown in Table 2.
- The focus was on events involving a transient (an "initiating event" in the parlance of PRA[12]) at one or more units at a single site. The search did not exclude but was not

[11] For example, although LERSearch is an excellent tool, it does not provide access to the Licensee Event Report (LER) associated with the dual unit loss of offsite power (LOOP) at Turkey Point in 1992 (Hurricane Andrew), nor to the LER associated with the dual unit LOOP at Surry in 2011 (caused by a tornado). Additionally, since its scope is limited to LERs, it cannot provide access to other documents types, such as SECYs.

[12] See NUREG-2122 [30] for definitions of PRA-related terms.

aimed at identifying degraded conditions that could affect the response of multiple units at a site during an accident, or at identifying events/conditions affecting multiple sites.
- The events were characterized in terms of the event date, facility involved, event extent, and event cause.[13]

4.2 USE CASE 1 TECHNICAL CHALLENGES

On the surface, it might seem that a search for multi-unit initiating events should be relatively straightforward. After all, surely a human analyst, upon reading an event summary, can readily determine whether that event involved initiating events at multiple units or not. However, there are two problems with this view. First, there is an enormous number of event reports to review. (For example, the project's corpus contains nearly 55,000 LERs covering the period 1980-2014.) Second, although determining whether an event involved multiple units is straightforward,[14] the event descriptions can require more careful reading to determine whether the event involved an initiating event or a degraded condition (i.e., whether it involved an event that triggered an accident, or an event that weakened the plant but did not trigger an accident).

Although the use of computer-aided searches can address the challenge of document volume, the unstructured form of the event reports presents a significant challenge to machine-based approaches. For example, Table 4 provides a list of multi-unit events identified as precursors[15] by the NRC's Accident Sequence Precursor (ASP) Program. Table 5 provides a list of additional events identified by Schroer [31] that involved multiple units but were not determined to be accident precursors. The last column in both tables contains key phrases from the associated LERs indicating that the event involved initiating events at multiple units. It can be seen that the phrases are not standardized. Moreover, sometimes the effects on different units are described in different places in the LER.

---

[13] As discussed in Ref. 6, it's worth noting that the original project plans called for the identification a broader range of event characteristics. Early team discussions led to the conclusion that the four characteristics chosen were sufficient for the purposes of this project.

[14] NRC Form 366 used for LERs has a box indicating if the event involved other units. (See the highlighted portion of Figure 6.)

[15] Per SECY-15-0124 [9], a precursor is defined to be an event with a conditional core damage probability (CCDP) or a change in core damage probability ("delta CDP" or ΔCDP) greater than or equal to 1×10^{-6}.

NRC FORM 366 (10-2011) U.S. NUCLEAR REGULATORY COMMISSION	APPROVED BY OMB: NO. 3150-0104 EXPIRES 10/31/2013
LICENSEE EVENT REPORT (LER) (See reverse for required number of digits/characters for each block)	Estimated burden per response to comply with this mandatory collection request: 80 hours. Reported lessons learned are incorporated into the licensing process and fed back to industry. Send comments regarding burden estimate to the FOIA/Privacy Section (T-5 F53), U.S. Nuclear Regulatory Commission, Washington, DC 20555-0001, or by internet e-mail to infocollects.resource@nrc.gov, and to the Desk Officer, Office of Information and Regulatory Affairs, NEOB-10202, (3150-0104), Office of Management and Budget, Washington, DC 20503. If a means used to impose an information collection does not display a currently valid OMB control number, the NRC may not conduct or sponsor, and a person is not required to respond to, the information collection.

1. FACILITY NAME	2. DOCKET NUMBER	3. PAGE
Surry Power Station, Unit 1	05000 - 280	1 OF 6

4. TITLE
Reactor Trip on Both Units Due to Loss of Offsite Power

5. EVENT DATE			6. LER NUMBER			7. REPORT DATE			8. OTHER FACILITIES INVOLVED	
MONTH	DAY	YEAR	YEAR	SEQUENTIAL NUMBER	REV NO.	MONTH	DAY	YEAR	FACILITY NAME	DOCKET NUMBER
									Surry Unit 2	05000 - 281
04	16	11	2011 -	001 -	00	06	14	11	FACILITY NAME	DOCKET NUMBER 05000

9. OPERATING MODE: N

10. POWER LEVEL: 100%

11. THIS REPORT IS SUBMITTED PURSUANT TO THE REQUIREMENTS OF 10 CFR §: *(Check all that apply)*

☐ 20.2201(b)	☐ 20.2203(a)(3)(i)	☐ 50.73(a)(2)(i)(C)	☐ 50.73(a)(2)(vii)
☐ 20.2201(d)	☐ 20.2203(a)(3)(ii)	☐ 50.73(a)(2)(ii)(A)	☐ 50.73(a)(2)(viii)(A)
☐ 20.2203(a)(1)	☐ 20.2203(a)(4)	☐ 50.73(a)(2)(ii)(B)	☐ 50.73(a)(2)(viii)(B)
☐ 20.2203(a)(2)(i)	☐ 50.36(c)(1)(i)(A)	☐ 50.73(a)(2)(iii)	☐ 50.73(a)(2)(ix)(A)
☐ 20.2203(a)(2)(ii)	☐ 50.36(c)(1)(ii)(A)	☒ 50.73(a)(2)(iv)(A)	☐ 50.73(a)(2)(x)
☐ 20.2203(a)(2)(iii)	☐ 50.36(c)(2)	☐ 50.73(a)(2)(v)(A)	☒ 73.71(a)(4)
☐ 20.2203(a)(2)(iv)	☐ 50.46(a)(3)(ii)	☐ 50.73(a)(2)(v)(B)	☐ 73.71(a)(5)
☐ 20.2203(a)(2)(v)	☐ 50.73(a)(2)(i)(A)	☐ 50.73(a)(2)(v)(C)	☐ OTHER Specify in Abstract below or in NRC Form 366A
☐ 20.2203(a)(2)(vi)	☒ 50.73(a)(2)(i)(B)	☐ 50.73(a)(2)(v)(D)	

12. LICENSEE CONTACT FOR THIS LER

FACILITY NAME	TELEPHONE NUMBER *(Include Area Code)*
B. L. Stanley, Director Safety and Licensing	(757) 365-2003

13. COMPLETE ONE LINE FOR EACH COMPONENT FAILURE DESCRIBED IN THIS REPORT

CAUSE	SYSTEM	COMPONENT	MANU-FACTURER	REPORTABLE TO EPIX	CAUSE	SYSTEM	COMPONENT	MANU-FACTURER	REPORTABLE TO EPIX
C	FK	BU	various	Y					

14. SUPPLEMENTAL REPORT EXPECTED		15. EXPECTED SUBMISSION DATE	MONTH	DAY	YEAR
☐ YES *(If yes, complete 15. EXPECTED SUBMISSION DATE)*	☒ NO				

ABSTRACT *(Limit to 1400 spaces, i.e., approximately 15 single-spaced typewritten lines)*

On April 16, 2011, at 1849 hours, with Surry Power Station Unit 1 at 100% reactor power and Unit 2 at 98.3% reactor power, an automatic reactor trip occurred on Unit 1 and on Unit 2 due to the loss of offsite power resulting from damage inflicted in the switchyard from a tornado. All automatic safety systems, including Emergency Diesel Generators, performed as designed. The loss of offsite power resulted in violating several Technical Specifications including unavailability of independent offsite power. Following the unit trips, Pressurizer/Pressurizer spray temperature difference and Pressurizer heatup rate were also exceeded. A Notification of Unusual Event was declared at 1855 hours due to loss of offsite power to both emergency busses on both units. This event is reportable pursuant to 10CFR50.73(a)(2)(iv)(A) since the event resulted in automatic actuation of reactor protection systems, 10CFR50.73(a)(2)(i)(B) for operation or condition prohibited by Technical Specifications, and 10CFR73.71(a)(4) for delayed implementation of compensatory measures for loss of power to one source of surveillance equipment.

Figure 6. Sample LER with highlighted multi-unit field

Table 4. Multi-Unit Precursor Events with Indicative Phrases (Page 1 of 2)[a]

Date	Site	Type	LER	Indicative Phrase(s)
6/22/82	Quad Cities	LOOP[b]	254/82-012	**Separated text, requires inference**: "Unit Two reactor tripped" AND "Due to the degraded mode of the Unit One emergency AC power system, a Generating Station Emergency Plan Unusual Event was declared". **Also could infer from**: "Unit 1/2 Diesel Generator tripped".
8/11/83	Salem	LOOP	272/83-033, 034	**Direct statements**: "both Salem units tripped", "Salem Units 1 and 2 Reactor Trips"
7/26/84	Susquehanna	SBO during test	388/84-013	**Separated text, requires inference**: "Unit 2 operating" AND "This resulted in a scram" AND "Unit 1 entered an LCO".
5/17/85	Turkey Point	LOOP	251/85-011	**Direct statement**: "An Unusual Event was declared for both Units 3 and 4"
7/23/87	Calvert Cliffs	LOOP	317/87-012	**Direct statement**: "resulting in both reactors tripping on loss of load"
3/20/90	Vogtle	LOOP	424/90-006, 425/90-002	**Direct statement**: "tripped Unit 1 RAT A and Unit 2 RAT B" **Also could infer from**: Unit 1 LER (424/90-006) "further description of the Unit 2 response to this event is provided in LER 50-425/1990-002" OR Unit 2 LER (425/90-002) "See Licensee Event Report 50-424/1990-006 for a discussion of the resulting effect on Unit 1".
8/24/92	Turkey Point	LOOP (Hurricane Andrew)	250/92-SO1	Missing LER. **Separated text requiring inference** in NUREG-1145, Vol. 9 (NRC Annual Report for 1992): (heading) "Turkey Point" AND "On August 24, 992, Class 4 Hurricane Andrew hit south Florida" AND "they eye of the storm passed over the Turkey Point site" AND "the licensee brought Units 3 and 4 to a' "hot shutdown" condition". Note that there are OCR errors that could affect a search.
12/31/92	Sequoyah	LOOP	327/92-027	**Direct statement**: "both units received a reactor trip signal"
10/12/93	Beaver Valley	LOOP	334/93-013	**Direct statements**: (title) "Dual Unit Loss of Offsite Power", "loss of offsite power to Units 1 and 2"
6/28/96	LaSalle	Reactor trip (fouled cooling water)	373/96-007, 008	**Direct statements**: (title) "Dual Unit Shutdown", "the units were shutdown"
6/29/96	Prairie Island	Loss of power to safeguards buses	282/96-012	**Direct statements**: (title) "Reactor Trips of Both Units", "both reactors tripped"
8/14/03	Fermi	LOOP (Northeast blackout)	341/03-002	N/A
8/14/03	Nine Mile Point	LOOP (Northeast blackout)	220/03-002, 410/03-002	None (even between NMP1 and NMP2)

[a]Sources: NUREG/CR-2497 [33], NUREG/CR-3491 [34], and the NUREG/CR-4674 series of reports (e.g., [35]), SECY papers on ASP program [9,36-42]

Table 4. Multi-Unit Precursor Events with Indicative Phrases (Page 2 of 2)[a]

Date	Site	Type	LER	Indicative Phrase(s)
8/14/03	Fitzpatrick	LOOP (Northeast blackout)	333/03-001	None
8/14/03	Ginna	LOOP (Northeast blackout)	244/03-002	N/A
8/14/03	Indian Point	LOOP (Northeast blackout)	247/03-005, 286/03-005	None (even between IP2 and IP3)
8/14/03	Perry	LOOP (Northeast blackout)	440/03-002	N/A
6/14/04	Palo Verde	LOOP	528/04-006	**Direct statements**: (title) “Three Unit trip”, “LOOP caused each reactor to trip”
9/25/04	St. Lucie	LOOP (Hurricane Jeanne)	335/04-004	**Direct statements**: (title) “Dual Unit Loss of Offsite Power”, “dual-unit LOOP occurred”
5/20/06	Catawba	LOOP	413/06-011	**Direct statements**: (title) “Reactor Trip of Both Catawba Units”, “both Catawba units tripped”
3/26/09	Sequoyah	partial LOOP	327/09-003	**Direct statements**: (title) “Units 1 and 2 Reactor Trip”, “Units 1 and 2 received an automatic reactor trip”
4/16/11	Surry	LOOP (tornado)	280/11-001	**Direct statements**: (title) “Reactor Trip on Both Units”, “reactor trip occurred on Unit 1 and on Unit 2”
4/27/11	Browns Ferry	LOOP (tornado)	259/11-001	**Direct statements**: (title) “Three-Unit Scram”, “automatic scrams of all three units”
8/23/11	North Anna	LOOP (earthquake)	338/11-003	**Direct statements**: (title) “Dual Unit Reactor Trip”, “automatic reactor trip of both Units”.
3/31/13	Arkansas Nuclear One	LOOP (U1) and trip (U2) (stator drop)	313/13-001	**Direct statements**: “structural damage to the ANO-1 and ANO-2 turbine buildings,” “loss of offsite power for ANO-1…ANO-2 automatically tripped off-line.”
4/17/13	LaSalle	LOOP (lightning)	373/13-002	**Direct statements**: “loss of offsite power and reactor scrams on both Units.”
5/25/14	Millstone	LOOP	336/14-006	**Direct statements**: (title) “Dual Unit Reactor Trip on Loss of Offsite Power,” “Both MPS2 and MPS3 experienced a turbine trip,” MPS declared an Unusual Event (UE) following the reactor trips”

[a]Sources: NUREG/CR-2497 [33], NUREG/CR-3491 [34], and the NUREG/CR-4674 series of reports (e.g., [35]), SECY papers on ASP program [9,36-42]

Table 5. Multi-Unit Non-Precursor Events with Indicative Phrases [31]

Date	Site	Type	LER	Indicative Phrase(s)
5/15/03	Comanche Peak	Dual unit trip	445/03-003	**Direct statements**: (title) "Reactor Trip on Units 1 and 2", "reactor trips on both units"
9/15/03	Peach Bottom	LOOP (lightning)	277/03-004	**Direct statements**: "Units 2 and 3 automatically scrammed"
9/18/03	Surry	Loss of buses (wind debris)	280/03-004	**Direct statements**: "a manual reactor trip was initiated on Unit 1 at 1728 and on Unit 2 at 1732"
2/15/07	Oconee	LOOP	269/07-001	**Direct statements**: "trip of Oconee Units 1 and 2"

As an even more challenging example, Figure 7 provides three sections of text from Appendix 3 of NUREG/CR-6738 [32], which provides a description of the well-known 1975 cable fire at Browns Ferry. It can be seen that the first reference to the Unit 1 reactor trip can be only be linked to Unit 1 by explanatory text on an earlier page and that the Unit reactor trip is mentioned on a separate page. Furthermore, the text highlights (resulting from a cursor drag attempting to grab relevant text) show that the actual text that would be reviewed by a software routine doesn't necessarily correspond to the text seen by a reader. For example, ICA 2.2 does not recognize the positional context of text appearing in a table. In the second example in Figure 7, the reader can easily separate the two entries of "At 12:5 1pm, operators manually scrammed the reactor from 704 MWe power level" and "It is not entirely clear why operators delayed the scram for 15 minutes after learning of the fire." In contrast, the literal text string as generated by optical character recognition (OCR) and provided to ICA 2.2 is "At 12:5 1pm, operators manually scrammed the It is not entirely clear why operators delayed the reactor from 704 MWe power level." The intervening text between "scrammed the" and "reactor" clearly presents a challenge to keyword-oriented searches.

4.3 USE CASE 1 APPROACH

The general approach for this use case followed the process described in Section 3.2.2.

The use case team was comprised of three SMEs and two software engineers. The first SME had, prior to this LTRP project, performed a manual search of LERs (reported for the period 2003-2013) in order to identify multi-unit events of potential interest for the previously mentioned, ongoing Level 3 PRA study (see Appendix B). The second SME had also performed a pre-LTRP project search of LERs (reported for the period 2000-2011). This was done as part of a University of Maryland Master's Thesis study aimed at developing and testing an event classification scheme aimed at addressing multi-unit dependencies in support of multi-unit PRA [31]. The third SME was the overall lead for this LTRP project. The lead software engineer was an expert with ICA 2.2, being responsible for, among other things, the development of ICA 2.2 applications for a range of NRC problems. The supporting software engineer was responsible for developing and testing facets based on input from the SMEs.

From Page A3-1:

Whether an event from the chain of events is typically included in a fire PRA is discussed where deemed appropriate. Lessons that may be gleaned from a specific event in the context of fire PRA are also provided. Unless otherwise noted, the event descriptions refer to events impacting Unit 1.

Time (rel. to ignition) (hr:min)	Event Description (Note 1)	Fire PRA Implications
Prior to the incident	The power cables for two 480 VAC boards from opposite safety trains were routed during construction, erroneously, inside the same cable	In a fire PRA, error in routing of cables is not taken into consideration. The actual discovery of such a construction error is rare. No other

From Page A3-5:

00:31	At 12:51pm, operators manually scrammed the reactor from 704 MWe power level.	It is not entirely clear why operators delayed the scram for 15 minutes after learning of the fire. In a fire PRA a scram immediately upon a report of an unsuppressed CSR fire would typically be assumed.

From Page A3-6:

--	On the Unit 2 control panel, operators noticed malfunctions on ECCS panel 9-3 and feedwater panel. Unit 2 RB fans were switched to low by the operators.	Typical fire PRAs consider the impact of a fire only on a single unit, even if that fire occurs in a common or shared plant area. In this case, the second unit also experienced some difficulties and was shut down. Simultaneous demand for multi-unit shutdown may introduce unique equipment demands that may not be covered by current fire PRAs.
00:40	At 1:00pm Unit 2 control room operators observed several annunciations regarding DC power and that one reactor protection M-G set had tripped. They proceeded to scram the Unit 2 reactor and initiate shutdown cooling. Unit 2 operator confirmed that all rods inserted.	

Figure 7. Excerpts from NUREG/CR-6738 [32] describing the Browns Ferry 1 and 2 fire

The first two SMEs were primarily involved in the first and third task steps: search problem specification, and test and refinement of the customized ICA 2.2 tool.[16] The third SME applied the final tool in a demonstration designed to develop the information needed to conduct the technology evaluation.

The demonstration was performed in two phases to exercise the ICA 2.2 tool in two different usage modes: informed search (i.e., a search where very specific information is known about the target documents), and basic search (i.e., a search where only general information is known about the target documents).[17] In all cases, the demonstration was limited to events involving reactor trips. This greatly reduced the number of LERs to be reviewed. (Note that of the 392 multi-unit LERs identified by Schroer [31], the large majority do not involve initiating events.)

4.3.1 Phase 1 – Informed Search

This phase, which involved searches of the corpus to find specific LERs for multi-unit events, was performed in two stages. The first stage, which helped the SME conducting the demonstration to become better acquainted with the ICA 2.2 tool, was aimed at finding the LERs for a 2011 dual-unit loss of offsite power (LOOP) at North Anna (caused by an earthquake) and a 2011 three-unit LOOP at Browns Ferry (caused by a tornado).

In general, the search process involved performing an initial search using selected facets, subfacets, and individual keywords. Progressive refinements of the search query, sometimes using additional user-supplied keywords to supplement the built-in keywords, eventually resulted in a manageable number of hits. At this point, a quick review of the contextual text supplied by ICA 2.2, or of the target documents, was usually sufficient to determine if the hits represented desired search results.

Table 6 shows how this stepwise search process was executed for the North Anna seismically-induced LOOP. Note that Step 1 used a built-in keyword phrase ("Licensee Event Reports - Idaho National Laboratory") contained in the facet "Document Source." Similarly, Step 2 used a built-in keyword phrase ("North Anna") contained in the subfacet "Multi-Unit" of the facet "Facility by Name." The last keyword, "earthquake," was added manually (recognizing that this was an informed search). The final number of documents resulting from Step 3 (i.e., 10) was judged small enough for manual review (to determine which hit corresponded to the desired LER).

---

[16] For example, recognizing that text references to a site containing multiple units would probably not be very effective in identifying multi-unit events, the subfacet "Multi-Docket LERs" was added to take advantage of the LER entries for "Other Units Affected" (see Figure 6).

[17] Exercise of ICA 2.2 in discovery mode is addressed in Section 6 of this report.

Table 6. ICA 2.2 Search Process for North Anna Seismically-Induced LOOP LER

Step	Incremental Query[a]	Hits	Time (s)[b]
0	"*:*"[c]	333,512	N/A
1	AND keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory"[d]	54,788	2.2
2	AND keyword::/"Facility by Name"/"Multi-Unit"/"North Anna"	987	2.4
3	AND "earthquake"	10	1.8

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.
[b]Approximate search execution times, based on hand timing.
[c]This initial query provided by the ICA 2.2 tool captures all of the documents in the project corpus.
[d]This query is intended to capture the LERs in the project corpus provided by the Idaho National Laboratory in support of the project.
[e]This query is intended to capture all LERs involving the North Anna plant (a two-unit site).
[f]This user-developed custom query is intended to capture all North Anna LERs using the keyword "earthquake."

Figure 8 shows a screen shot of the final ICA 2.2 result. The desired LER is the first document in the list of hits. Note that careful reading of the contextual text can help distinguish between the target LER and the fifth hit in Figure 8, which is for a related LER. However, at least in this case, it is actually easier to identify the target LER by retrieving both documents – using the hyperlinks provided in the ICA 2.2 interface – and doing a quick review.

For comparison purposes, informed searches for the North Anna and Browns Ferry LERs were also conducted using LERSearch and Adobe Acrobat. The latter search employed a manually-created index of the LERs included in the project corpus. Some of the general capabilities of the three tools are shown in Table 7.

The second stage of Phase 1 involved a search of the corpus for the LERs for all multi-unit initiating event precursors identified by the NRC's ASP Program (see Table 4). This stage used a user-constructed search query building on the keywords included in the ICA 2.2 tool (see Figure 9).[18] Note that the ICA 2.2 interface facilitates query construction: the contents of the query window, which show the latest search string (see Figure 4), can be copied and pasted into a conventional text editor, modified, and pasted back into the query window.

[18] This query was constructed to capture LER numbers in various formats (dck/yr-num, dckyrnum, and dckyearnum). The wildcard "*" was used to capture standardized filenames, e.g., 3382011003R00.pdf, which is the corpus file containing the target LER.

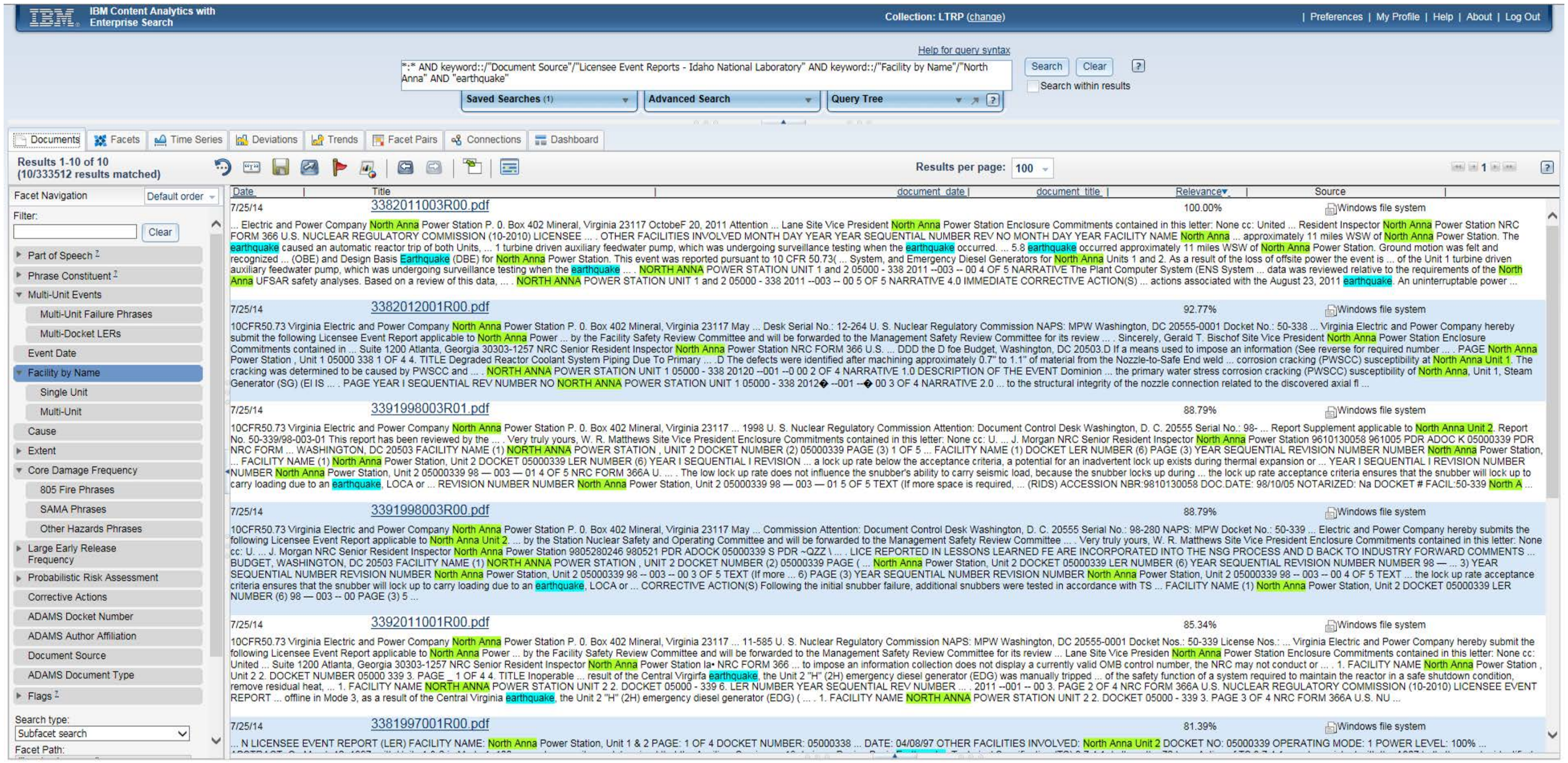

Figure 8. Results of informed search for North Anna seismically-induced LOOP LER

Table 7. Keyword Search Capabilities of ICA 2.2, LERSearch, and Adobe Acrobat

<table>
<tr><th>ICA</th><th>LERSearch</th><th>Adobe Acrobat</th></tr>
<tr><td>Exact match/AND
Synonyms (prefix ~)
Linguistic base (postfix ~)
Logical NOT (prefix -)
Wildcard indefinite length (*)
Wildcard single character (?)
Exact phrase (“”)
Facet value(s) (/fname/v1/…/vn/)
Boost relevance (postfix ^bvalue)
Fuzzy match (postfix ~fvalue)
Logical OR ((term 1 | term 2))
Proximity (WITHIN nword)
Ordered proximity (WITHIN nword INORDER)
Subset ((terms) ANY value)
Domain restriction, e.g.,
title:
site:
url:
link:
field: (defined for a collection)</td><td>Logical AND (AND, &&)
Logical OR (OR, “,”)
Logical NOT (NOT, prefix -)
Wildcard indefinite length (*)
Wildcard single character (?)
Exact phrase (“”)
Fuzzy match (word ~fvalue)
Proximity (“phrase” ~nword)
Range ([term1 TO term2], {term1 TO term2}</td><td>Exact match
Match any of the words
Match all of the words
Boolean query
Domain restriction
Date created
Date modified
Author
Title
Subject
Filename
Keywords
Bookmarks
Comments</td></tr>
</table>

“*:*” AND keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" AND (254/82-012 OR 25482012 OR 2541982012* OR 272/83-033 OR 27283033 OR 2721983033* OR 272/83-034 OR 27283034 OR 2721983034* OR 388/84-013 OR 38884013 OR 3881984013* OR 251/85-011 OR 25185011 OR 2511985011* OR 317/87-012 OR 31787012 OR 3171987012* OR 424/90-006 OR 42490006 OR 4241990006* OR 425/90-002 OR 42590002 OR 4251990002* OR 250/92-SO1 OR 25092SO1 OR 2501992SO1* OR 327/92-027 OR 32792027 OR 3271992027* OR 334/93-013 OR 33493013 OR 3341993013* OR 373/96-007 OR 37396007 OR 3731996007* OR 373/96-008 OR 37396008 OR 3731996008* OR 282/96-012 OR 28296012 OR 2821996012* OR 341/03-002 OR 34103002 OR 3412003002* OR 220/03-002 OR 22003002 OR 2202003002* OR 410/03-002 OR 41003002 OR 4102003002* OR 333/03-001 OR 33303001 OR 3332003001* OR 244/03-002 OR 24403002 OR 2442003002* OR 247/03-005 OR 24703005 OR 2472003005* OR 286/03-005 OR 28603005 OR 2862003005* OR 440/03-002 OR 44003002 OR 4402003002* OR 528/04-006 OR 52804006 OR 5282004006* OR 335/04-004 OR 33504004 OR 3352004004* OR 413/06-001 OR 41306001 OR 4132006001* OR 327/09-003 OR 32709003 OR 3272009003* OR 280/11-001 OR 28011001 OR 2802011001* OR 259/11-001 OR 25911001 OR 2592011001* OR 338/11-003 OR 33811003 OR 3382011003* OR 313/13-001 OR 31313001 OR 3132013001* OR 373/13-002 OR 37313002 OR 3732013002* OR 336/14-006 OR 33614006 OR 3362014006*)

Figure 9. Query for multi-unit precursor LERs

4.3.2 Phase 2 – Basic Search

This phase involved two separate searches for multi-unit initiating events that exercised the ICA 2.2 tool in a more exploratory mode, i.e., without prior knowledge regarding which specific events involved multiple units.

The first search focused on the LERs in the project corpus. After some exploratory searches to gain a sense of the number of potentially relevant documents in the corpus, a search was performed using relevant keywords provided under the “Document Source” and “Multi-Unit Events” facets. The former facet was used to focus on LERs; the latter facet was used to narrow in on events involving multi-unit reactor scrams/trips.[19] The search steps used to develop the results discussed in Section 4.4.2.1 are shown in Table 8.

The second search aimed at finding ASP-related SECY papers referring to multi-unit initiating events. This search was performed upon recognizing that recent papers have explicitly identified a number of precursors involving multiple units. The search was only aimed at identifying relevant SECY papers; the papers themselves typically provide the LER numbers for the events. The search steps used to develop the results discussed in Section 4.4.2.2 are shown in Table 9.

[19] During the search, it was determined that there were some potential problems with the characterization of documents in the corpus. For example, an early search for LERs with multiple plant docket entries – Box 8 in Figure 6 – came up with a hit for the River Bend Station (a single-unit plant), although the associated LER has no entry for Box 8. To date, the problems identified appear to lead to false positives that are eliminated during refined searches, and therefore don’t significantly affect the results of this demonstration analysis.

Table 8. ICA 2.2 “Basic Search” Process for Multi-Unit LERs

Step	Incremental Query[a]	Hits	Time (s)[b]
0	“*:*”[c]	333,512	N/A
1	AND (keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" OR keyword::/"Document Source"/"Licensee Event Reports - ADAMS")[d]	63,714	1.8
2	AND (keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000185,05000271" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000206,05000361,05000362" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000206,05000361" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000206,05000362" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000230,05000280" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000235,05000285" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000237,05000249" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000245,05000336,05000423" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000247,05000286" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000250,05000251" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000254,05000265" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000259,05000260,05000296" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000259,05000260" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000259,05000296" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000260,05000296" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000266,05000301" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000269,05000270,05000287" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000269,05000270" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000269,05000287" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000272,05000311" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000275,05000323" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000277,05000278" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000280,05000281" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000282,05000306" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000295,05000304" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000313,05000318" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000313,05000368" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000315,05000316" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000317,05000318" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000321,05000366" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000324,05000325" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000327,05000328" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000333,05000338" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000334,05000412" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000335,05000389" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000336,05000366" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000338,05000339" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000348,05000364" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000352,05000353" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000361,05000362" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000369,05000370" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000373,05000374" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000387,05000388" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000387,05000397" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000413,05000414" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000424,05000425" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000445,05000446" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000454,05000455" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000454,05000456" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000456,05000457" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000498,05000499" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000528,05000529,05000530" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000528,05000529" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000528,05000530" OR keyword::/"Multi-Unit Events"/"Multi-Docket LERs"/"05000529,05000530")[e]	8,335	9.0

Table 8. ICA 2.2 "Basic Search" Process for Multi-Unit LERs (cont.)

Step	Incremental Query[a]	Hits	Time (s)[b]
3	AND (keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"trips on both units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"trips of both Units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"tripped on both units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"trip of both units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"trip of both Units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"trip for both units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"scrams on both Units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"Trips on Both Units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"Trips of Both Units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"Trip on Both Units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"Trip for both units" OR keyword::/"Multi-Unit Events"/"Multi-Unit Failure Phrases"/"Trip On Both Units")	23	6.5

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.

[b]Approximate search execution times, based on hand timing.

[c]This initial query provided by the ICA 2.2 tool captures all of the documents in the project corpus.

[d]This query is intended to capture all of the LERs in the project corpus.

[e]This user-developed custom query is intended to capture all LERs involving multiple units.

[f]This user-developed custom query is intended to capture all multi-unit LERs involving reactor trips.

Table 9. ICA 2.2 "Basic Search" Process for ASP SECYs Reporting Multi-Unit Initiating Event Precursors

Step	Incremental Query[a]	Hits[b]
0	"*:*"[c]	333,512
1	AND keyword::/"Document Source"/"Commission SECY Paper - ADAMS"[d]	8,609
2	AND ASP AND CCDP[e]	67
3	AND "Dual unit" OR "dual-unit"[f]	5

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.

[b]The associated search execution times were very short (on the order of a few seconds).

[c]This initial query provided by the ICA 2.2 tool captures all of the documents in the project corpus.

[d]This query is intended to capture all of the documents profiled as SECY papers in the project corpus (including enclosures to the SECYs).

[e]This user-developed custom query is used to focus on SECY papers discussing the ASP program and results. The query was developed after determining (using the ICA 2.2 facet tab) that none of the SECY papers in the corpus used any of the built-in keywords in the "Multi-Unit Failure Phrases" subfacet.

[f]This user-developed custom query (not included in the tool's facets) was used based on knowledge of how recent ASP SECY papers had characterized multi-unit initiating events. A later test confirmed that a search omitting the "dual-unit" phrase misses one of the target SECYs.

4.4 USE CASE 1 RESULTS

4.4.1 Informed Search Results

4.4.1.1 Search for a Specific Multi-Unit LER

When searching for the specified LERs (for the North Anna seismically-induced LOOP of 8/23/11 and the Browns Ferry tornado-induced LOOP of 4/27/11), the customized ICA 2.2 tool was extremely quick and effective. Queries were executed in a few seconds,[20] and the overall process (including time to think about what keywords to use to narrow the search) took on the order of 1 to 2 minutes. Both searches resulted in a small number of hits; a quick read of the contextual text and/or candidate LERs was sufficient to identify the target document.

Similar searches using LERSearch were also very quick and effective. The searches (executed using the plant name field and keywords: “earthquake” for North Anna and “tornado” for Browns Ferry) were completed in a few seconds, resulted in a small number of hits, and provided descriptive document titles that made it easy to determine which hit was the desired LER. Also, unlike the ICA 2.2 results, the LERSearch results clearly indicated the event date.[21]

The search using Adobe Acrobat was somewhat slower (taking tens of seconds to execute) and less efficient (the contextual text did not always provide sufficient information to quickly identify the target document).

4.4.1.2 Search for a Specified Set of Multi-Unit LERs

Similar to the search for single LERs, when provided with the LER numbers associated with the events listed in Table 4, the customized ICA 2.2 tool was quite successful in identifying those LERs. The search required the development of a custom query (see Figure 9), but this relatively simple task only required a few minutes using the ICA 2.2 interface in combination with a standard text editor. The search query itself took about 1.5 minutes to execute.

As shown in Table 10, the search identified two false positives (these involved LERs that involved single-unit events but referenced the target LERs), and missed two events (because the LERs were not in the project corpus). However, the search process also helped identify an LER (for a LOOP at Catawba on 5/20/06) whose number was incorrectly entered in SECY-07-0176, Enclosure 2 [39].

---

[20] It should be recognized that all ICA 2.2 search times reported are appropriate for the project corpus. Larger databases, such as the ADAMS Main Library, will take longer to search.

[21] The ICA 2.2 interface provides a display field for the document date, but this date does not necessarily correspond to the event date. Further, as shown in Figure 8, many of the corpus documents do not have entries for the document date. As discussed in Ref. 6, it proved surprisingly difficult to reliably extract event dates from the LERs.

Table 10. Results of ICA 2.2 Informed Search for Multi-Unit LERs

LER	ASP Multi-Unit?	Notes
254/82-012R00	Y	Quad Cities 6/22/82 LOOP
272/83-033R00	Y	Salem 8/11/83 LOOP
272/83-034R00	Y	Salem 8/11/83 LOOP
388/84-013R00	Y	Susquehanna 7/26/84 SBO
251/85-011R00	Y	Turkey Point 5/17/85 LOOP[a]
317/87-012	Y	Calvert Cliffs 7/23/87 LOOP
311/89-013	N	False Positive – Salem 2 6/10/89 Loss of Main Condenser, refers to 8/11/83 event as a similar occurrence
424/90-006	Y	Vogtle 3/20/90 LOOP
425/90-002	Y	Vogtle 3/20/90 LOOP
250/92-SO1	Y	Miss – Turkey Point 8/24/92 LOOP (Hurricane Andrew). LER not included in corpus.
327/92-027	Y	Sequoyah 12/31/92 LOOP
334/93-013	Y	Beaver Valley 10/12/93 LOOP
272/94-007	N	False Positive – Salem 1 4/7/94 reactor trip, refers to grass intrusion reported for 8/11/83 event.
282/96-012	Y	Prairie Island 6/29/96 loss of power to safeguards buses
373/96-007	Y	LaSalle 6/28/96 reactor trip
373/96-008	Y	LaSalle 6/28/96 reactor trip
220/03-002R00/01	Y	8/14/03 Northeast Blackout – effect on Nine Mile Point 1
244/03-002R00	Y[b]	8/14/03 Northeast Blackout – effect on Ginna[b]
247/03-005R00	Y	8/14/03 Northeast Blackout – effect on Indian Point 2
286/03-005R00	Y	8/14/03 Northeast Blackout – effect on Indian Point 3
333/03-001R00	Y	8/14/03 Northeast Blackout – effect on Fitzpatrick
341/03-002R00/01	Y[b]	8/14/03 Northeast Blackout – effect on Fermi 2[b]
410/03-002R00	Y	8/14/03 Northeast Blackout – effect on Nine Mile Point 2
440/03-002R00/01	Y[b]	8/14/03 Northeast Blackout – effect on Perry[b]
335/04-004R00	Y	St Lucie 9/25/04 LOOP
528/04-006R00/01	Y	Palo Verde 6/14/04 LOOP
413/06-001R01[c]	Y	Catawba 5/20/06 LOOP
327/09-003R00/01	Y	Sequoyah 3/26/09 LOOP
259/11-001R00	Y	Browns Ferry 4/27/11 LOOP
280/11-001R00	Y	Surry 4/16/11 LOOP
338/11-003R00	Y	North Anna 8/23/11 LOOP
313/13-001R00/01	Y	Arkansas Nuclear One 3/31/13 stator drop
373/13-002R00/01/02	Y	LaSalle 4/17/13 LOOP
336/14-006	Y	Miss – Millstone 5/25/14 LOOP LER not included in corpus.

[a]Discovered through search process that LER number was originally mis-typed. NUREG/CR-4674, V2 (pp. E-5, E-61) provides correct LER number.

[b]Ginna, Fermi, and Perry are single units but were involved in a multi-site event.

[c]Discovered through search process that LER number provided in SECY-07-0176, Enclosure 2 (413/06-011) is incorrect. Correct number is as shown in table.

4.4.2 Basic Search Results

4.4.2.1 LER Search

Unlike the informed search, this search did not provide the customized ICA 2.2 tool with extremely specific information for multi-unit LERs of interest (i.e., the LER numbers). The search results, developed using a 3-step search process (see Table 8 for the specific queries and results), are shown in Table 11. For comparison purposes, the results of the informed search (see Section 4.4.1) and those of the earlier manual search performed by Schroer [31] (covering only LERs for the years 2000-2011) are also provided.

It can be seen that the search process shown in Table 8 identified only 4 of the 32 of the ASP multi-unit initiating event precursors occurring between 1982 and 2014. The search also provided 4 false positives, i.e., LERs for events that did not involve multiple units. (Most of these false positives involve an LER for a single unit event that refers to an LER for a multi-unit event.)

Note that Step 2 of the search identified 8,335 LERs. It seems likely that the target LERs (missed by the 3-step search) are included in the 8,335, and that better results could have been obtained by developing additional keywords based on the information in Table 4. However, such an effort would have been complex (especially for events where the indicative phrases are widely separated in the LER), and was not judged to be necessary for the purpose of this feasibility study. Indeed, the complexity of such a refinement is in itself an indicator of the level of the ICA 2.2 tool with respect to the needs of Use Case 1.

On the positive side, the customized ICA 2.2 tool identified 6 multi-unit initiating events not identified as precursors by the ASP program, and, perhaps more significantly, a dual unit precursor (a tornado-induced LOOP at Surry on 4/6/11) not identified by the manual search and not accessible through LERSearch.[22]

4.4.2.2 ASP SECY Search

Using the search process shown in Table 9, the customized ICA 2.2 was quick and effective at identifying all of the SECY papers (or their enclosures) that referred to dual unit events, i.e., SECY-07-0176 [38], SECY-10-0125 [39], SECY-12-0133 [40], SECY-13-0107 [41], and SECY-14-107 [42]. Of course, it did not identify a relevant SECY not in the corpus (SECY-15-0124 [9]), nor did it identify SECYs for which multi-unit events can only be inferred from tables containing entries indicating that different units were affected on the same date (i.e., SECY-05-0192 [36] and SECY-06-0208 [37]). Addressing this latter situation would likely require custom programming well beyond the scope or needs of this feasibility study.

[22] Note that the LER (280/11-001) is publicly available.

Table 11. Results of ICA 2.2 Searches for Multi-Unit LERs (Page 1 of 2)

LER	Multi-Unit	ASP	Manual[a]	ICA 2.2 (Informed)[b]	ICA 2.2 (Basic)[c]	Notes
254/82-012	X	X	N/A	X		Quad Cities 6/22/82 LOOP
272/83-033	X	X		X		Salem 8/11/83 LOOP
272/83-034	X	X		X		Salem 8/11/83 LOOP
388/84-013	X	X		X		Susquehanna 7/26/84 SBO
251/85-011	X	X		X		Turkey Point 5/17/85 LOOP[a]
369/85-034	X				X	McGuire 11/2/85 dual unit trip
306/86-002					X[d]	Prairie Island 2 5/20/86 (LER refers to previous trips on both units)
317/87-012	X	X		X		Calvert Cliffs 7/23/87 LOOP
311/89-013				X[d]		Salem 2 6/10/89 Loss of Main Condenser (refers to 8/11/83 LER)
424/90-006	X	X		X		Vogtle 3/20/90 LOOP
425/90-002	X	X		X		Vogtle 3/20/90 LOOP
338/92-007					X[d]	North Anna 3/6/92 missed surveillances
250/92-SO1	X	X				Turkey Point 8/24/92 LOOP (Hurricane Andrew) (LER not included in corpus)
327/92-027	X	X		X		Sequoyah 12/31/92 LOOP
334/93-013	X	X		X		Beaver Valley 10/12/93 LOOP
272/94-007				X[d]		Salem 1 4/7/94 reactor trip (refers to 8/11/83 LER)
275/94-016	X[d]				X[d]	Diablo Canyon 8/15/94 partial LOOP (no reactor trips)
445/95-002	X				X	Comanche Peak 5/5/95 LOOP
373/96-007	X	X		X		LaSalle 6/28/96 reactor trip
373/96-008	X	X		X		LaSalle 6/28/96 reactor trip
282/96-012	X	X		X	X	Prairie Island 6/29/96 loss of power to safeguards buses
275/96-012	X				X	Diablo Canyon 8/10/96 LOOP
275/96-013					X[d]	Diablo Canyon 8/10/96 outside Technical Specifications (refers to LOOP LER)
369/97-009	X				X	McGuire 9/6/97 dual unit trip

[a]"Definite" multi-unit events identified by Schroer [31] for the period 2000-2011

[b]Search performed with knowledge of event LER numbers (see Section 4.3.1)

[c]Search performed without knowledge event LER numbers (see Section 4.3.2)

[d]Not a multi-unit initiating event

Table 11. Results of ICA 2.2 Searches for Multi-Unit LERs (Page 2 of 2)

LER	Multi-Unit	ASP	Manual[a]	ICA 2.2 (Informed)[b]	ICA 2.2 (Basic)[c]	Notes
325/00-001	X[d]		X[d]			Brunswick 3/3/00 Unit 1 LOOP (no reactor trip), Unit 2 LCO[e], EDG[f] starts
275/01-001	X[d]		X[d]			Diablo Canyon 4/5/01 LOOP, EDG starts (no reactor trips)
445/03-003	X		X		X	Comanche Peak 5/15/03 LOOP
220/03-002	X	X		X		8/14/03 Northeast Blackout – effect on Nine Mile Point 1
244/03-002	X[g]	X		X		8/14/03 Northeast Blackout – effect on Ginna
247/03-005	X	X		X		8/14/03 Northeast Blackout – effect on Indian Point 2
286/03-005	X	X		X		8/14/03 Northeast Blackout – effect on Indian Point 3
333/03-001	X	X		X		8/14/03 Northeast Blackout – effect on Fitzpatrick
341/03-002	X[g]	X		X		8/14/03 Northeast Blackout – effect on Fermi 2
410/03-002	X	X		X		8/14/03 Northeast Blackout – effect on Nine Mile Point 2
440/03-002	X[g]	X		X		8/14/03 Northeast Blackout – effect on Perry
277/03-004	X		X			Peach Bottom 9/15/03 grid disturbance
280/03-004	X		X			Surry 9/18/03 hurricane debris
528/04-006	X	X	X	X		Palo Verde 6/14/04 LOOP
335/04-004	X	X		X		St Lucie 9/25/04 LOOP
296/04-002	X[d]		X[d]			Browns Ferry 11/23/04 Unit 3 scram, Unit 2 turbine speed perturbation
237/05-003	X[d]		X[d]			Dresden 6/23/05 declared LOOP (no trip)
352/06-001			X[d]		X[d]	Limerick 12/9/05 partial LOOP, EDG starts
413/06-001	X	X	X	X		Catawba 5/20/06 LOOP
269/07-001	X		X		X	Oconee 2/15/07 grid failure
327/09-003	X	X		X		Sequoyah 3/26/09 LOOP
275/11-003	X[d]		X[d]			Diablo Canyon 3/11/11 staff evacuation from intake structure
280/11-001	X	X		X	X	Surry 4/16/11 LOOP
259/11-001	X	X	X	X		Browns Ferry 4/27/11 LOOP
338/11-003	X	X	X	X	X	North Anna 8/23/11 LOOP
313/13-001	X	X	N/A	X		Arkansas Nuclear One 3/31/13 stator drop
373/13-002	X	X	N/A	X	X	LaSalle 4/17/13 LOOP
336/14-006	X	X	N/A			Millstone 5/25/14 LOOP (LER not included in corpus)

[a]"Definite" multi-unit events identified by Schroer [31] for the period 2000-2011

[b]Search performed with knowledge of event LER numbers (see Section 4.3.1)

[c]Search performed without knowledge event LER numbers (see Section 4.3.2)

[d]Not a multi-unit initiating event

[e]LCO = Limiting Condition of Operation

[f]EDG = Emergency Diesel Generator

[g]Ginna, Fermi, and Perry are single units but were involved in a multi-site event

4.4.3 Characterization of Multi-Unit Events

As discussed in Section 4.1, the objective of Use Case 1 was to evaluate the effectiveness and efficiency of ICA 2.2 in helping users identify and characterize past U.S. operational events involving multiple reactors. Sections 4.4.1 and 4.4.2 discuss the event identification portion of the use case. Regarding characterization, per the discussion in Section 4.1, the focus was on the Event Date, Facility Name, Event Extent, and Event Cause.

During the course of the use case, it became clear to the SMEs that the customized ICA 2.2 tool is not ideal for automatically generating this latter type of information. For example, the tool is well suited for determining how many LERs involve a particular facility. However, it is not as direct when trying to determine the involved facility for each LER that matches a specific search query. If the contextual text provided by the tool does not provide the desired information, the user has to download and review (a hopefully small number of) search-identified documents.

Assuming that the customized ICA 2.2 tool has already been used to identify multi-unit LERs, the ability of the tool to support the characterization of the associated events is as follows.

- ***Event Date.*** The event date can be often found in the contextual text, and sometimes is provided as a document property. (Figure 10 provides a screen shot showing both. Note that the figure highlights the ICA 2.2 button used to toggle the document properties display.) Otherwise, the user needs to download and review the target document.

- ***Facility Name.*** The facility name sometimes shows up in contextual text, but sometimes does not. (See Figure 11.) If not, the user needs to download and review the target document.

- ***Event Extent.*** As shown in Figure 12, selecting keywords from the “Extent”/“Plant Systems” subfacet will highlight affected systems in the contextual text, providing a quick visual indication as to whether the chosen systems were affected. However, a more holistic view of the event will generally require download and review of the target document.

- ***Event Cause.*** In principle, the keywords in the facet “Cause” can be used to highlight appropriate contextual text (similar to the treatment of event extent). In this feasibility study, the actual keyword list (see Figure 13) was left at a highly preliminary stage of development and so no example results are presented. However, it is useful to note the following.

    - The ICA 2.2 feature of highlighting keywords in the contextual text helps the user to determine if the event involved a particular cause. (For example, the contextual text in Figure 8 highlights the keyword “earthquake,” which was used in the search query).

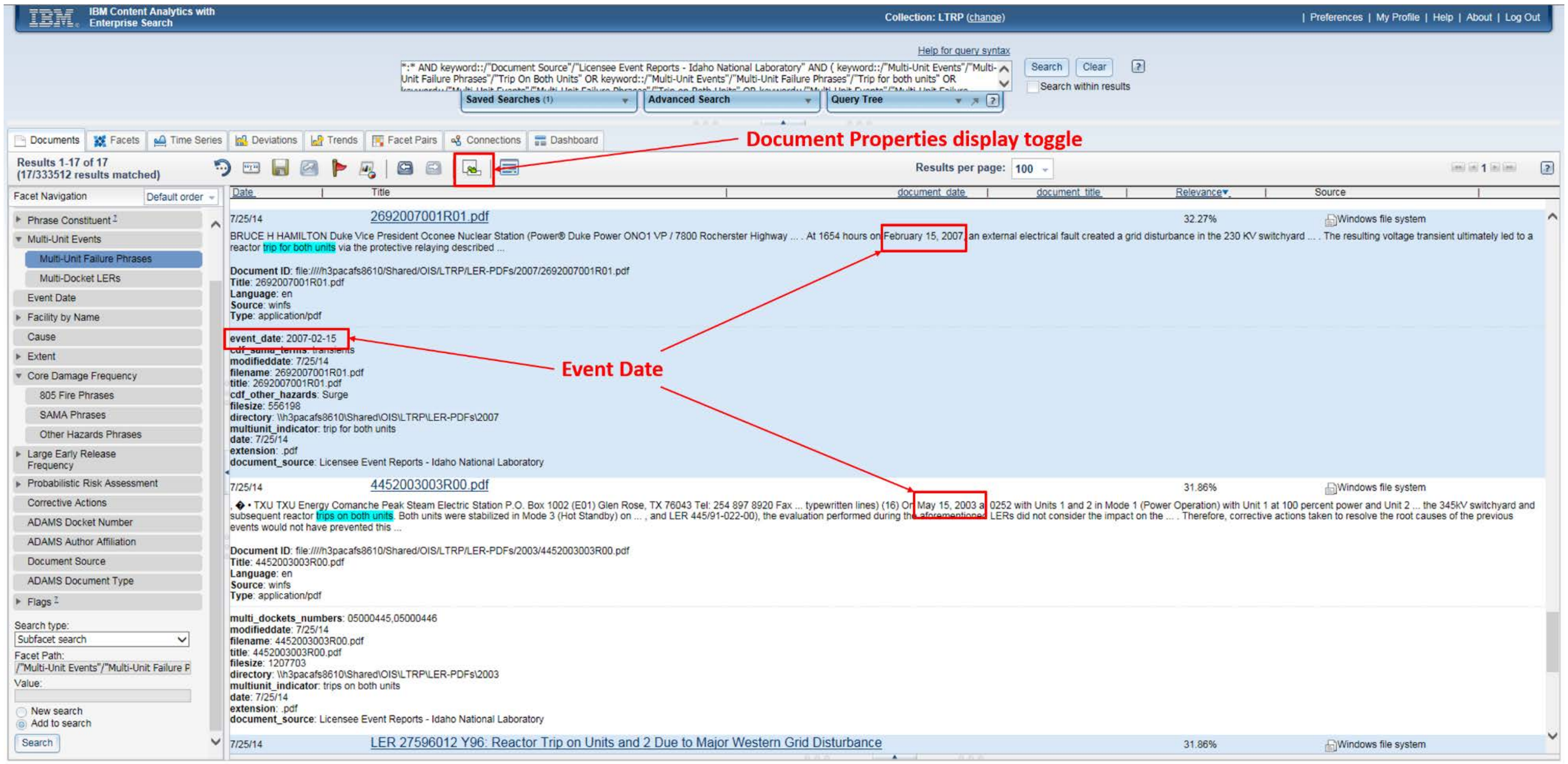

Figure 10. Event date

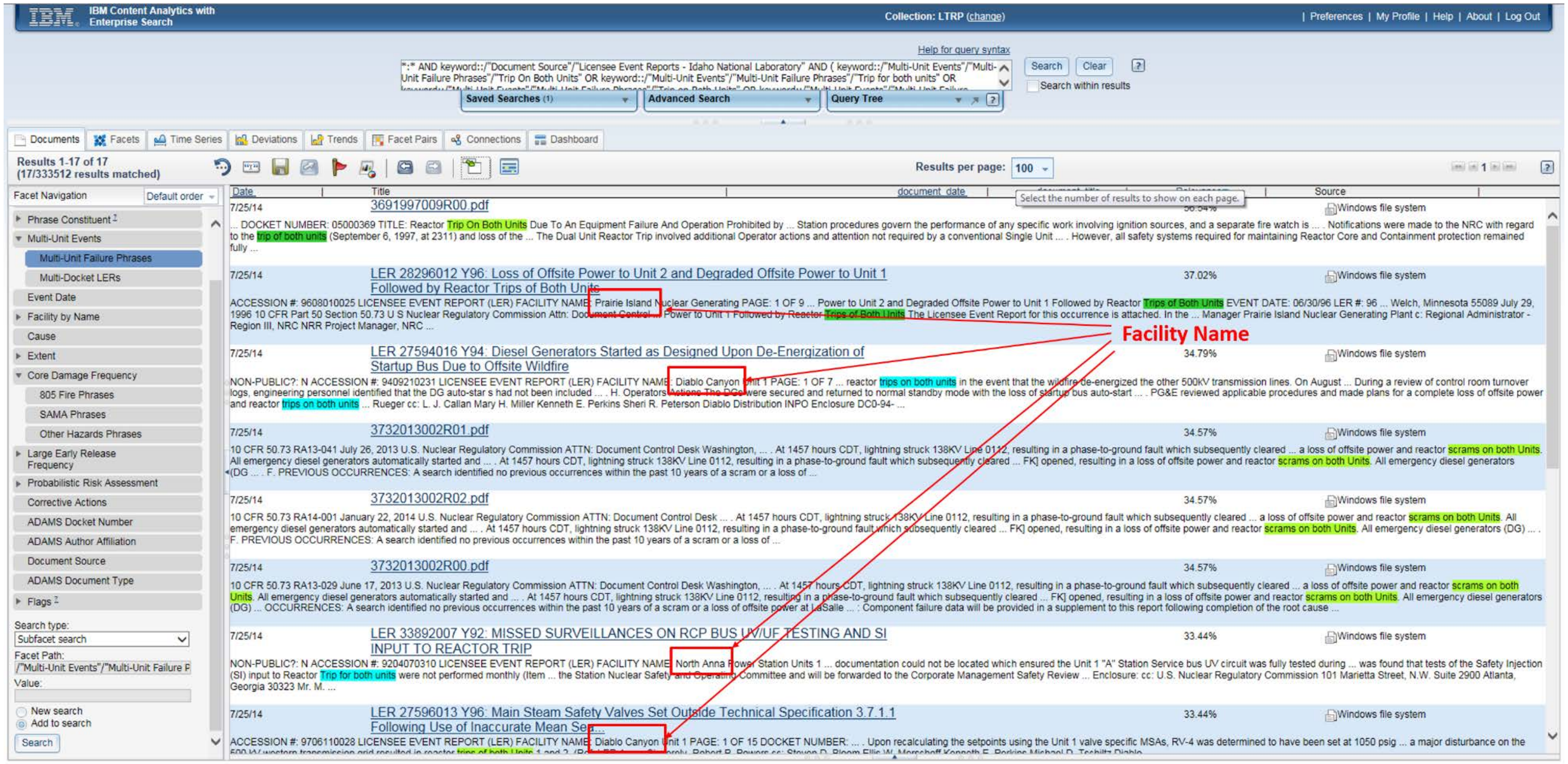

Figure 11. Facility name

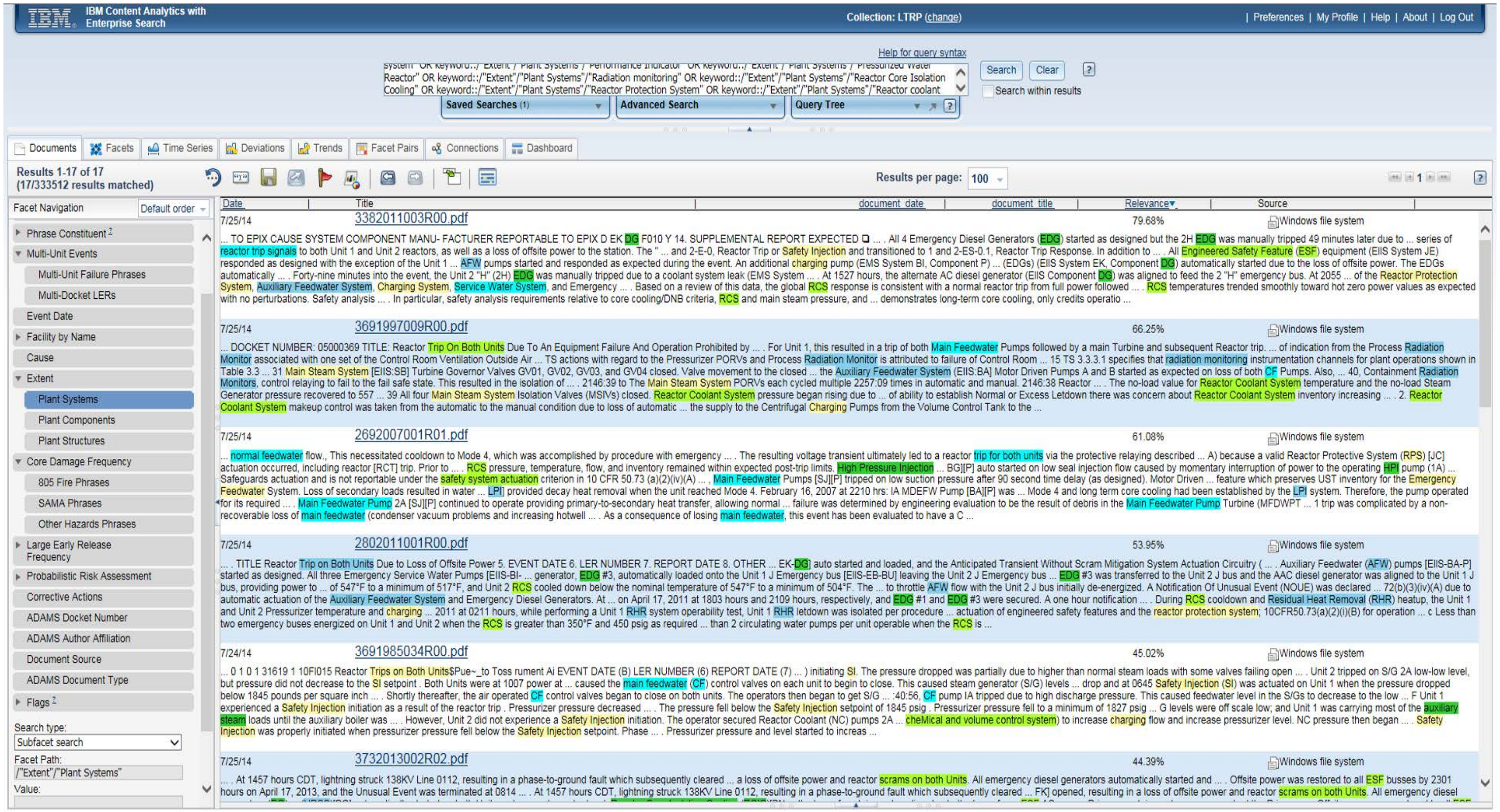

Figure 12. Event extent

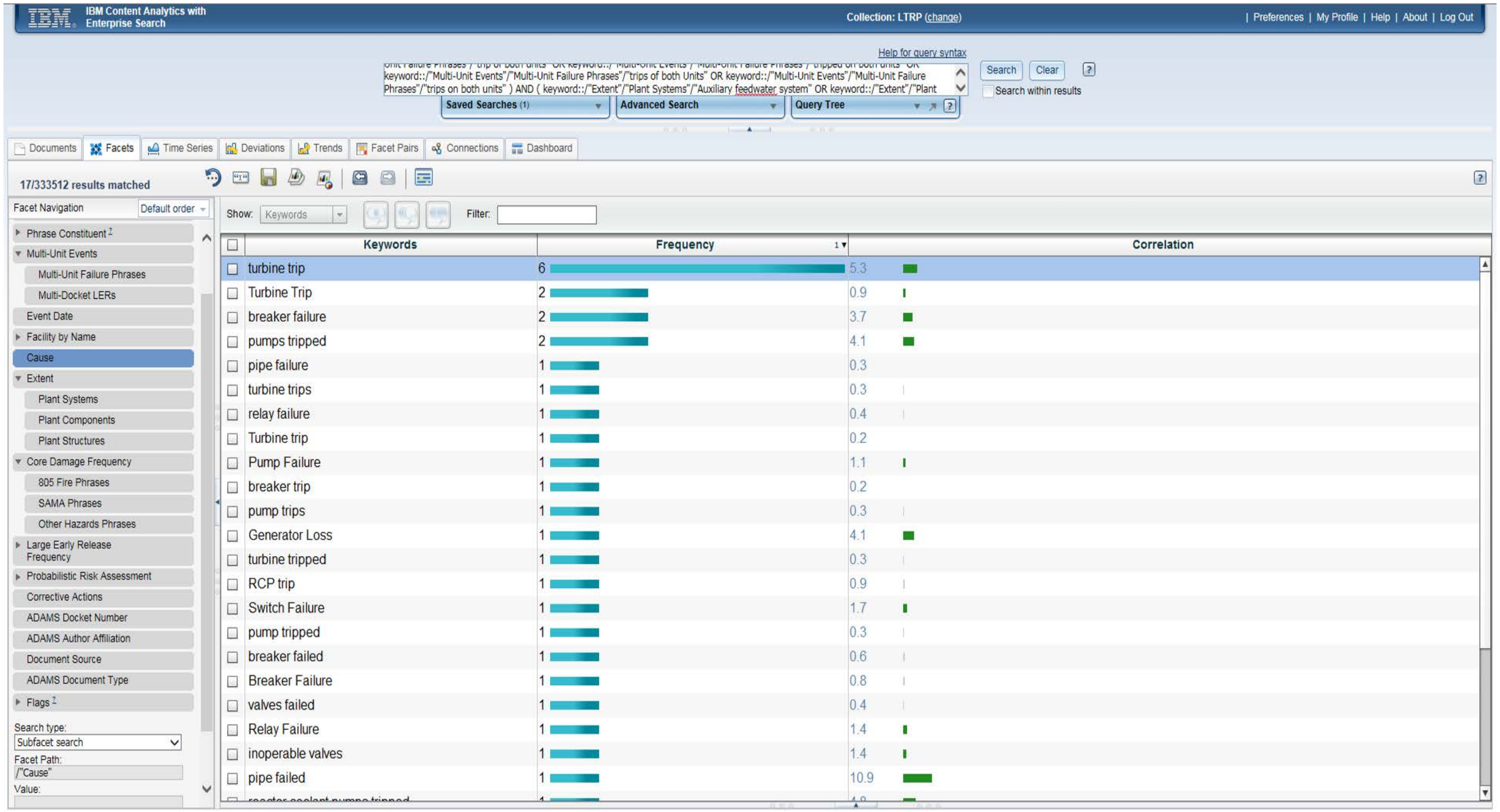

Figure 13. Event cause (current keywords)

- o The ICA 2.2 feature enabling a joint view of facet/subfacets can be useful (when appropriate keyword lists are constructed) in determining the relative frequency of various combinations of cause-related terms. This is further discussed in Section 6 of this report.

- o Although not developed for this project, the subfacet “Part of Speech”/”Verb” supplied with the customized ICA 2.2 tool contains the keyword “cause.” Selection of this keyword will capture documents not only this keyword but also related terms (e.g., “caused,” “causing”). Moreover, the contextual text nearby these terms can provide an indication of the event cause. (See, for example, Figure 14, which shows the results of a search of INL LERs involving the North Anna plant.)

4.5 USE CASE 1 CONCLUSIONS AND REMARKS

This use case has explored the effectiveness and efficiency of ICA 2.2, with limited customization to support the application, in helping users identify and characterize past U.S. operational events involving multiple reactors.

With respect to multi-unit event identification, when provided with highly-discriminating information (e.g., unique characteristics such as the occurrence of an earthquake or tornado, specific event identifiers such as LER number), the tool enabled effective and efficient searches. The search results were as complete as could be expected (search misses were due to missing documents in the corpus, rather than tool deficiencies) and resulted in very few false positives. The tool was easy to use and provided rapid responses (often within a few seconds) to queries.

When provided with less specific information, the searches were less successful – they only identified a small number of relevant events and also identified a fair number of false positives. Improved keyword lists better reflecting the variety of key terms used in the LERs would likely help, but more advanced programming (e.g., to draw inferences across widely separated text) is likely needed to ensure that the searches are effective and efficient. Such additional effort was not judged to be necessary for the purposes of this feasibility study.

With respect to multi-unit event characterization, the customized ICA 2.2 tool provides a number of aids (principally highlighted contextual text) that help users identify event characteristics of interest (e.g., event date, facility name, and event extent). However, these aids are not helpful for all LERs; document download and review remains the surest approach to collect the desired information. In this light, the primary value of the ICA 2.2 tool is in identifying the best documents to download and review.

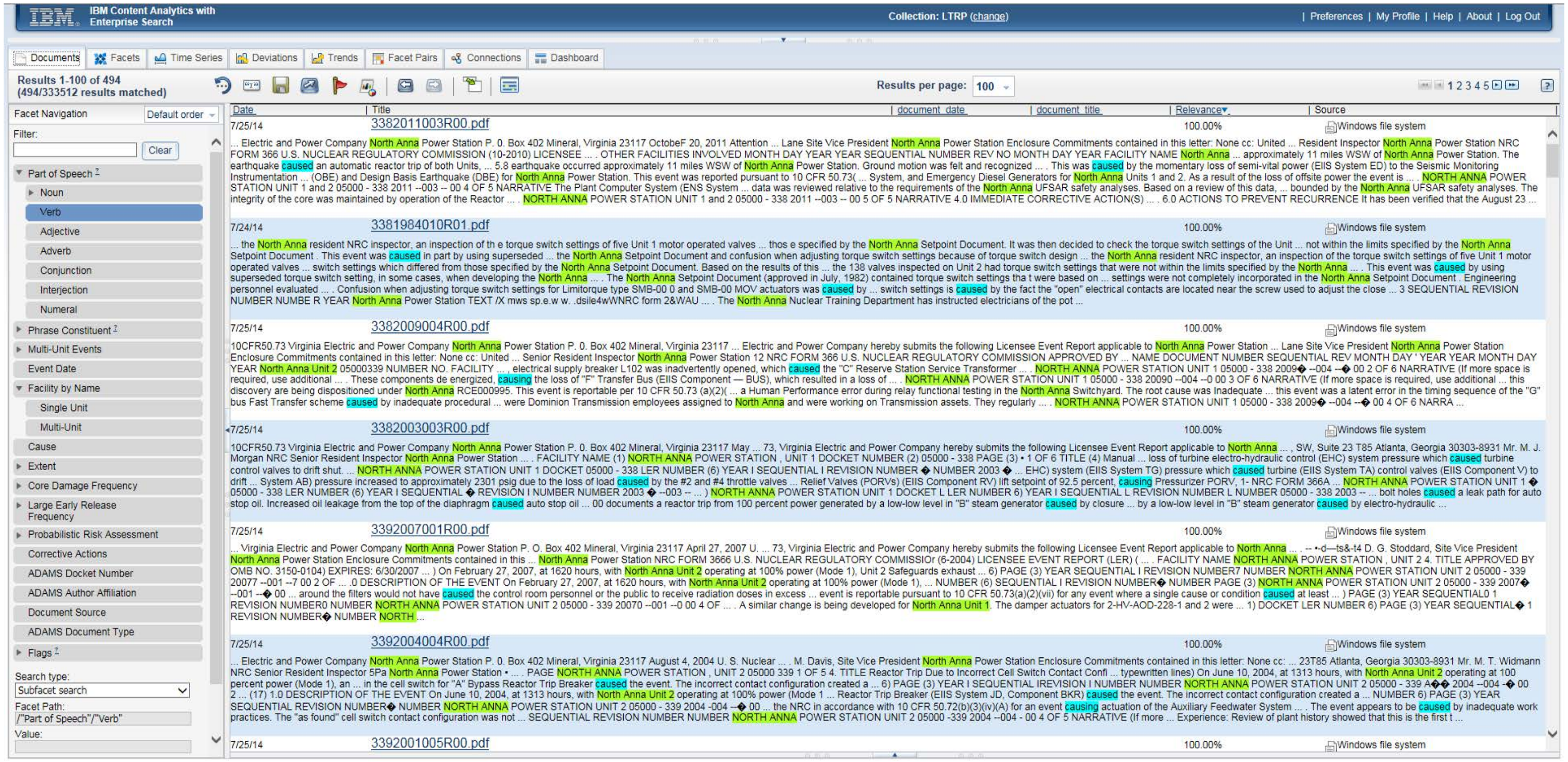

Figure 14. Highlighted contextual text highlighting “cause” and related terms

Two other tools available to NRC staff to identify and characterize multi-unit LERs are LERSearch and the pdf library search capabilities provided with Adobe Acrobat. LERSearch is also extremely effective and efficient for simple searches. However, as compared with ICA 2.2, its advanced query capabilities are somewhat less powerful, its search space is restricted to LERs, and it lacks the ability to save searches. (This last point becomes especially important when refining a search query, and when performing multiple searches.) Adobe Acrobat searches of the library of LERs used in this project are slower than those of ICA 2.2 or LERSearch, less flexible, and less helpful. (Even though contextual text is provided, users typically will need to download and review documents to identify targeted information.)

Overall, the customized ICA 2.2 tool appears to have potential for future use as an event-search tool. Even in its current feasibility-demonstration state, it can support more efficient searches of LERs than currently possible through LERSearch or ADAMS (P8 or Enterprise Search).[23] With further development of its facets/subfacets/keywords (see Appendix C), and perhaps some custom programming (e.g., to take advantage of structured data such as report tables), it might provide NRC users with an even more powerful search tool to address PRA-related information needs. We recognize that such developments are likely to involve non-trivial levels of effort, and may not be judged worthwhile compared with the expected benefits to be gained.

[23] For this purpose, the restricted corpus of the project is actually a benefit, as it reduces the search time required by the more general purpose ADAMS tools.

5. USE CASE 2 – LICENSEE PRA CHARACTERIZATION

As discussed in a survey of international practices regarding the use and development of PRA [43], in the U.S., the adoption of a risk-informed approach is generally voluntary for regulatory applications involving operating reactors – there is no legal requirement for an operating plant licensee to have a PRA for its plants[24] or to submit such a PRA (or its results) to the NRC for review. However, if a licensee chooses to adopt a risk-informed approach, then PRA results must be included as part of the submittal for regulatory approval. For example, if a licensee wishes to transition a plant's deterministic fire protection program to a risk-informed, performance-based program per the requirements of 10 CFR 50.48(c),[25] the licensee's license amendment request (LAR) must, among other things, provide current results from its PRA supporting the acceptability of the transition request. Licensees applying for plant license renewals also typically submit PRA results in support of evaluations required for environmental assessments. The NRC staff's reviews of these evaluations, which include the PRA results, are documented in plant-specific supplements to NUREG-1437 [45].

It can be seen that this voluntary approach to risk-informed applications causes NRC to receive plant-specific risk information on an irregular basis. Moreover, the information for the overall operating fleet is distributed across a variety of documents (typically risk-informed LARs and license renewal requests). Furthermore, because the plants and their PRAs typically change over time, the risk information for a given plant can vary from submittal to submittal.

To address these challenges, an analyst tasked with the development of a summary set of current PRA results for all plants must first identify the document containing the latest set of PRA results for each plant and must then find those results within the document.[26] In simple cases, the results are contained in a summary table somewhere within the document. In more difficult cases, the results are embedded in the document text. Thus, the analyst's task, while not conceptually difficult, can be quite labor intensive. (Recent performances have required several staff-days of effort.)

---

[24] As one partial exception, the calculation of the Mitigating Systems Performance Index (MSPI) (a required, risk-informed element of the NRC's Reactor Oversight Program) requires that the licensee use a plant-specific, limited scope PRA (addressing events occurring during power operation). Note also that plants licensed under 10 CFR 52 are required to have PRAs, although they are not required to submit these PRAs to the NRC.

[25] This rule is commonly referred to by the National Fire Protection Association (NFPA) standard endorsed by the rule: NFPA 805 [44].

[26] For plants that have not undertaken any risk-informed application or requested license renewal, the most recent information available to the NRC may be from the Individual Plant Examination (IPE) and Individual Plant Examination of External Events (IPEEE) programs of the mid-1990s. Even for plants participating in risk-informed applications or license renewal, the plant PRAs may be limited to the treatment of internal events, and so the most recent information on the risk from other hazards may be that developed for the IPEEE program.

Recognizing that content analytics tools in general, and ICA 2.2 in particular, are more than just high-powered search tools, nevertheless the question addressed by this use case is whether the use of such tools could help analysts in developing the desired set of current PRA results.

5.1 USE CASE 2 OBJECTIVE AND SCOPE

The objective of Use Case 2 was to evaluate the ability of ICA 2.2 to help analysts efficiently identify documents containing the most recent risk information for operating plants. To limit staff and contractor resource requirements, and in keeping with the exploratory nature of LTRP projects, the following scope limitations were employed.

- The project corpus was limited to the document types shown in Table 2. At the time of the performance of this use case, the corpus contained around 240,000 documents.[27]
- The task was limited to the consideration of CDF (at-power operation, consideration of all hazards).
- The task was limited to information from four representative plants: Brunswick 1, Calvert Cliffs 1, Wolf Creek 1, and LaSalle 2.

The task focused on CDF because this is an extremely useful metric in current risk-informed applications, and because it is expected that lessons learned from a search for CDF would likely be relevant in searches for other risk metrics (e.g., large early release frequency – LERF).

Brunswick and Calvert Cliffs were selected because they: a) have recent LARs to transition the plant's deterministic fire protection program to a risk-informed, performance-based program per the requirements of 10 CFR 50.48(c); and b) also have been approved for license renewal, as documented in appropriate plant-specific supplements to NUREG-1437 [45]. Both the NFPA 805 LARs and the NUREG-1437 supplements (e.g., [46]) contain relevant CDF information; the latter, which discuss the environmental impact of the license renewal, provide the CDF information in a discussion of potential severe accident mitigation alternatives (SAMAs).

Wolf Creek and LaSalle were selected to test ICA 2.2 under more information-limited conditions. The former has a SAMA analysis but not an NFPA 805 analysis, and the latter has neither.

Some use-case relevant characteristics for the four plants are shown in Table 12.

[27] Subsequent to the completion of this use case, the corpus was expanded to around 330,000 documents. Section 5.4.2 discusses the results obtained with the updated corpus.

Table 12. Search-Relevant Characteristics of Selected Plants

	NFPA 805 LAR		NUREG-1437 Supplement		
Plant	**Date**	**Quantified Hazards**	**Date**	**Quantified Hazards**	**Notes**
Brunswick 1	2012	Internal Internal Flood Fire Seismic Wind External Flood	2006	Internal Internal Flood Fire Wind	• NFPA 805 LAR (non-public) • NUREG-1437 Supplement 25 [46] • CDF for internal hazards provided in 2012 Standardized Plant Analysis Risk (SPAR) model (non-public)
Calvert Cliffs 1	2013	Internal Internal Flood Fire Seismic Wind	1999	Internal Internal Flood Fire Seismic Wind	• NFPA 805 LAR (non-public)NUREG-1437 Supplement 1 [47] • CDF for internal hazards provided in 2012 Standardized Plant Analysis Risk (SPAR) model (non-public)
Wolf Creek 1	N/A	N/A	2008	Internal Internal Flood Fire Other External	• NUREG-1437 Supplement 32 [48] • CDF for internal hazards provided in 2012 Standardized Plant Analysis Risk (SPAR) model (non-public)
LaSalle 1	N/A	N/A	N/A	N/A	• CDF for internal hazards provided in 2012 Standardized Plant Analysis Risk (SPAR) model (non-public) • CDF for internal hazards provided by 2005 LAR for a change in plant Technical Specifications [49] • CDF estimates provided by 1992 Risk Methods Integration and Evaluation Program (RMIEP) [50]. Significant contributors to CDF: fire, internal hazards, internal floods, and seismic.

Note: This table presents the pre-search understanding of the availability of recent, all-hazard CDF information for each plant. As discussed later in this report, this project’s searches revealed more recent information for Wolf Creek and LaSalle.

5.2 USE CASE 2 TECHNICAL CHALLENGES

For an automated search tool, a key challenge for this use case is the identification of documents containing pertinent information (e.g., the most recent estimate CDF for a given plant). The tool needs to:

- recognize the variety of non-standardized phrases that refer to numerical estimates of the plant CDF (see, for example, Table 13), and
- recognize that CDF estimates often appear in tables (e.g., see Figure 15)

For tables, the tool needs to determine the table structure (which may be obvious visually but not obvious to a text-oriented tool), understand the meaning of the table structure (e.g., that the middle column of Figure 15 contains the CDF estimates), and understand the meaning of qualifiers (such as the "internal events" parenthetical in the last line of Figure 15).

For a semi-automated, human-in-the-loop tool such as ICA 2.2, the need to meet the above challenges is significantly reduced. However, to be efficient and practical, the tool needs to produce a relatively small number of hits (both documents and hits within a document) requiring manual review.

Table 13. Examples of Indicative Text for Plant CDF

Source	Indicative Text
NUREG-1437, Supplement 25 [46]	"The baseline core damage frequency (CDF) for the purpose of the SAMA evaluation is approximately 4.19 x 10^{-5} per year."
LAR for Technical Specification Change [49]	"The base CDF for the LSCS Unit 2 PRA is 6.64E-6/yr…"
Evaluation of Integrated Leak Rate Test Extension [51]	"…the total Internal Events Core Damage Frequency (CDF) = 1.61 E-5/year for Unit 1 and CDF = 1.41 E-5/year for Unit 2."

Table 5-3. BSEP Core Damage Frequency for Internal Events

Initiating Event	CDF (per year)	Percent Contribution to CDF
Loss of offsite power (LOOP) – dual unit	1.47×10^{-5}	35.1
Turbine trip	1.14×10^{-5}	27.3
Main steam isolation valve (MSIV) closure/loss of condenser vacuum	4.78×10^{-6}	11.4
Loss of direct current (DC) panel	3.18×10^{-6}	7.6
Loss of alternating current (AC) emergency bus	2.39×10^{-6}	5.7
Loss of control rod drive (CRD)	1.72×10^{-6}	4.1
LOOP – single unit	1.01×10^{-6}	2.4
Other	1.01×10^{-6}	2.4
Internal floods	8.80×10^{-7}	2.1
Loss of reactor building closed cooling water (RBCCW)	4.60×10^{-7}	1.1
Interfacing systems loss of coolant accident (ISLOCA)/excessive LOCA	3.40×10^{-7}	0.8
Total CDF (internal events)	**4.19×10^{-5}**	**100**

Figure 15. Example table identifying plant CDF [46]

5.3 USE CASE 2 APPROACH

The general approach followed the process described in Section 3.2.2.

The task team was comprised of one SME and two software engineers. The SME was a PRA analyst who had, prior to the project, performed manual searches of the ADAMS Main Library for the information of interest to the use case. The software engineers were the same individuals who had worked on Use Case 1.

Similar to Use Case 1, the SME used the customized ICA 2.2 tool to search the corpus and identify potential problems. Following discussions with the software engineers, the latter developed refinements for agreed-upon issues. The potential problems generally involved either a failure to identify corpus documents known to contain the desired information, or an excessive number of "false positives" (i.e., documents identified by the tool that did not contain the desired information). The refinements ranged from complete changes to the search strategy,[28] through

[28] An early approach tried by the team involved focusing on the exponential notation typically used in reporting CDFs. For example, recognizing that these CDFs are typically very small numbers, indicative tokens for a reported CDF of $1x10^{-4}$/ry could be the character strings "$1x10^{-4}$," "1E-4," "1E-04," "$1.0x10^{-4}$," "1.0E-4," and so forth. However, since exponential notation is also widely used in non-PRA contexts, this approach yielded an excessive number of false positives and was not further pursued.

the development of new facets,[29] to modifications of the list of search phrases in a given facet. In some cases, it was determined that the corpus did not contain key documents, and the corpus was updated.

Figures 16 and 17 illustrate some of the keywords developed for the “CDF Phrases” and “SAMA Phrases” subfacets, respectively. In the case of the latter subfacet, the keywords were developed from CDF-relevant portions of plant-specific analyses of Severe Accident Mitigation Alternatives – SAMAs – documented in supplements to NUREG-1437.) Note that the facet and subfacet labels and contents were to assist project diagnostics (e.g., comparisons of results from risk-informed LARs versus SAMA analyses).. Should the tool be revised for broader staff use, some relabeling (and possibly reorganization of keywords within facets) would likely be helpful.

Also similar to Use Case 1, once the customized ICA 2.2 tool was finalized, it was exercised in two modes: informed and basic.

In the informed mode, it was assumed that the user knows that CDF information (for all hazards) is often provided in a plant’s risk-informed LAR (or associated documents) if the plant has applied for approval of a risk-informed application. Furthermore, the NFPA 805 LARs are fairly recent and should represent up-to-date information. If a plant has not made a risk-informed submittal, or if the submittal does not address the CDF from all hazards, the SAMA analyses typically provide this information. (Although many of the analyses are dated, they are more up-to-date than the IPE and IPEEE analyses.) Table 14 shows an example informed mode search for CDF information.

In the basic mode, it was assumed the user does not know about the above sources of information and starts with a “blind search” of the database. Table 15 shows an example basic mode search for CDF information. (It was assumed that the analyst knows that documents containing information on total CDF are likely to have information on the contributions from specific hazards, including fire. The search also took advantage of the ICA 2.2 graphical interface feature which facilitates the selection of documents with specified date ranges.)

Note that Tables 14 and 15 include fairly complex queries regarding the term “core damage frequency” and its variants. In hindsight, such complexity was probably not needed because almost invariably, corpus documents referring to core damage frequency will liberally employ the acronym CDF. Thus, both the informed and basic searches for this use case could have been performed with very simple queries, likely yielding the same results.[30]

---

[29] The facet “805 Fire Phrases” was developed upon recognizing that NFPA 805 LARs (and the associated staff requests for additional information – RAIs – and evaluations) are a useful source of recent CDF information (for other hazards as well as fire).

[30] In some PRA-relevant documents, the acronym “CDF” can also stand for “cumulative distribution function.” Therefore, it is possible that spelled-out variants on “core damage frequency” (to avoid false positives) might prove useful. However, we did not investigate the degree of added value.

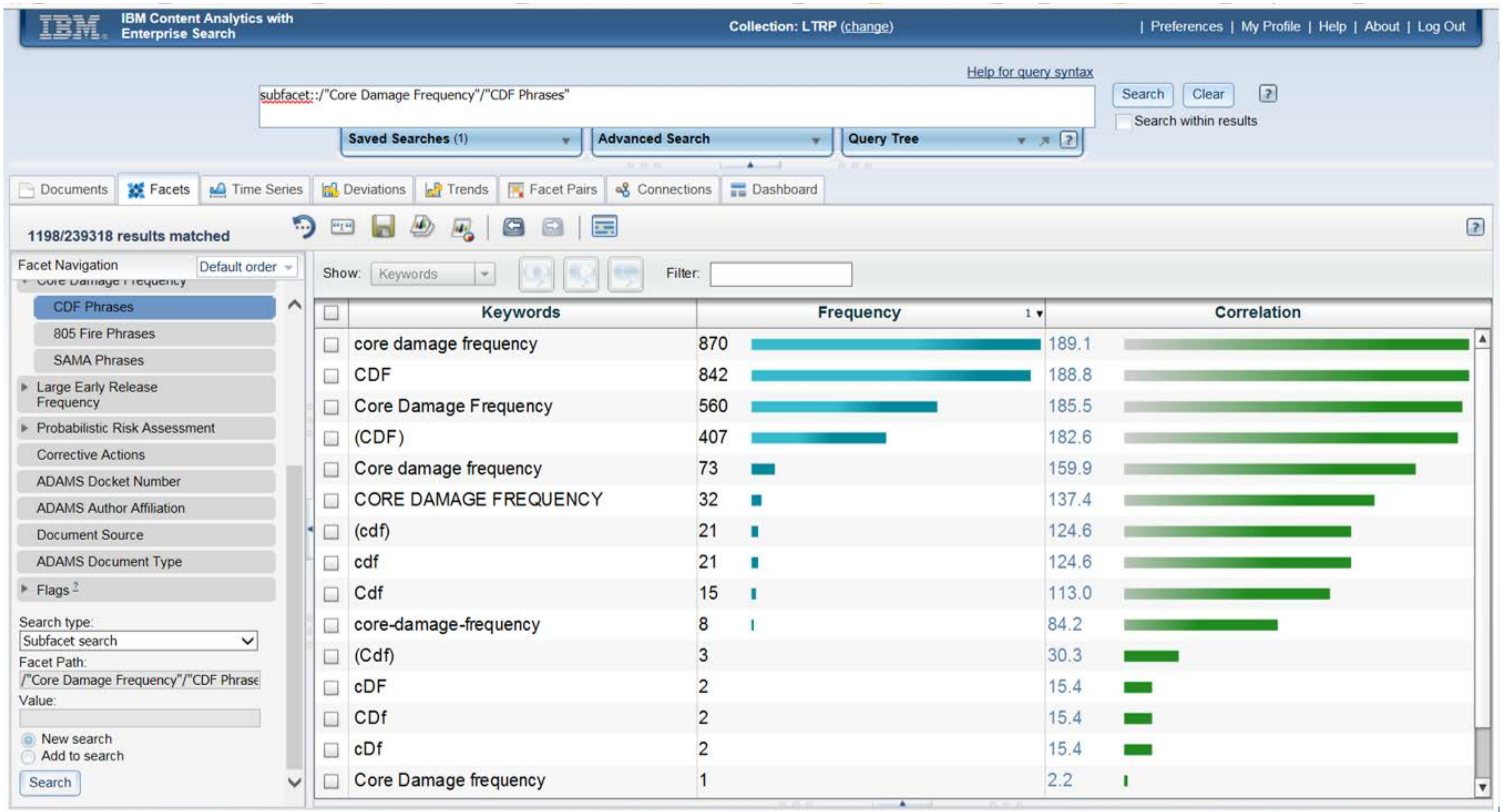

Figure 16. Keywords in "CDF Phrases" subfacet

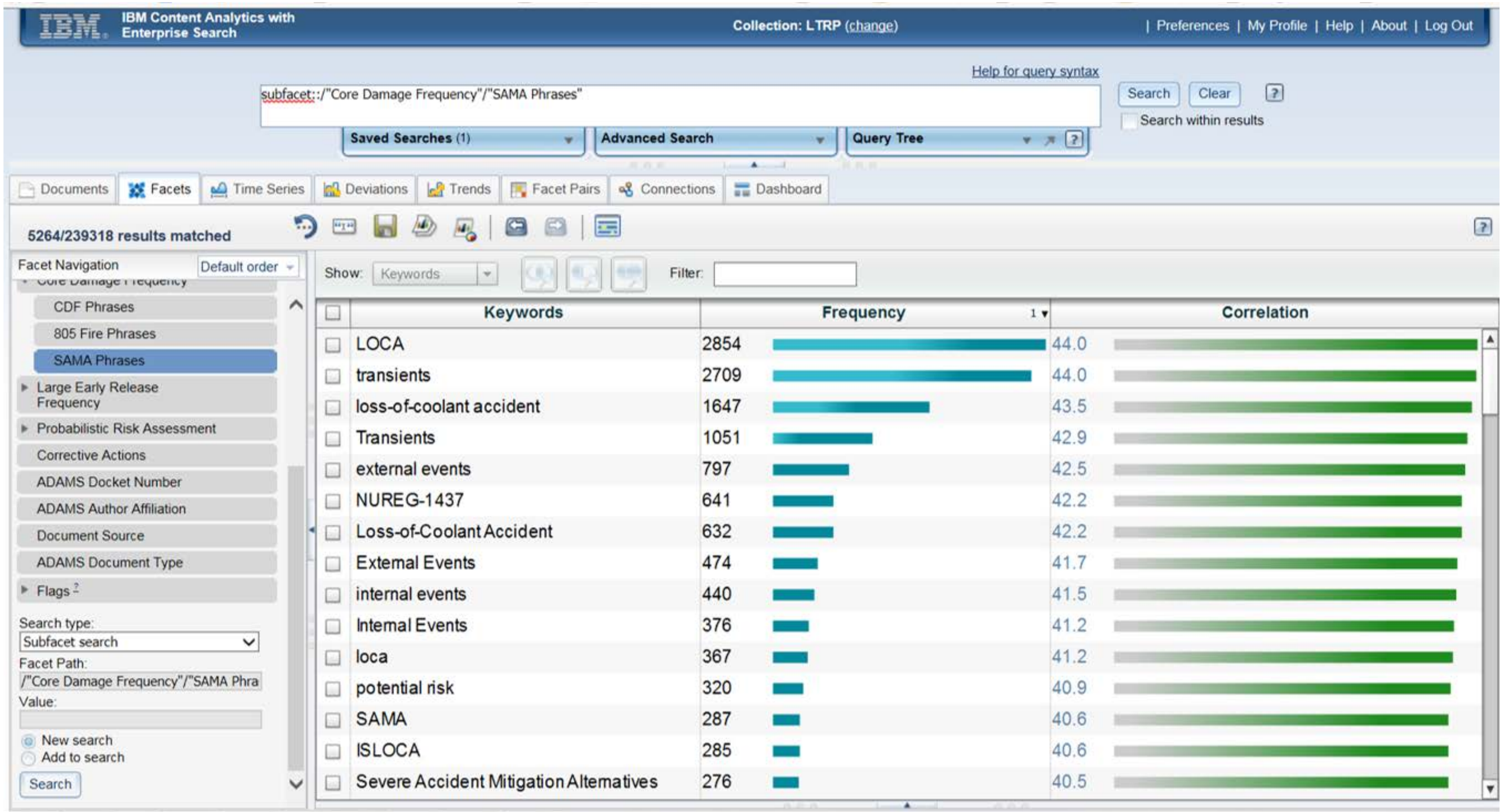

Figure 17. Keywords in “SAMA Phrases” subfacet

Table 14. ICA "Informed Search" Process for Plant-Specific CDFs (for Brunswick)

Step	Incremental Query[a]	Hits
0	"*:*"	239,318
1	AND keyword::/"Core Damage Frequency"/"805 Fire Phrases"/"NFPA 805"	178
2	AND (keyword::/"Core Damage Frequency"/"CDF Phrases"/"CDF" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Core Damage Frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"(CDF)" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"core damage frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Core damage frequency")	92
3	AND "Brunswick"	2

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.

Table 15. ICA "Basic Search" Process for Plant-Specific CDFs (for Brunswick)

Step	Incremental Query[a]	Hits
0	"*:*"	239,318
1	AND (keyword::/"Core Damage Frequency"/"CDF Phrases"/"core damage frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"CDF" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Core Damage Frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"(CDF)" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Core damage frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"CORE DAMAGE FREQUENCY" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"(cdf)" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"cdf" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Cdf" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"core-damage-frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"(Cdf)" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"cDF" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"CDf" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"cDf" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Core Damage frequency" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"CdF" OR keyword::/"Core Damage Frequency"/"CDF Phrases"/"Core-Damage-Frequency")	1,198
2	AND "fire"	817
3	AND "Brunswick"	83
4	AND (document_date>="2012-01-01" document_date<="2013-12-31")	7

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.

For each demonstration plant and search mode, the evaluation addressed search:

- effectiveness (whether the desired CDF information could be found); and
- efficiency (the level of effort required to find this information).

Regarding search effectiveness, it is important to recognize that ICA 2.2 is not designed to directly generate precise answers to search questions such as those posed in this use case (e.g., what is the CDF for Plant X?). Rather, it will help the user find the answer by: a) identifying candidate documents that might contain the answer, and b) providing contextual information (e.g., document titles, document descriptions, document dates, contextual text segments) helping the user to quickly determine if a candidate document should be further investigated. The user still needs to open the document and determine if it actually contains the answer.

Regarding search efficiency, the evaluation considered the number of search steps, the search execution time, and the number of potential "false positives."

The evaluation did not look for additional benefits provided by the tool (e.g., unexpected lessons resulting from the search process, insights developed from the application of the contents analytics features highlighted in Figure 4). The content analytics features are discussed in Section 6 of this report.

To provide a comparison with alternate approaches available to NRC staff, the ICA 2.2 searches were repeated using:

- ADAMS P8;
- Web-Based ADAMS;
- ADAMS Enterprise Search; and
- Google.

The ADAMS Main Library, which the staff currently accesses through the ADAMS P8 application, contains roughly 2 million documents. The ADAMS P8 interface enables users to find documents via a structured file system. It also provides users with simple and advanced tools that can find documents by searching through profiling data (e.g., type, title, author, and date) as well as document contents.

The Publicly Available Records System (PARS) Library, which contains several hundred thousand documents, can be accessed through the Web-Based ADAMS application. The website interface (http://adams.nrc.gov/wba/), similar to ADAMS P8, enables users to find documents via a structured file system or using search tools.

ADAMS Enterprise Search (ES) is an improved tool currently being deployed for use by NRC staff. The tool, which was undergoing development during much of this project, employs modern search tool technologies that are also used in ICA 2.2. ADAMS ES presents the user with an interface similar to that for ICA 2.2. Unlike ICA 2.2, ADAMS ES is specifically a search tool – it lacks some of the analytics features provided by ICA 2.2. Being a tool aimed for general applications, it also lacks the custom facets developed to facilitate the searches performed in this project.

The search tool Google is included in this study as a representative, widely-used Internet search engine. It is ubiquitous, easy-to-use, and provides ready access to documents outside of the ADAMS libraries. Some staff use it as a first-choice search tool even for NRC documents. Unlike the other search tools discussed above, the results of a Google search can change depending on the searches performed by other users. The search was performed only for the Brunswick plant. As discussed later in this report, the use of Google for this application is time-consuming and the insights from the Brunswick demonstration search are likely representative of those for other plant searches.

Note that the non-ICA searches are based on larger databases than this project's corpus and so their results are not directly comparable with those developed using the customized ICA 2.2 tool. Note also that current CDF estimates (both licensee- and NRC-generated) are often only available in non-public documents, thereby automatically handicapping (at least in principle) Web-based ADAMS and Google. Nevertheless, the results and observations from these searches are instructive.

5.4 USE CASE 2 RESULTS

5.4.1 Search Results

The results for the ICA 2.2 searches are summarized in Table 16. The results for the non-ICA 2.2 searches are summarized in Table 17.

In general, the customized ICA 2.2 tool was both effective and efficient for the task problem. The tool helped the user find a document containing the desired information (recent CDFs for all hazards) with relatively little effort. The required search processes were straightforward and the searches were executed quickly.[31] Relatively few false positives were generated and the document information provided by the tool (e.g., titles and contextual text segments) was useful in identifying potential target documents.

Subjectively, the ICA 2.2 tool was easy to use and the wait times for processing were quite acceptable. The tool's different options for entering and modifying queries (via selection of facet keywords, graphical highlighting of such items as date ranges, or direct text entry in the query window) were useful and appreciated.

Note that the above results are based on the application of the ICA 2.2 tool to a limited corpus. Application of the tool to the ADAMS Main Library (which is roughly ten times the size of the corpus) could lead to significantly increased search execution times and to a larger number of potential false positives requiring document download and review.

The ADAMS-based non-ICA 2.2 tools were also effective in identifying useful documents. In some cases, because they were applied to NRC's current libraries, they identified documents not included in the project corpus. (Such documents were received by the NRC after the corpus was constructed.) In the case of Wolf Creek, initial title-based searches were unsuccessful, but this was due to the omission of the phrase "Wolf Creek" in the ADAMS document title.

---

[31] Note that the ICA 2.2 search sessions performed in this task involved iterative applications of search queries followed by query refinements (aimed at identifying a small set of target documents to review in more detail). The execution of each search query typically only required a few seconds – more time was needed to modify the queries than to execute.

Table 16. ICA 2.2 Search Results (1 of 2)

Plant	Search Mode	Results	Notes
Brunswick 1	Informed	• Effectiveness – Found target document. • Efficiency – 3-step process (see notes) identified two candidate documents in <1 minute (search time); ICA 2.2 enables quick identification of document needing detailed review.	• Search is based on knowledge that non-public NFPA 805 LAR contains recent CDF information for all hazards.[a] • Search process: see Table 14. • Based on contextual text from Document View window, both documents are promising; one document appears to be an older version of the other and can be rejected. • It is not clear from information provided by ICA 2.2 whether the selected document actually contains the desired information; document download and review is required. [b]
	Basic	• Effectiveness – Found target document. • Efficiency – 4-step process (see notes) identified 7 candidate documents in <1 minute (search time); ICA 2.2 enables quick identification of document needing detailed review.	• Search is based on assumption that a document containing total CDFs is likely to also contain references to CDFs from key contributors (including fire). • For the project corpus, target document isNFPA 805 LAR. • Search is limited to 2012 and later to ensure results are recent. • Search process: see Table 15. • Four most recent documents are generic reports and can be rejected based on title; based on document review, fifth document is target document.
Calvert Cliffs 1	Informed	• Effectiveness – Found target document. • Efficiency – 4-step process (similar to that for Brunswick) identified two candidate documents in <1 minute (search time); ICA 2.2 enables quick identification of document needing detailed review.	• Search is based on knowledge that non-public NFPA 805 LAR contains recent CDF information for all hazards. • Search process is similar to that for Brunswick (informed mode). • Based on information provided by ICA 2.2, one of the two candidate documents is an ACRS report from 1999 and can be rejected.
	Basic	• Effectiveness – Found target document. • Efficiency – 4-step process (see notes) identified 6 candidate documents in in <1 minute (search time); ICA 2.2 enables quick identification of document needing detailed review.	• Search is based on assumption that a document containing total CDFs is likely to also contain references to CDFs from key contributors (including fire). • Target document is NFPA 805 LAR. • Search is limited to 2012 and later to ensure results are recent. • Search process is similar to that for Brunswick (basic mode). • First document is target document.

[a] "Target document" is the document in the corpus believed, based on the author's past experience as well as the combined results of the searches performed in this project, to contain the most recent CDF information for all hazards.

[b] A similar note applies to all search tools reviewed in this paper.

Table 16. ICA 2.2 Search Results (2 of 2)

<table>
<tr><th>Plant</th><th>Search Mode</th><th>Results</th><th>Notes</th></tr>
<tr><td rowspan="2">Wolf Creek 1</td><td>Informed</td><td>• Effectiveness – Found target document.
• Efficiency – 3-step process (see notes) identified 7 candidate documents in <1 minute; quick review of document titles and dates eliminated all but target document.</td><td>• Search is based on assumption (correct in this case) that SAMA analysis is suitably recent and knowledge; target document is NUREG-1437, Supplement 32 [48].
• Search process:
- Search using keyword “NUREG-1437” selected from sub-facet “SAMA Phrases” (results in 641 documents)
- Add all keywords from sub-facet “CDF Phrases” to query and search within results (results in 195 documents)
- Manually add (AND) “Wolf Creek” to query and search within results (results in 7 documents)
• Target document is identified by title.</td></tr>
<tr><td>Basic</td><td>• Effectiveness – Found target document.
• Efficiency – 4-step process (see notes) identified 14 candidate documents in <1 minute (search time); ICA 2.2 enables quick identification of document needing detailed review.</td><td>• Search is based on assumption that a document containing total CDFs is likely to also contain references to CDFs from key contributors (including fire).
• For the project corpus, target document is Ref. 48.
• Search range covers 2008 to present to ensure results capture target document.
• Search process is similar to that for Brunswick (basic mode).
• Nine most recent documents are generic reports and can be rejected based on title; tenth document is target document.</td></tr>
<tr><td rowspan="2">LaSalle 1</td><td>Informed</td><td>N/A</td><td>Informed search not performed due to lack of a NFPA 805 LAR or a supplement to NUREG-1437 (see Table 12).</td></tr>
<tr><td>Basic</td><td>• Effectiveness – Found target document.
• Efficiency – 4-step process (see notes) identified 96 candidate documents in <1 minute (search time); document download and review is needed to confirm that 1992 RMIEP study provides most recent information.</td><td>• Search is based on assumption that a document containing total CDFs is likely to also contain references to CDFs from key contributors (including fire).
• Target document is RMIEP study summary [50].
• Search range covers 1992-present (to capture RMIEP results).
• Search process is similar to that for Brunswick (basic mode).
• ICA 2.2 provides sufficient information (via the document title, confirmed by the tool’s ADAMS Document Type facet) to indicate that all 96 documents are NRC reports – none were developed by the licensee.
• Based on spot checks, it appears all documents in corpus (including, most recently, NUREG-1437 Vol. 3, Rev. 1 [45], which was published in 2013) refer to information developed in the 1992 RMIEP study. (This can only be confirmed by a comprehensive document download and review.)
• Non-ICA 2.2 searches, which have access to more recent reports beyond those in the project corpus, find a more recent document for LaSalle. See Table 17.</td></tr>
</table>

Table 17. Non-ICA 2.2 Search Results (1 of 4)

Plant	Search	Search Strategy	Results	Notes
Brunswick 1	ADAMS P8	Informed Mode, Advanced Search	• Effectiveness – Found target document. • Efficiency – Search (~1 minute) identified 38 candidate documents. Document names helped identify potentially useful documents.	• Searched for “Brunswick” in the document title and “CDF” and “NFPA 805” in the document content. • Target document is an NRC letter authorizing the NFPA 805 license amendment [52]. This is more recent than the NFPA 805 LAR (the target document for the ICA 2.2 search). • Document names indicate document contents but download and review is needed. • Shifting “NFPA 805” from content query to title query degrades search – doubles search time and finds 4 candidate documents (all false positives).
	Web-Based ADAMS	Basic Mode, Content Search	• Effectiveness – Found target document. • Efficiency – Search (almost instantaneous) identified 29 candidate documents. Document titles helped identify potentially useful documents.	• Searched for “Brunswick” in the document title and “CDF” in the document content, with document dates limited to 2013 and beyond. • Target document is Ref. 52. • Search was completed in ~2 sec. • Document titles indicate document contents, but download and review are needed.
	ADAMS ES	Informed Mode Search	• Effectiveness – Found target document. • Efficiency – Search (~30 sec) identified 86 candidate documents. Document titles and contextual text help limit documents for download and review.	• Search query: “NFPA 805” AND “CDF” AND “Brunswick,” dates limited to 2012-2015. • Document titles and contextual text are much more useful than titles alone; only a handful of documents need to be downloaded and reviewed.
	Google (Public)	Informed Mode Search	• Effectiveness – Found a document (the non-sensitive portion of an NFPA 805 LAR) that appeared to contain desired information distributed over document; detailed review is needed to confirm usefulness. • Efficiency – Searches are rapid but document downloading and review is laborious. Insufficient aids are provided to help users determine document content or, in some cases, document date.	• Basic search (“Brunswick steam electric plant” “CDF”) resulted in 1870 hits. More informed queries reduce down results. • Providing key phrases appearing in the title of the NRC letter authorizing the plant’s NFPA 805 license amendment [52] (found by Web-Based ADAMS) produced 45 results but not the letter. • Searches were completed in fractions of seconds, but document review took several minutes. • Document search was hindered by: lack of dates for some documents and a software bug preventing a return to Google search from some opened NRC documents.

Table 17. Non-ICA 2.2 Search Results (2 of 4)

Plant	Search	Search Strategy	Results	Notes
Calvert Cliffs 1	ADAMS P8	Informed Mode, Advanced Search	• Effectiveness – Found presumed target document. • Efficiency – Search (~1 minute) identified 20 candidate documents.	• Searched for "Calvert Cliffs" in the document title and "CDF" and "NFPA 805" in the document content. • Presumed target document for Informed Mode search is the NFPA 805 LAR. Actual target document was found by Basic Mode search using Web-Based ADAMS (see below). • Quick document reviews provide pretty clear indication that more recent candidate documents are unlikely to have desired information.
	Web-Based ADAMS	Basic Mode, Content Search	• Effectiveness – Found target document. • Efficiency – Search (almost instantaneous) identified 27 candidate documents. Review required for a number of documents – this took ~10 minutes.	• Searched for "Calvert Cliffs" in the document title and "CDF" in the document content, with document dates limited to 2013 and beyond. • Target document is a licensee calculation provided as an attachment to another document [51]. This is more recent than the NFPA 805 LAR. • Search was completed in ~2 sec. • Document titles indicate document contents, but download and review are needed.
	ADAMS ES	Informed Mode Search	• Effectiveness – Found presumed target document. • Efficiency – Search (~2 min) identified 66 candidate documents. Document titles and contextual text help limit documents for download and review.	• Search query: "NFPA 805" AND "CDF" AND Calvert Cliffs," dates limited to 2012-2015. • Presumed target document for Informed Mode search is the NFPA 805 LAR. Actual target document was found by Basic Mode search using Web-Based ADAMS (see above). • Document titles and contextual text are much more useful than titles alone; only a handful of documents need to be downloaded and reviewed.

Table 17. Non-ICA 2.2 Search Results (3 of 4)

Plant	Search	Search Strategy	Results	Notes
Wolf Creek 1	ADAMS P8	Basic Mode, Advanced Search	• Effectiveness – Found target documents (see notes). • Efficiency – Search (~ 1 min) identified 11 pages of results – too many to sort by date. Several minutes needed to review titles, download and review selected documents.	• Used Basic Search mode because an Informed Search (looking for NUREG-1437 Supplement 32 [48]) should be trivial. • Searched for "CDF" and "fire" and "Wolf Creek" in the document content, with document dates limited to 2008 and beyond. (A search using "Wolf Creek" in the document title did not identify Ref. 48 because the ADAMS document title omitted this phrase.) • There are three target documents: Ref. 48 (which appears to provide the latest licensee results), and two non-public documents.
	Web-Based ADAMS	Basic Mode, Content Search	• Effectiveness – Did not find NUREG-1437 Supplement 32 [48] or any later document with desired information. • Efficiency – Search (almost instantaneous) identified 280 candidate documents. A few minutes are needed to review titles, download and review selected documents.	• Searched for "CDF" and "fire" and "Wolf Creek" in the document content, with document dates limited to 2008 and beyond. • Target document is NUREG-1437 Supplement 32 [48]. • Although the list of candidate documents is long, scrolling and checking by title is relatively quick; only a few documents need to be downloaded and scanned for content.
	ADAMS ES	Informed Mode Search	• Effectiveness – Found target documents (see notes). • Efficiency – Search (~ 1 min) identified 416 candidate documents. Document titles allow easy screening of most, but review process is laborious.	• Search query: "CDF" AND "fire" AND Wolf Creek," dates limited to 2008-2015. • Same three target documents as for ADAMS P8 search. • Checking titles and contextual text on a single page of results is quick. Changing page views of results is the most time consuming portion of review.

Table 17. Non-ICA 2.2 Search Results (4 of 4)

Plant	Search	Search Strategy	Results	Notes
LaSalle 1	ADAMS P8	Basic Mode, Advanced Search	• Effectiveness – Found target document. • Efficiency – Search (~2 min) identified 14 candidate documents. Review required for a number of documents.	• Used Basic Search mode because LaSalle has not submitted an NFPA 805 LAR and it does not have a NUREG-1437 supplement. • Searched for "LaSalle" in the document title and "CDF" in the document content, with document dates limited to 2013 and beyond. • Target document is licensee's environmental analysis in support of license renewal [53]. This document was provided to the NRC after this project's corpus was constructed. • Document titles (and knowledge that desired information is likely found in environmental reports since these provide inputs to the staff analyses documented in the NUREG-1437 supplements) helps identify documents for download and review.
	Web-Based ADAMS	Basic Mode, Content Search	• Effectiveness – Found target document. • Efficiency – Search (almost instantaneous) identified 11 candidate documents. Review required for a number of documents.	• Searched for "LaSalle" in the document title and "CDF" in the document content, with document dates limited to 2013 and beyond. • Target document is licensee's environmental analysis in support of license renewal [53]. This document was provided to the NRC after this project's corpus was constructed. • Search was completed in ~2 sec. • Document titles (and knowledge that desired information is likely found in environmental reports since these provide inputs to the staff analyses documented in the NUREG-1437 supplements) helps identify documents for download and review.
	ADAMS ES	Basic Mode Search	• Effectiveness – Found target document. • Efficiency – Search (~1 min) identified 153 candidate documents. Title review largely sufficient to identify document.	• Search query: "CDF" AND "LaSalle" with document dates limited to 2013 and beyond. • Document titles (and knowledge that desired information is likely found in environmental reports since these provide inputs to the staff analyses documented in the NUREG-1437 supplements) helps identify documents for download and review. Contextual text helps further screen documents.

Regarding efficiency:

- Web-Based ADAMS was extremely quick (especially when performing a Content Search) and provided enough information (via document titles) to enable rapid screening of most documents.

- ADAMS P8 was a little slower than Web-Based ADAMS in executing searches (time scales of 1-2 minutes versus a few seconds) but had the advantage of being able to search the entire ADAMS Main Library. Subjectively, ADAMS P8 was, at times, frustratingly slow to the point of appearing unresponsive in performing actions (e.g., sorting documents by date, changing page views) needed to scroll through a long list of search results.

- ADAMS ES, which also accesses the ADAMS Main Library, appeared to be comparable in speed to ADAMS P8. (Of the four searches conducted, ADAMS ES was faster in two, comparable in one, and slower in one.) However, the additional information provided by ADAMS ES (notably the contextual text provided for each document) as well as its different options for modifying search queries (via keyword selection, graphical input, or direct text-based entry in the query window) made this tool more efficient and easy to use than ADAMS P8. Note also that the ADAMS ES search strategy, which considers both document content and structured metadata (e.g., document titles), would appear to make ADAMS ES more immune to data entry problems such as that identified for Wolf Creek.

The Google-based searches conducted in this task were neither effective nor efficient. A search containing several indicative keywords did not identify a publicly available document containing current CDF information for Brunswick [52]. Although searches were conducted in fractions of a second, the tool provided insufficient information (e.g., indications of document content and date) to help users narrow down the large number of hits typically generated. Thus, the overall search process required the downloading and review of numerous documents. The process was further hindered by an apparent software problem that prevented the user from returning to the Google search window from some open documents.

5.4.2 Updated Search

The searches discussed in the preceding section were performed mid-way through the project. Since then, the project corpus has been expanded (from nearly 240,000 documents to over 330,000 documents) and some of the facets and subfacets of the customized ICA 2.2 tool have been modified.

To provide an indication of the effect of these changes on the results, the ICA 2.2 searches (both informed and basic) were repeated using the final project corpus and the final version of the customized ICA 2.2 tool. The updated searches (see Tables 18 and 19 for example searches used for Brunswick) generally provided the same results as obtained earlier.

- All of the updated ICA 2.2 informed searches were effective and efficient – all found the target document in a relatively short amount of time.
- Most of the updated ICA 2.2 basic searches were effective and efficient: three of the four searches und the target document in a relatively short amount of time.
    - All of the searches were executed in a few seconds; most of the time was spent on deciding what search query to use.)
    - In the case LaSalle, the simple search strategy used for the other three plants (i.e., to find documents containing the keywords CDF, fire, and [plant name], and then to exclude documents referring to other plants) was unsuccessful. (The target document – the LaSalle RMIEP study – likely references other plants.) Either a revised search strategy or a more intensive review of intermediate search results would be needed to identify the target document.

Table 18. Updated ICA "Informed Search" Process for CDFs (for Brunswick)

Step	Incremental Query[a]	Hits
0	"*:*"	333,512
1	AND keyword::/"Core Damage Frequency"/"805 Fire Phrases"/"NFPA 805"	569
2	AND (CDF OR "Core Damage Frequency" OR "core damage frequency" OR "Core damage frequency")	199
3	AND "Brunswick"	9

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.

Table 19. Updated ICA "Basic Search" Process for CDFs (for Brunswick)

Step	Incremental Query[a]	Hits
0	"*:*"	333,512
1	AND CDF[b]	3,733
2	AND fire	2,345
3	AND "Brunswick"	146
4	AND -(keyword::/"Facility by Name"/"Arkansas Nuclear 1" OR keyword::/"Facility by Name"/"Beaver Valley 1" OR keyword::/"Facility by Name"/"Beaver Valley 2" OR keyword::/"Facility by Name"/"Braidwood 1" OR keyword::/"Facility by Name"/"Braidwood 2" OR keyword::/"Facility by Name"/"Browns Ferry 1" OR keyword::/"Facility by Name"/"Browns Ferry 2" OR …[c]	29

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.
[b]Keywords not enclosed in double quotes are case-insensitive.
[c]The full query involves the NOT operator to all plant names except Brunswick.

5.5 USE CASE 2 CONCLUSIONS AND REMARKS

The ICA 2.2 tool, as customized for this use case, proved to be effective and efficient in identifying target documents containing the desired information (recent CDFs for a given set of plants). The tool was easy to use and generally helped the user find the target documents in a short amount of time. With minor revisions of the facets (see Appendix C for specifics), the tool should be useful to a broader range of staff.

Other search tools available to the staff (e.g., ADAMS P8, Web-Based ADAMS, and ADAMS Enterprise Search) also proved to be effective and efficient in identifying the target documents. As compared with ADAMS P8 and Web-Based ADAMS, the ICA 2.2 tool provided significant advantages through its advanced interface which, among other things, facilitated the construction and saving of complex queries and provided highlighted contextual text that helped the user more rapidly determine the relevance of a particular hit. As compared with ADAMS Enterprise Search, which employs an interface similar to that for ICA 2.2, the customized facets developed for ICA 2.2 were of some use, but may not have been needed for the simple searches involved in this use case. (The customized facets were of greater use in Use Cases 1 and E.)

It should be noted that the Use Case 2 results are based on a search process that takes advantage of the recency of CDF results developed for NFPA 805 applications, and the standardized reporting of these results. (Later NFPA 805 LARs report CDF contributions from various hazards in a standard table in a standard section of the LAR. This consistent approach makes it easy for a user to rapidly review a search-identified document to see if it contains the desired CDF information.) More general CDF searches may need to consider a wider range of phrases and reporting formats.

6. USE CASE E – DISCOVERY AND EXPLORATION

The questions behind Use Cases 1 and 2 (What are some key multi-unit events worth further examination? What are the current CDFs for U.S. plants?) are fairly specific and can be answered using direct (and often quite simple) search queries. ICA 2.2 is a useful tool for performing such searches, but it is primarily designed to support more complex, open-ended explorations of available data [5].[32] Although the original plans of this LTRP project did not include any activities aimed at testing these capabilities, late in the project it was judged worthwhile to add a limited scope use case to support a broader evaluation of the tool.

6.1 USE CASE E OBJECTIVES AND SCOPE

The objective of this use case was to provide insights regarding the discovery/exploration capabilities of ICA 2.2. To limit the time and resource requirements, the following scope limitations were employed.

- The use case employed the same project corpus and customized ICA 2.2 tool developed for Use Cases 1 and 2 – no additional modifications were made to the corpus or the tool to support this use case, even if a particular search led to inconclusive results.
- The use case involved exploration of a small number of topics (described in Section 6.3).

6.2 USE CASE E CHALLENGES

The principal challenge for this use case involved changing the tool user's mindset from one of "searching" to that of "exploring." Since the latter is not strongly aligned with typical staff tasks and, therefore, typical staff uses of available data, this challenge proved more difficult than might be expected.[33] The approach described in the following section can be viewed as a compromise: it addresses a broader question than typically addressed in staff activities, but it isn't completely open-ended.

A related, secondary challenge involved determining how to use the analytics tools provided by ICA 2.2 to facilitate the exploration process. For example, a key question is how to generate information that suggests where it might be interesting to look next (as opposed to information that helps narrow a search for a particular document or particular document content).

[32] Similar to browsing the stacks of a technical library or "surfing" the website of an organization, the specific endpoint of a content-analytics guided exploration of a database may not be known at the beginning of the activity. The goal is to extract useful insights from the mass of available information, rather than to obtain the answer to a specific pre-defined question.

[33] To some extent, this challenge played a role throughout the project as for quite a while, the SMEs did not fully appreciate the principal focus of ICA 2.2 or how it worked. See Section 7 for further discussion.

6.3 USE CASE E APPROACH

For this use case, the sole participant was an SME (the project technical lead).

The broad question addressed was "What do the data in the corpus tell us about the following topics?"

- External events
    - Reported events
    - Analyses
- Ice storms at or near the Vogtle Nuclear Power Plant[34]
- PRA-relevant content of NRC Inspection Reports

The topics were not developed through any systematic process, but do reflect questions of potential interest to PRA analysts.

For each topic, the SME used the facets in the customized ICA 2.2 tool, sometimes in combination with additional keywords, to perform an initial search. Using the results of this search (principally hits and facet keyword frequency counts, but also more advanced analytics such as correlations between facets), follow-on search questions (perhaps exploration of sub-topics) were identified and pursued. As might be expected, the exploration sometimes led to situations where the query led to a large number of hits whose relevance could only be determined through document download and review. Given the scoping nature of this use case, the exploration was generally terminated at this point.

For some topics, upon completion of the exploration process, a number of the analytics features of ICA 2.2 (Time Series, Trends, Facet Pairs, Connections – see Figure 4) were applied to see what insights (e.g., confirmation of current understanding, surprises) or suggestions for further exploration they might provide.

To provide an example, Tables 20-22 illustrate the exploration process used for the first topic (reported external events). The exploration processes used for all topics are provided in Appendix D.

---

[34] This subject is picked as an exploratory topic that may be of interest to the NRC's Level 3 PRA project [27,28], recognizing that severe ice storms can occur in the Southeast, and that such storms are not typically addressed in detail in current PRAs.

Table 20. Exploration Process for External Event Occurrences – Initial Phase

Step	Incremental Query[a]/Description	Hits
0	“*:*”	333,512
1	AND subfacet::/"Core Damage Frequency"/"Other Hazards Phrases"[b]	69,551
2	AND (keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" OR keyword::/"Document Source"/"Licensee Event Reports - ADAMS")[c]	9,017
3	AND trip*[d]	3,220
4	AND (keyword::/"Extent"/"Plant Systems"/"Reactor Protection System" OR keyword::/"Extent"/"Plant Systems"/"Reactor coolant system" OR keyword::/"Extent"/"Plant Systems"/"Engineered Safety Feature" OR keyword::/"Extent"/"Plant Systems"/"Reactor Pressure Vessel" OR keyword::/"Extent"/"Plant Systems"/"Condensate and feedwater" OR keyword::/"Extent"/"Plant Systems"/"Safety Injection" OR keyword::/"Extent"/"Plant Systems"/"Residual Heat Removal" OR keyword::/"Extent"/"Plant Systems"/"Emergency Diesel Generator" OR keyword::/"Extent"/"Plant Systems"/"Auxiliary feedwater system" OR keyword::/"Extent"/"Plant Systems"/"Emergency Core Cooling System" OR keyword::/"Extent"/"Plant Systems"/"High-pressure Safety Injection" OR keyword::/"Extent"/"Plant Systems"/"Service water system" OR keyword::/"Extent"/"Plant Systems"/"Charging System" OR keyword::/"Extent"/"Plant Systems"/"Reactor Core Isolation Cooling" OR keyword::/"Extent"/"Plant Systems"/"High-pressure Core Spray" OR keyword::/"Extent"/"Plant Systems"/"Pressurized Water Reactor" OR keyword::/"Extent"/"Plant Systems"/"Boiling Water Reactor" OR keyword::/"Extent"/"Plant Systems"/"High-pressure Coolant Injection" OR keyword::/"Extent"/"Plant Systems"/"Radiation monitoring" OR keyword::/"Extent"/"Plant Systems"/"Performance Indicator" OR keyword::/"Extent"/"Plant Systems"/"Emergency Feedwater" OR keyword::/"Extent"/"Plant Systems"/"Low-pressure Coolant Injection" OR keyword::/"Extent"/"Plant Systems"/"Heating, Ventilating, and Air Conditioning" OR keyword::/"Extent"/"Plant Systems"/"Component cooling water system" OR keyword::/"Extent"/"Plant Systems"/"Electrical distribution" OR keyword::/"Extent"/"Plant Systems"/"Main steam system" OR keyword::/"Extent"/"Plant Systems"/"Emergency safety functions actuation signals" OR keyword::/"Extent"/"Plant Systems"/"Fire Protection System" OR keyword::/"Extent"/"Plant Systems"/"Chemical and volume control system" OR keyword::/"Extent"/"Plant Systems"/"Auxiliary steam system" OR keyword::/"Extent"/"Plant Systems"/"Emergency Power System" OR keyword::/"Extent"/"Plant Systems"/"Rod control system" OR keyword::/"Extent"/"Plant Systems"/"Rod position indication" OR keyword::/"Extent"/"Plant Systems"/"Safety System Failure" OR keyword::/"Extent"/"Plant Systems"/"Steam generator water level control" OR keyword::/"Extent"/"Plant Systems"/"Safety System Actuation" OR keyword::/"Extent"/"Plant Systems"/"Low-pressure Safety Injection" OR keyword::/"Extent"/"Plant Systems"/"Reactor trip signals" OR …)[e]	3,066
5	Review keyword frequency counts for - Subfacet “Other Hazard Phrases”[f] - Subfacet “SAMA Phrases”[g]	See notes f and g

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and that of queries for preceding steps.

[b]Returns hits for documents referring to earthquakes, floods, high wind events, etc.

[c]Limits the search to LERs.

[d]Returns hits for “trip,” “trips,” “tripped,” etc.

[e]Partial query shown. The full query returns hits for various plant systems identified in the customized ICA 2.2 tool.

[f]See Figure 18. Shows that “surge” is the most frequently occurring keyword. The references to “external electrical loads” also appear interesting.

[g]See Figure 19. Potentially unexpected results (considering that LERs are the source) include the frequent references to loss of coolant accidents (LOCA) and to CDF.

Table 21. Exploration Process for External Event Occurrences – Exploratory Phase 1

Step	Incremental Query[a]/Description	Hits
6A	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"surge" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Surge")[b]	660[c]
6B	AND (keyword::/"Core Damage Frequency"/"SAMA Phrases"/"LOCA" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"loss-of-coolant accident" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"ISLOCA" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"Loss-of-Coolant Accident" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"loca" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"LOSS-OF-COOLANT ACCIDENT" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"Loss-of-coolant accident" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"Loca")[d]	476[e]
6C	AND (keyword::/"Core Damage Frequency"/"SAMA Phrases"/"total CDF" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"estimated core damage frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"estimated Core Damage Frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"CDF contribution" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"calculated core damage frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"core damage frequency contributions" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"total core damage frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"(CDF) contribution" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"total Core Damage Frequency")[f]	20[g]
6D	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"External Electrical Load event" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external operating events" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external electrical load/turbine trip is an analyzed event" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"External Load event" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external electrical load accident" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external electrical load and/or turbine trip event")[h]	42[i]

[a]Query strings are provided as increments to the query from Step 4 (Table 20).

[b]Returns hits for LERs referring to surge.

[c]Review of the contextual text for the first several LERs indicates the references are to surge tanks, surge lines, and surge capacitors – further review appears uninteresting with respect to external events.

[d]Returns hits for LERs referring to LOCA.

[e]The first several references are to detailed ASP analyses (whose results are presented in various ASP SECY papers). This search process provides an unexpected path to find such analyses.

[f]Returns hits for LERs referring to CDF.

[g]Document reviews for the relatively small number of LERs show that the CDF references are generally indications that an analysis was done – the quantitative results are not presented.

[h]Returns hits for LERs referring to external electrical loads and similar terms.

[i]Contextual text review indicates that the references are to the implications of analyses, and appear uninteresting with respect to external events.

Table 22. Exploration Process for External Event Occurrences – Exploratory Phase 2

Step	Incremental Query[a]/Description	Hits
6E1	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"seismic" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"flooding" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"industrial" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Seismic" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"wind" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Surge" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"flood" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"earthquake" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Industrial" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornado" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Flood" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Hurricane" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Earthquake" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Flooding" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"floods" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Tornado" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"hurricane" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"earthquakes" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Wind" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"SEISMIC" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"high wind" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"other external" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Outside Design Basis EVENT" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornados" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"FLOOD" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornado that causes substantial plant equipment damage and a loss" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"seismic event or tornado caused a loss" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornado event is reduced")[b]	1,346
6E2	AND keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory"[c]	985
6E3	Review - Keyword frequency counts for subfacet "Other Hazard Phrases"[d] - Time Series chart[e] - Trends Plots[f]	See notes d, e, f

[a]Query strings are provided as increments to the query from Step 4 (Table 20).Based on the results of previous searches, intent is to not look for LERs containing uninteresting keywords in the Other Hazards Phrases subfacet (e.g., surges). (Note that search does not exclude such LERs.)

[b]Returns hits for documents referring to selected external hazards of interest (e.g., earthquakes, floods, high wind events).

[c]Excludes LERs not obtained from INL.

[d]See Figure 20. The frequent references to seismic and flooding events are not surprising.

[e]See Figure 21. The Time Series plot indicates a major change at 1988.

[f]See Figure 22. Separate trend charts are provided for keywords that differ only in capitalization.

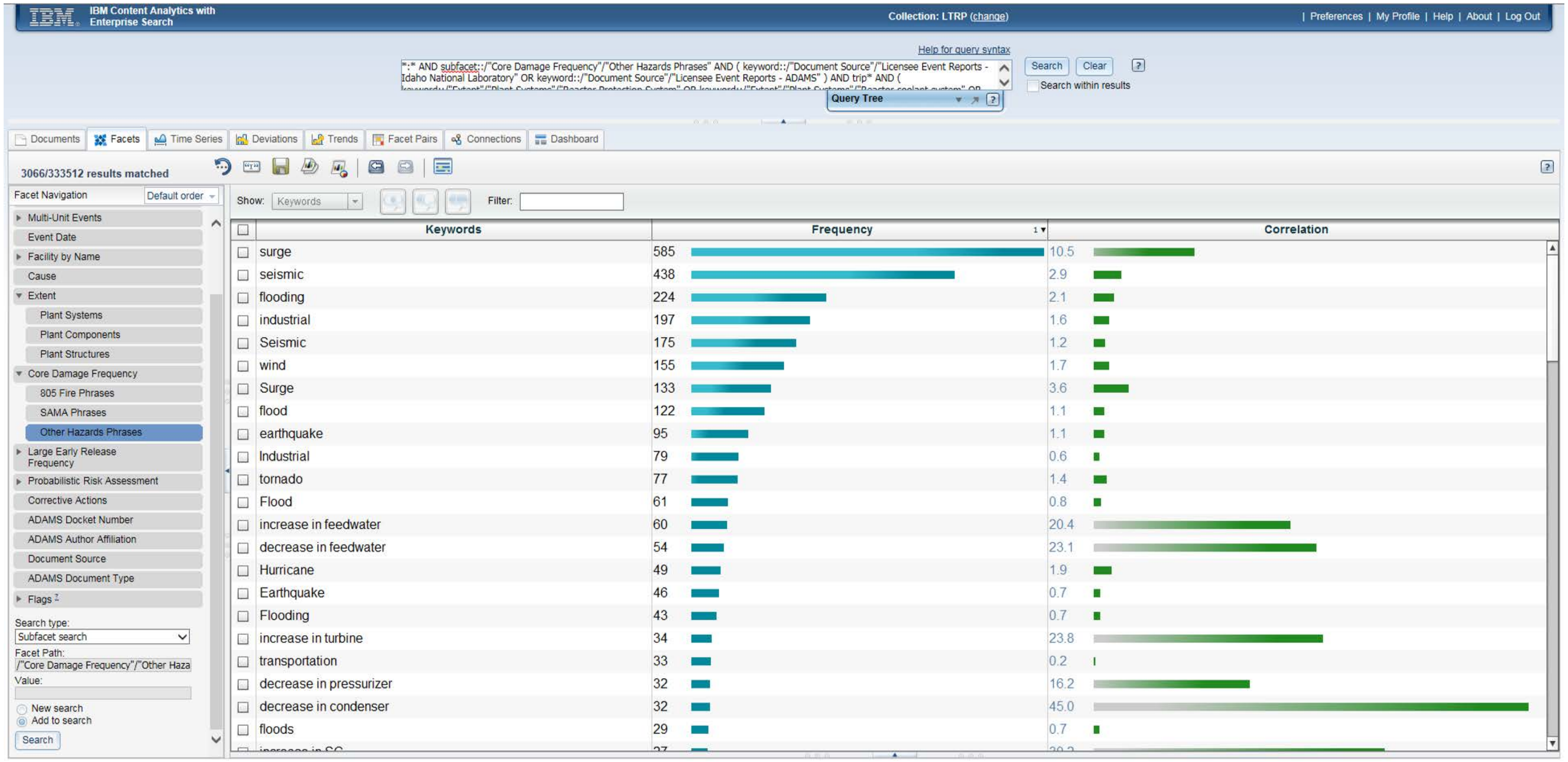

Figure 18. Reported external events – Other Hazards Phrases keyword frequencies

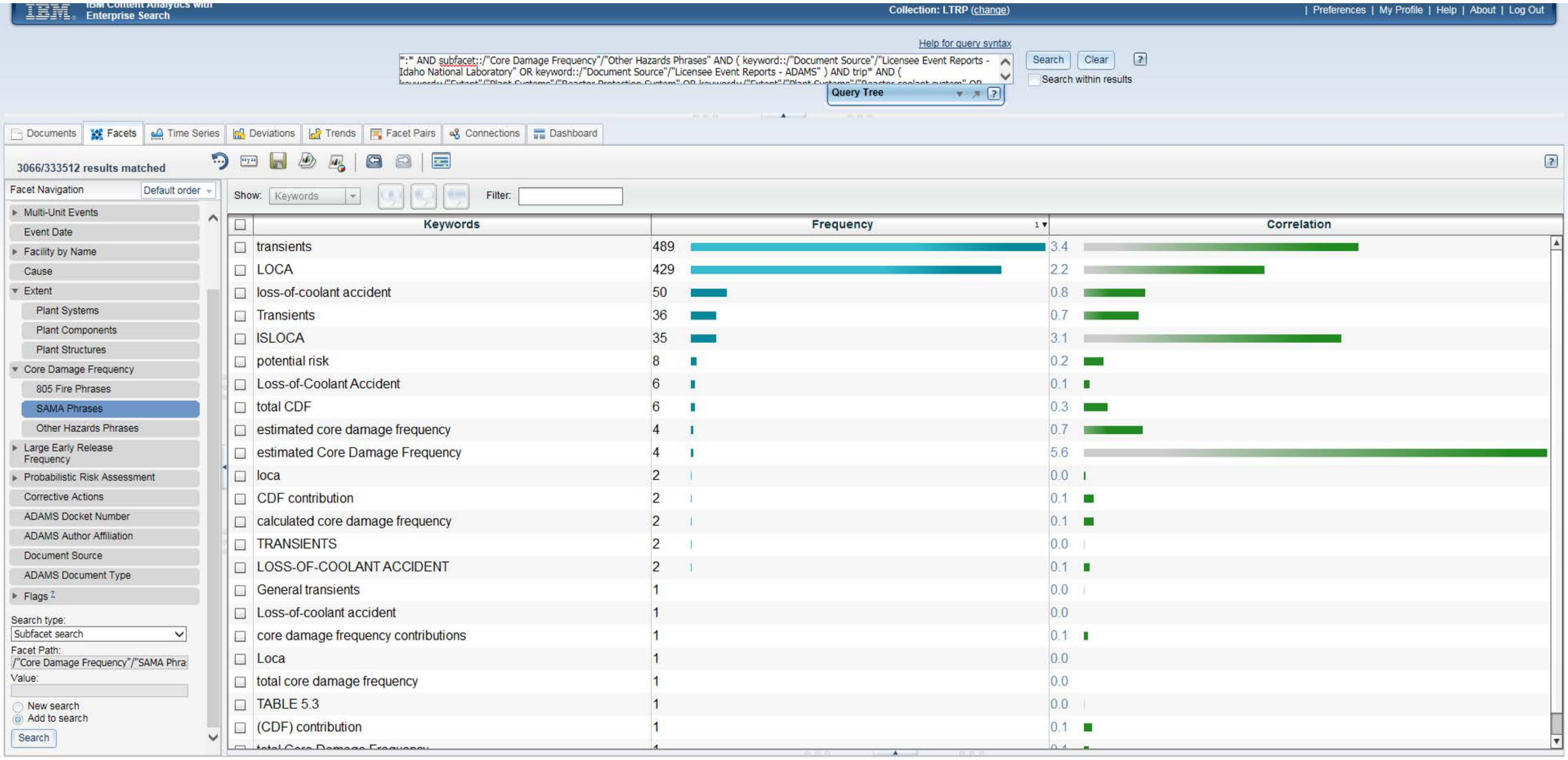

Figure 19. Reported external events – SAMA Phrases keyword frequencies

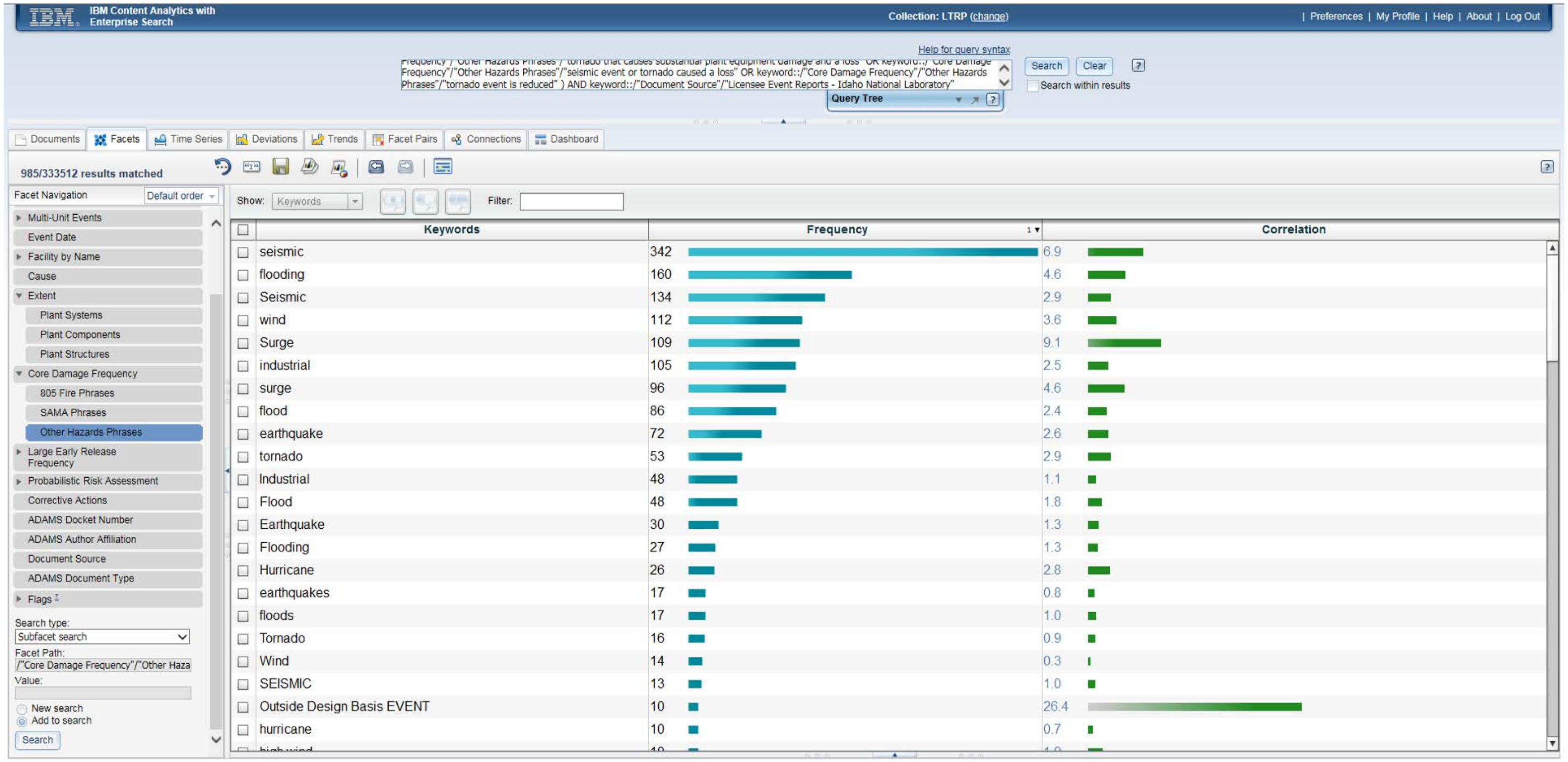

Figure 20. External event LERs – external hazard keyword frequencies

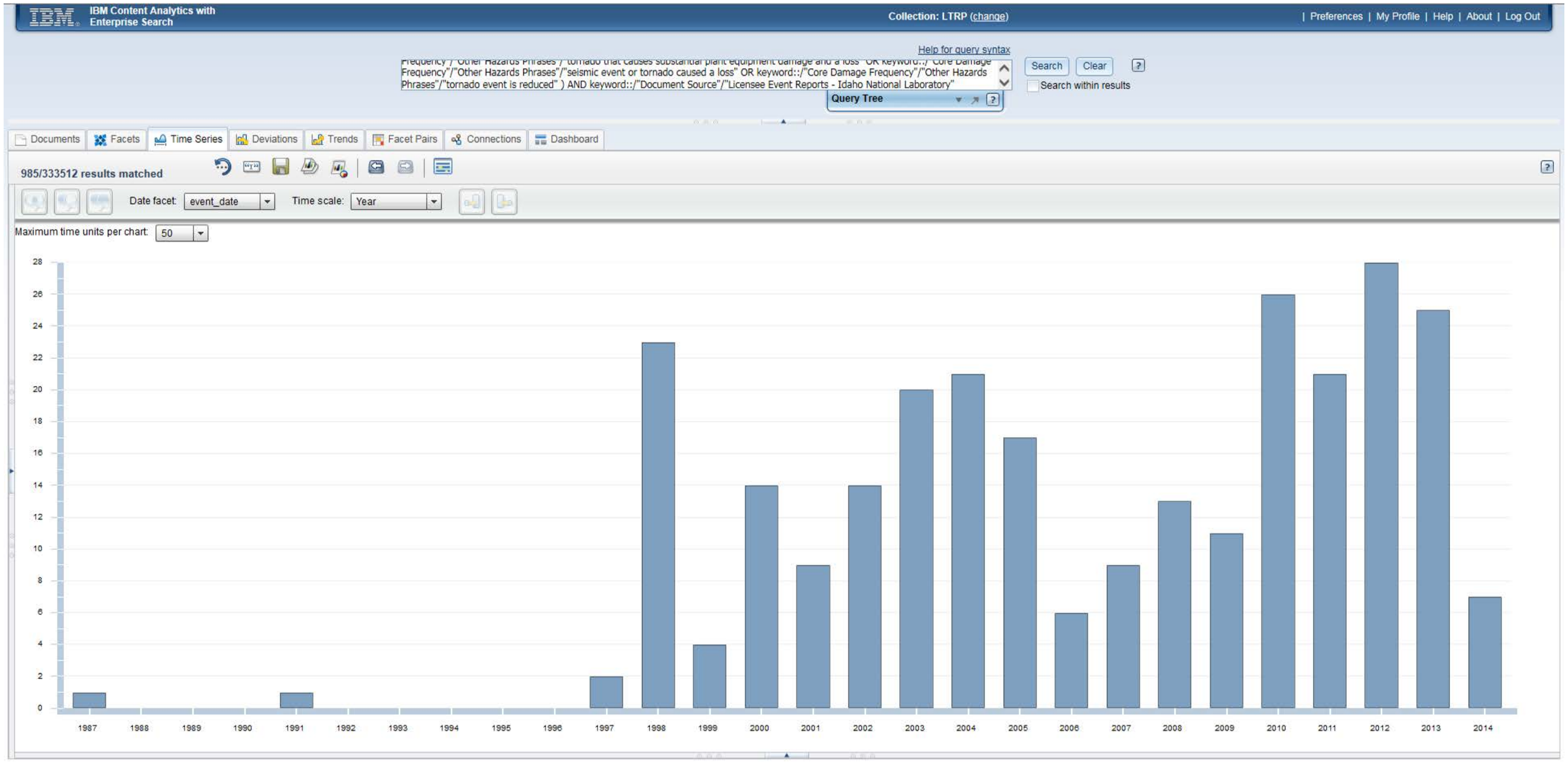

Figure 21. External event LERs – time series chart

t

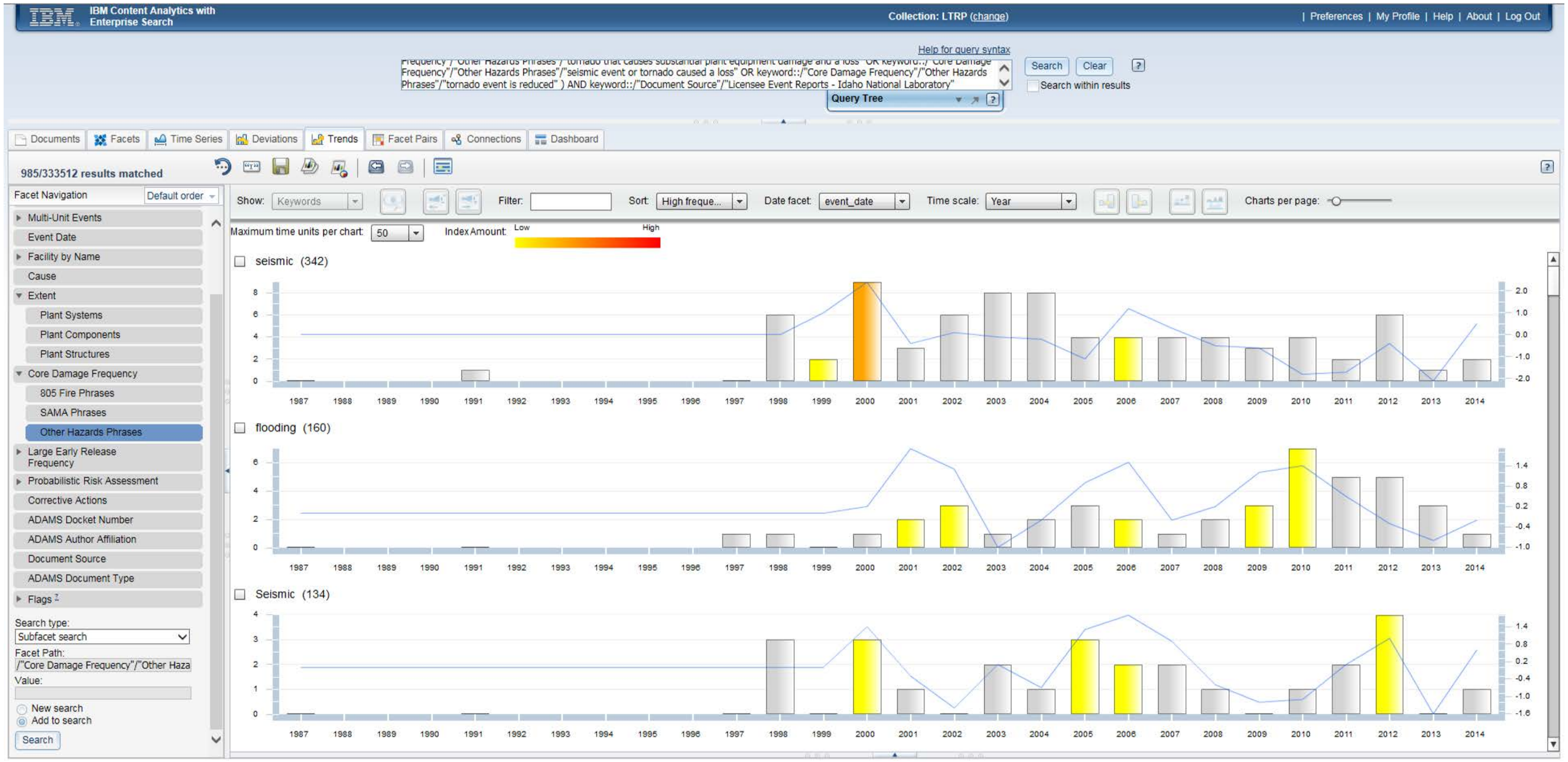

Figure 22. External event LERs – trend charts

6.4 USE CASE E RESULTS

Each of the topic explorations led to observations that: supported prior understandings (i.e., were not surprising), provided interesting information and even surprises suggesting areas for follow-up, provided seemingly spurious correlations requiring further exploration for explanation, or provided indications suggesting potential areas for improvement in the customized ICA 2.2 tool.

Some of the confirmatory observations included:

- The lack of information in LERs regarding quantitative CDF estimates. (LER references to CDF are generally high-level and qualitative, typically indicating that an analysis of CDF was performed to support assessments of the significance of the event.)
- The small number of LERs (10) regarding ice storms.
- Inspection report indications that PRA has been used to determine the importance of inspection findings.
- A high strength of relationship between the Turkey Point plant and hurricanes.

The more interesting observations included:

- An unexpected path (via a search for operating events involving external hazards) to find detailed staff analyses of ASP events.
- A relatively higher fraction of LERs that include seismic- and flooding-related keywords as compared with those including keywords related to high winds.
- A major discontinuity at 1988 in the annual number of external-hazard related LERs (see Figure 21).[35]
- References to several licensee full-scope, Level 3 PRAs.
- References to ice storms in inspection reports (for plants in the Southeast) not captured by LERs. Some of these references indicated the loss of emergency sirens (a potentially important event for a Level 3 PRA).
- Inspection Report identification of a number of specific manual actions determined to be important in the plant PRA.

---

[35] A quick check indicates that this observation applies to all LERs (not just those associated with external hazards). Further exploration indicates that the discontinuity might be at least partly explained by the January, 1988 publication of NUREG-1022, Rev. 1 (which clarified the 30-day reporting requirement for LERs) [54]. Note that the plot includes only LERs for which the event date could be determined from the associated pdf file. As discussed in Ref. 5, a custom software routine for determining the date was developed for the project. The discontinuity at 1988 may be due to changes in LER format (which might affect the performance of the routine) as well in changes in actual reporting.

The seemingly spurious correlations included:

- A large number of hits involving documents referring to storm surge and to fire protection – further investigation (involving document download and review) showed these largely arose from one plant's FSAR.
- A number of hits involving documents referring to ice storms and to the Vogtle plant – further investigation showed that in many of the documents, the references to ice storms and to Vogtle appeared in independent discussions.

The indications for potential areas of improvement included:

- A large number of hits associated with surge tanks and surge lines when searching for operating events involving external hazards. (This suggests a refinement of the keywords in the "Other Hazards Phrases" subfacet.)
- Separate trend charts for keywords that differed only in capitalization (e.g., "seismic" vs. "Seismic" – see Figure 22).
- The Inspection Report keyword frequency chart for "Facility by Name" provides a high count for "Summer." This count likely includes multiple references to the season, and not to the Virgil C. Summer plant, and suggests improvements in identifying plant names.

With respect to the utility of the ICA 2.2 to support database explorations, the interface was easy to use, the response to queries was suitably quick, a number of the facets/subfacets/keywords were helpful (even though they were developed specifically to support Use Cases 1 and 2), and, as in the previous use cases, the contextual text provided with search results was quite useful. Of the content analytics provided, the keyword frequency counts (accessed through the Facets tab) and the Time Series plots were helpful. The Trends view and Facet Pairs view might be useful following some additional work to group similar keywords. Regarding the Connections view, insufficient work was done in this task to determine its potential value.

6.5 USE CASE E CONCLUSIONS AND REMARKS

This limited-effort use case has demonstrated that the ICA 2.2 tool can help a user explore the corpus for potentially useful information and insights regarding topics that the tool was not explicitly designed to address. However, further work is needed to take reasonable advantage of a number of the content analytics features provided by ICA 2.2.

7. COMMENTARY – ORACLES VERSUS AIDES

At the beginning of this project, encouraged by the implications of the IBM Watson Jeopardy! demonstration and the natural language capabilities of personal assistant software (e.g., Apple's Siri, Microsoft's Cortana), the project SMEs hoped that ICA 2.2 would be able to provide direct answers to such natural language questions as "What are some key multi-unit events worth further examination?" (Use Case 1) or "What is the CDF for Plant X?" (Use Case 2). As the project progressed, it became clear that ICA 2.2 is not targeted at this kind of problem.

First, as discussed in Section 6, ICA 2.2 is largely intended to support database exploration. When employed in a direct question/answer mode, it can generate informative intermediate results (e.g., which LERs involving multiple units are referenced in ASP SECY papers) and potentially useful statistics (e.g., how many documents include references to total CDF). However, in general, the user must review contextual text or review linked documents to answer a posed question. Furthermore, given the natural language variations in source documents (e.g., see Tables 4 and 13 and Figure 7), significant effort (well beyond that employed in this technology evaluation project) is needed to ensure that the search results are reasonably complete (without including an excessive number of false positives).

Second, and related to the point above, ICA 2.2 is designed as a human-in-the-loop tool. Thus, in search mode, the tool does not function as an oracle that provides final answers to a user's questions. Rather, it acts as an aide, providing: a) information that suggests, as the search progresses, the next steps a user might take to refine a search, and then b) links facilitating download and review of documents that might contain the answers.

Due to the focus of ICA 2.2, we do not have any empirical data relevant to the current effectiveness and efficiency of commercial, off-the-shelf software to (after appropriate customization) directly answer questions of the sort underlying Use Cases 1 and 2.[36] However, given the complexities revealed in the two use cases, it appears likely that the development of an industrial grade, "push button" solution will require considerable SME and software engineer involvement. Moreover, such a solution, by not involving the SME as an integral part of the actual search process:

- may not take full advantage of SME skills (e.g., recognizing words and numbers despite faulty OCR, recognizing the data relationships implied by a tabular structure) and knowledge (e.g., to recognize apparent conflicts between documents),
- may generate results not fully trusted by the SME, and

[36] Ref. 55 provides an interesting discussion of the status of and challenges being faced by Watson (and other artificial intelligence tools) in the medical field.

- will minimize the SME learning benefits associated with formulating and refining a search (including learning from efforts to develop a search strategy, lessons from “failed searches,” and useful information and insights from intermediate search results).

Thus, although the Watson Jeopardy! demonstration seems to indicate that automated solutions for both Use Cases 1 and 2 are achievable, initiatives to develop such solutions for these problems, or any other PRA- or risk-informed decision making related problems, will need to consider the above costs as well as potential benefits.

8. PROJECT CONCLUSIONS, OBSERVATIONS, AND RECOMMENDATIONS

8.1 CONCLUSIONS

In this project, we have employed three case studies (“use cases”) to investigate the ability of a particular content analytics tool, ICA 2.2 (customized with problem-specific facets), to support searches and database explorations of interest to PRA and risk-informed decision making. Based on the results and observations from the use cases, we draw the following conclusions.

- The customized ICA 2.2 tool is generally effective and efficient in identifying target documents of interest to the use cases. In the one test situation where the tool is not effective (a basic, uninformed search for LERs involving multi-unit events), additional tool refinements (particularly the updating of tool facets) would likely improve tool performance.
- The tool is capable of identifying, with relatively little user effort, target documents not identified by alternate means (e.g., labor-intensive manual searches).
- The tool is capable of supporting more open-ended explorations of the database that lead to potentially interesting insights and suggest avenues for further exploration.
- The human-in-the-loop, stepwise search approach underlying ICA 2.2 is comfortable to use, at least for the corpus and use cases tested. Feedback from queries is quick (typically on the order of a few seconds) and informative, and document downloads (when more detailed information is needed) are also quick.
- The initial development and subsequent refinement of a useful tool requires extensive interactions between the SMEs and the software engineers to ensure mutual understanding of: a) the technical problem(s) targeted by the tool,[37] b) examples of a successful search,[38] and c) the intent and capabilities of the tool.[39]
- Although the customized ICA 2.2 tool has been developed only to support this LTRP project’s technology evaluation, the tool appears to be capable of assisting staff interested in extracting PRA-relevant lessons from operational experience documents.

---

[37] Although not used in this project due to higher-priority demands on project staff, we expect that interactive working sessions involving both SMEs and software engineers would be more effective than the process actually followed (in which, after initial discussions of needs, constraints, and possible solutions, the software engineers developed a prototype tool, the SMEs tested the prototype and provided feedback, the software engineers developed improvements to address higher priority problems, etc.).

[38] These examples of success are critical, as it is difficult or even practically impossible to imagine all possible natural language variants.

[39] For example, as discussed in Section 7, it is critical to recognize that ICA 2.2 is designed to act as an aide, rather than an oracle.

    - As compared with LERSearch (the current tool of choice), the ICA 2.2 interface provides additional capabilities (e.g., supporting the development of complex searches, the saving of these searches, and the rapid screening of search results through contextual text). The ICA 2.2 tool also provides access to potentially useful documents beyond LERs.
    - As compared with more general, ADAMS-based tools (e.g., ADAMS Enterprise Search), the reduced size and pre-indexing of the customized corpus leads to significantly more rapid searches.
- With improvements to the tool facets (see Appendix C), the tool could be more effective and efficient for Use Cases 1 and 2; and will likely also be helpful in other (non-Use Case 1 or 2) searches. Note that because the tool does not have the capability to automatically generate facets from examples of searches, some of these improvements will require a non-trivial effort to construct and test facets.
- Further work, perhaps requiring major programming effort, could significantly increase the power and tool ease of use. This includes work to:
    - take advantage of data structures in technical documents, including document sections, structures within text passages (e.g., subordinate clauses), and tables; and
    - add capabilities (e.g., dragging, ctrl- and shift-clicking) to facilitate selection of ranges of keywords.
- For staff tasks calling for automatic generation of direct answers to natural language questions, tools other than ICA 2.2 should be explored. It should be recognized that, depending on the characteristics of the particular search problem, the costs (both resource- and knowledge-related) of developing and implementing such tools may be considerable, and that the human-in-the-loop approach of ICA 2.2 may actually be more effective and efficient.

8.2 ADDITIONAL OBSERVATIONS

In addition to the preceding conclusions, we make the following observations.

- In general, problems with database documents (e.g., due to errors in the documents, OCR faults, or faulty document profiling) can hinder text-based searches by any tool. For many cases, the keywords of interest occur multiple times within a document, and so database problems may not significantly affect search results. However, cases can arise (e.g., when searching for a document with a specific identifier) when such problems do become critical. If it is important that the search identify all documents matching a specified query, considerable effort may be needed to ensure that potential errors in the documents are identified and handled by the tool.
- As a related matter, it is important to recognize that the ability of KE tools to find desired

information is naturally limited by the availability of that information. The KE tools can enable more rapid identification, review, and processing of relevant documents, but such documents need to be in the corpus.

- It is also worth noting that analytics-based approaches, which employ numerical measures (e.g., keyword frequency counts) to indicate importance, may not point a user towards a document that is the only one that addresses the specific subject of interest. (Of course, such a document can be identified with an appropriate search query.)

- The willingness of users to pursue searches (or explorations) using any tool depends on, among other things, the time required to obtain informative feedback for each query. To help ensure rapid yet helpful feedback when using ICA 2.2, it may be useful to:
    - focus applications on problems that can be addressed with a limited corpus; and
    - provide users with tips for developing queries that generate quicker responses.[40]
- For ICA 2.2 and similar tools, document download and review is an integral part of the search process. Download by hyperlink is straightforward. However, the review portion can be resource intensive. For Use Case 1, the review was aided by the title and summary sections of LERs. For Use Case 2, the review was aided for LARs that provided standardized tables of CDF information in standard document sections. Thus, although ICA 2.2 has been developed to deal with unstructured data, it can be seen that the overall search process benefits from structured data.

8.3 RECOMMENDATIONS

Based on the preceding conclusions and observations, we make the following recommendations.

- NRC staff unaware of ICA 2.2 but performing tasks that could benefit from the customized ICA 2.2 tool (e.g., operational experience reviews to support research activities) should be informed of the tool and, if they are interested, be provided sufficient training to help them start using the tool.[41]
- After cognizant staff have gained some experience with the tool, they should consider: a) working with OCIO to develop facet improvements (and perhaps custom programming) needed to better address their tasks, and b) forming a Community of Practice to facilitate the sharing of lessons learned, best practices, etc.[42]
- Given the importance of structured information within technical documents, OCIO should consider developing ICA 2.2 enhancements that address this type of information.
- Given the potential value of advanced KE tools for NRC technical staff applications, and the need for SME involvement when developing such tools, RES staff should continue to monitor developments in this area (e.g., through the revised LTRP program [56]).

---

[40] For example, queries that lead to large numbers of hits provide results more rapidly than queries that lead to small numbers of hits. Query response time can also be affected by the amount of contextual text provided with each hit. (This amount can be adjusted by the user.)

[41] Note that the project team has already implemented this recommendation for some staffers.

[42] As indicated earlier, NRR/DIRS/IOEB is currently a user of ICA 2.2.

REFERENCES

1. U.S. Nuclear Regulatory Commission, "Use of Probabilistic Risk Assessment Methods in Nuclear Activities: Final Policy Statement," *Federal Register*, Vol. 60, p. 42622 (60 FR 42622), August 16, 1995.
2. U.S. Nuclear Regulatory Commission, A Proposed Risk Management Regulatory Framework, *NUREG-2150*, 2012.
3. D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty, "Building Watson: an overview of the DeepQA Project," *AI Magazine*, **31**, No. 3, pp.59-79 (2010).
4. U.S. Nuclear Regulatory Commission, "Office of Nuclear Regulatory Research Long-Term Research Program," *NUREG/BR-0506*, 2012.
5. W.-D. Zhu, et al., "IBM Content Analytics Version 2.2: Discovering Actionable Insight from Your Content, Second Edition," International Business Machines Corporation, Armonk, NY, 2011.
6. N. Siu, M. Tobin, P. Appignani, K. Coyne, G. Young, and S. Raimist, "Knowledge engineering tools – ready to support risk-informed decision making?," *Proceedings of ANS PSA 2015 International Topical Meeting on Probabilistic Safety Assessment and Analysis*, Sun Valley, ID, April 26-30, 2015. (Available from ADAMS, Accession Number ML ML15028A162)
7. T. Friedlhuber, M. Hibti, and A. Rauzy, "A method to compare PSA models in a modular PSA," *Proceedings of ANS PSA 2015 International Topical Meeting on Probabilistic Safety Assessment and Analysis*, Sun Valley, ID, April 26-30, 2015.
8. U.S. Nuclear Regulatory Commission, "Probabilistic Risk Assessment and Regulatory Decision Making: Some Frequently Asked Questions," *NUREG-2201*, 2016.
9. U.S. Nuclear Regulatory Commission, "Status of the Accident Sequence Precursor Program and the Standardized Plant Analysis Risk Models," SECY-15-0124, 2015. (ADAMS ML15187A434)
10. M.S. Rosenwald, "Serious reading takes a hit from online scanning and skimming, researchers say," Washington Post, April 26, 2014. (Available from https://www.washingtonpost.com/local/serious-reading-takes-a-hit-from-online-scanning-and-skimming-researchers-say/2014/04/06/088028d2-b5d2-11e3-b899-20667de76985_story.html)
11. N. Siu, D. Marksberry, S. Cooper, K. Coyne, and M. Stutzke, "PSA technology challenges revealed by the Great East Japan Earthquake," *Proceedings of PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident*, Tokyo, Japan, April 15-17, 2013. (ADAMS Accession Number ML13038A203.)
12. U.S. Government Accountability Office, "Nuclear Regulatory Commission: Natural Hazard Assessments Could Be More Risk-Informed," *GAO-12-465*, 2012.

13. U.S. Nuclear Regulatory Commission, “Perspectives Gained from the Individual Plant Examination of External Events (IPEEE) Program,” *NUREG-1742*, 2002.
14. J.-M. Lanore, et al., “IRSN analysis of post-Fukushima "Hardened Safety Core": use of PSA insights,” *Proceedings of PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident*, Tokyo, Japan, April 15-17, 2013.
15. T. Takada, “Robustness and resilience as extentions [sic] of risk concept after Fukushima event,” *Proceedings of PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident*, Tokyo, Japan, April 15-17, 2013.
16. E. Vial, V. Rebour, and B. Perrin, “Severe storm resulting in partial plant flooding in ‘Le Blayais’ Nuclear Power Plant,” *Proceedings of International Workshop on External Flooding Hazards at Nuclear Power Plant Sites* (jointly organized by Atomic Energy Regulatory Board of India, Nuclear Power Corporation of India, Ltd., and International Atomic Energy Agency), Kalpakkam, Tamil Nadu, India, August 29 – September 2, 2005.
17. International Atomic Energy Agency, “The Fukushima Daiichi Accident: Report by the IAEA Director General,” STI/PUB 1710, 2015.
18. F. Ferrante, S. Sancaktar, J. Mitman, and J. Wood, “An assessment of large dam failure frequencies based on US historical data,” *Proceedings of ANS PSA 2011 International Topical Meeting on Probabilistic Safety Assessment and Analysis*, Wilmington, NC, March 13-17, 2011, on CD-ROM, American Nuclear Society, LaGrange Park, IL, 2011.
19. “Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications,” *ASME/ANS RA-Sa-2009, Addendum A to RA-S-2008*, ASME, New York, NY, American Nuclear Society, La Grange Park, IL, 2009.
20. U.S. Nuclear Regulatory Commission, “Procedural and Submittal Guidance for the Individual Plant Examination of External Events (IPEEE) for Severe Accident Vulnerabilities, Final Report,” *NUREG-1407*, 1991.
21. U.S. Nuclear Regulatory Commission, “Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants,” *NUREG-75/087*, 1975.
22. “Addenda to ASME/ANS RA-S-2008, Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications,” *ASME/ANS RA-Sb-2013*, ASME, New York, NY, American Nuclear Society, La Grange Park, IL, 2013.
23. T.R. Nellenbach, “Licensee Event Report 2011-003, Revision 1, for the Fort Calhoun Station,” *LIC-11-0039*, letter to U.S. Nuclear Regulatory Commission, May 16, 2011. (ADAMS ML111370123)
24. U.S. Nuclear Regulatory Commission, “Advanced Knowledge Engineering Tools to Support Risk-Informed Decision Making,” Solicitation RES-13, July 1, 2013. (Available from www.fbo.gov.)

25. *Webster's Third New International Dictionary, Unabridged*, P. B. Gove, Ed., G. & C. Merriam Co., Springfield, MA, 1969.
26. K. N. Fleming, "On the issue of integrated risk – a PRA practitioners perspective," *Proceedings of ANS. International Topical Meeting on Probabilistic Safety Analysis (PSA '05)*, San Francisco, CA, September 11-15, 2005.
27. U.S. Nuclear Regulatory Commission, "Options for Proceeding with Future Level 3 Probabilistic Risk Assessment (PRA) Activities," *SECY-11-0089*, Washington, DC, 2011. (ADAMS ML11090A039)
28. A. Kuritzky, N. Siu, K. Coyne, D. Hudson, and M. Stutzke, "L3PRA: Updating NRC's Level 3 PRA insights and capabilities," *Proceedings IAEA Technical Meeting on Level 3 Probabilistic Safety Assessment*, Vienna, Austria, July 2-6, 2012, International Atomic Energy Agency, Vienna, Austria, 2013. (ADAMS ML12173A092)
29. U.S. Nuclear Regulatory Commission, "Technical Analysis Approach Plan for Level 3 PRA Project, Rev 0b," Washington, DC, 2013. (ADAMS ML13296A064)
30. U.S. Nuclear Regulatory Commission, "Glossary of Risk-Related Terms in Support of Risk-Informed Decisionmaking," *NUREG-2122*, 2013.
31. S. Schroer and M. Modarres, "An event classification schema for evaluating site risk in a multi-unit nuclear power plant probabilistic risk assessment," *Reliability Engineering and System Safety*, **117**, 40-51(2013).
32. S.P. Nowlen, M. Kazarians, and F. Wyant, "Risk Method Insights Gained from Fire Incidents," *NUREG/CR-6738*, 2001.
33. J.W. Minarick and C.A. Kukielka, "Precursors to Potential Severe Core Damage Accidents: 1969-1979, A Status Report," *NUREG/CR-2497*, 1982.
34. W.B. Cottrell, J.W. Minarick, P.N. Austin, E.W. Hagen, and J.D. Harris, "Precursors to Potential Severe Core Damage Accidents: 1980-1981, A Status Report," *NUREG/CR-3591*, 1984.
35. J.W. Minarick, J.D. Harris, P.N. Austin, J.W. Cletcher, and E.W. Hagen, "Precursors to Potential Severe Core Damage Accidents: 1985, A Status Report," *NUREG/CR-4674, Vol. 1*, 1986.
36. U.S. Nuclear Regulatory Commission, "Status of the Accident Sequence Precursor (ASP) Program and the Development of Standardized Plant Analysis Risk (SPAR) Models," SECY-05-0192, 2005. (ADAMS ML052700542)
37. U.S. Nuclear Regulatory Commission, "Status of the Accident Sequence Precursor (ASP) Program and the Development of Standardized Plant Analysis Risk Models," SECY-06-0208, 2006. (ADAMS ML061920242)
38. U.S. Nuclear Regulatory Commission, "Status of the Accident Sequence Precursor Program and the Development of Standardized Plant Analysis Risk Models," SECY-07-0176, Enclosure 2, 2007. (ADAMS ML072070464)

39. U.S. Nuclear Regulatory Commission, “Status of the Accident Sequence Precursor Program and the Standardized Plant Analysis Risk Models,” SECY-10-0125, Enclosure 1, 2010. (ADAMS ML102100524)
40. U.S. Nuclear Regulatory Commission, “Status of the Accident Sequence Precursor Program and the Standardized Plant Analysis Risk Models,” SECY-12-0133, Enclosure 1, 2012. (ADAMS ML12220A608)
41. U.S. Nuclear Regulatory Commission, “Status of the Accident Sequence Precursor Program and the Standardized Plant Analysis Risk Models,” SECY-13-0107, Enclosure 1, 2013. (ADAMS ML13232A092)
42. U.S. Nuclear Regulatory Commission, “Status of the Accident Sequence Precursor Program and the Standardized Plant Analysis Risk Models,” SECY-14-0107, Enclosure 1, 2014. (ADAMS ML14230A090)
43. Organisation for Economic Co-operation and Development, “Use and Development of Probabilistic Safety Assessment: An Overview of the Situation at the End of 2010,” *NEA/CSNI/R(2012)11*, Nuclear Energy Agency, Paris, France, 2012.
44. National Fire Protection Association, “Performance-Based Standard for Fire Protection for Light Water Reactor Electric Generating Plants, 2001 Edition” *NFPA 805*, Quincy, MA, 2001.
45. U.S. Nuclear Regulatory Commission, “Generic Environmental Impact Statement for License Renewal of Nuclear Plants,” *NUREG-1437*, *Rev. 1*, 2013.
46. U.S. Nuclear Regulatory Commission, “Generic Environmental Impact Statement for License Renewal of Nuclear Plants, Supplement 25 Regarding Brunswick Steam Electric Plant, Units 1 and 2, Final Report,” *NUREG-1437*, *Supplement 25*, 2006.
47. U.S. Nuclear Regulatory Commission, “Generic Environmental Impact Statement for License Renewal of Nuclear Plants, Supplement 1 Regarding Calvert Cliffs Nuclear Power Plant, Final Report,” *NUREG-1437*, *Supplement 1*, 1999.
48. U.S. Nuclear Regulatory Commission, “Generic Environmental Impact Statement for License Renewal of Nuclear Plants, Supplement 32 Regarding Wolf Creek Generating Station, Final Report,” *NUREG-1437*, *Supplement 32*, 2008.
49. “Request for a License Amendment to Extend the Completion Times Related to Technical Specifications associated with Residual Heat Removal Service Water, Diesel Generator Cooling Water and the Opposite Unit Division 2 Diesel Generator,” J.A. Bauer, Exelon Generation, letter to U.S. Nuclear Regulatory Commission, April 13, 2005. (ADAMS ML051040149)
50. A.C. Payne, Jr., “Analysis of the LaSalle Unit 2 Nuclear Power Plant: Risk Methods Integration and Evaluation Program (RMIEP), Summary,” *NUREG/CR-4832*, *Vol. 1*, 1992.
51. M. Johnson, “Calvert Cliffs Nuclear Power Plant: Evaluation of Risk Significance of Permanent ILRT Extension,” Jensen Hughes, February 17, 2015. (ADAMS ML15051A410)

52. “Brunswick Steam Electric Plant, Units 1 and 2- Issuance of Amendment Regarding Transition to a Risk-Informed, Performance-Based Fire Protection Program in Accordance with 10 CFR 50.48(c) (TAC NOS. ME9623 and ME9624),” letter from A. Hon, U.S. Nuclear Regulatory Commission, to W.R. Gideon, Brunswick Steam Electric Plant, January 28, 2015. (ADAMS ML14310A808)
53. “LaSalle, Units 1 and 2 – Applicant’s Environmental Report, Operating License Renewal Stage, LaSalle County Station, Appendix C-End,” Exelon Generation Company, December 9, 2014. (ADAMS ML14343A897)
54. U.S. Nuclear Regulatory Commission, “Event Reporting Guidelines: 10 CFR 50.72 and 50.73,” *NUREG-1022, Rev. 1*, 1988.
55. B. Keim, “IBM’s Dr. Watson Will See You … Someday,” IEEE Spectrum, May 29, 2015. (Available from http://spectrum.ieee.org/biomedical/diagnostics/ibms-dr-watson-will-see-yousomeday)
56. U.S. Nuclear Regulatory Commission, “Revision to the Agency’s Long-Term Research Program and Related Reporting to the Commission,” COMSECY-16-0006, February 29, 2016. (ADAMS ML16021A353)

APPENDIX A – NUREG-2150 RECOMMENDATIONS

Risk Management Regulatory Framework

- The NRC should formally adopt the proposed Risk Management Regulatory Framework through a Commission Policy Statement.

Power Reactors

- The set of design-basis events and accidents should be reviewed and revised, as appropriate, to integrate insights from the power reactor operating history and more modern methods, such as probabilistic risk assessment (PRA).
- The NRC should establish through rulemaking a design-enhancement category of regulatory treatment for beyond-design-basis accidents. This category should use risk as a safety measure, be performance-based (including the provision for periodic updates), include consideration of costs, and be implemented on a site-specific basis.
- The NRC should reassess methods used to estimate the frequency and magnitude of external hazards and implement a consistent process that includes both deterministic and PRA methods. Consideration of the risks from beyond-design-basis external hazards should be included in the proposed design-enhancement category.
- The NRC should develop and implement guidance for use in its security regulatory activities that uses a common language with safety activities and harmonizes methods with risk assessment and the proposed risk-informed and performance-based defense-in-depth framework.

Nonpower Reactors (NPR)

- The proposed defense-in-depth framework should be applied to the NPR licensing process to ensure that the current amount of defense in depth is appropriate given the relatively small radioactive hazard. This application should include safety and security licensing matters.
- The NRC should evaluate the utility of performing a pilot risk assessment, including consideration of external hazards, using modern risk assessment methods at an NPR. This evaluation would assess the value of the risk insights gained from the risk assessment on the basis of possible safety enhancements and possible contributions to a more efficient and effective risk-informed and performance-based regulatory framework for NPRs.

Materials

- The NRC materials program should continue to apply risk insights and performance-

based considerations, as appropriate, in rulemaking, guidance and policy development, and implementation in accordance with the proposed risk management framework. This consideration should include both safety and security licensing processes.
- The development and rollout of the recommended Risk Management Policy Statement should be closely coordinated with the leadership of the Agreement States.

Low-Level Waste (LLW)

- The NRC should adopt the concept of risk management to the LLW program, as well as any revisions proposed to 10 CFR Part 61 (including performance assessment requirements) and related guidance documents.
- The NRC should develop an explicit characterization of how defense in depth, within the proposed risk management framework, applies to the LLW program and build this into current and future staff guidance documents and into training and development activities for the staff.
- The NRC should include environmental reviews within the scope of its risk management framework.

High-Level Waste (HLW)

- Any future revisions to the regulatory framework for geologic disposal of HLW should be done in accordance with the proposed risk management framework to ensure that risk information continues to be appropriately considered in the development of requirements and appropriately reflect any future HLW disposal paradigm.

Uranium Recovery

- Notwithstanding the current uncertainty associated with the EPA rulemaking, the NRC should adopt the proposed risk management regulatory framework to the uranium recovery program to provide greater efficiency, effectiveness, and predictability in policy development and regulatory decisionmaking.
- The NRC should work closely with the Agreement States and the regulated community to guide implementation of risk management in the uranium recovery program.
- The NRC should include environmental reviews within the scope of its risk management framework.

Fuel Cycle

- The fuel cycle regulatory program should continue to evaluate the risk and the associated defense-in-depth protection by using insights gained from ISAs. ISAs should continue to evolve to support regulatory decisionmaking.

Spent Nuclear Fuel (SNF) Storage

- While elements of the proposed risk management approach have been used in the SNF storage regulatory approach to evaluate the acceptable level of risk and the sufficiency of defense in depth (physical barriers, controls or margins) more consistently, the NRC should develop the necessary risk information, the corresponding decision metrics, and numerical guidelines. This is important in guiding further changes to the existing SNF storage regulatory approach and the evaluation of strategies for extended SNF storage activities.
- As part of the implementation of the proposed risk management regulatory framework, the NRC should more consistently consider the concept of defense in depth explicitly and evaluate its proper use in the SNF storage regulatory program. The NRC should also improve appropriate parts of staff.

Transportation

- Considering the strong international regulatory basis for transportation and the need to conform U.S. standards to those of the IAEA and other member states, application of the proposed risk management framework should focus on implementation guidance.
- The risk management process should be used to influence the future outcome of IAEA deliberations on proposed changes in international transportation regulations.
- The NRC should explore the value of using risk insights to justify regulations different from the IAEA's for domestic use only, such as regulations dealing with domestic storage and transportation of high burnup fuel. Risk information could be used to develop a more flexible approach toward implementing and making gradual changes to current transportation regulations.

APPENDIX B – MANUAL SEARCH FOR MULTI-UNIT EVENTS

One purpose of Use Case 1 was to compare the required effort and results of using the customized ICA 2.2 tool with those of a previous manual search performed in support of the NRC's ongoing Level 3 PRA study.

The manual search used the LER database maintained for the NRC by INL. LERs are reports submitted to the NRC from the plants in accordance with the requirements of 10 CFR 50.73. Examples of reportable events are losses of safety related systems or reactor trips. The LERs document the root cause, mitigating actions, and corrective actions that the plant has taken. The manual search included LERs reported over the period 2003-2013 and took under four weeks to complete. The general process used the following steps.

- Search the INL LER database to generate a spreadsheet of events for all sites that have two or more units and for which the word "trip" appeared in the report. (The latter keyword was used in an effort to exclude degraded conditions.)
- Sort search results by date and select all of the events that have either:
    - Both unit docket numbers on the LER, OR
    - Two separate LERs that were dated within 24 hours of each other at one site.

Once the list of events was compiled and sorted, each LER was reviewed to determine if it represented a genuine multi-unit event, or if it did not (i.e., if it was a "false positive"). An example of a multi-unit event that this search found was the multi-unit loss-of-offsite-power at Nine Mile Point Station and at Indian Point during the 2003 East Coast blackout. Examples of false positives include occasions where the LER listed both docket numbers (i.e., the LER was "dual docketed") but the second unit was not called out in the text of the report, or only called out to say that there was no effect on the unit, or where two events from a single unit occurred within the search-specified 24 hour timeframe.

From the initial search, there were approximately 50 hits that met all of the requirements listed above. Upon further review, it was determined that about 30-40 percent of the results were false positives, leaving about 30-35 actual multi-unit events.

In general, the manual search identified several loss of offsite power (LOOP) affecting both units, events where one unit was in shutdown with equipment out of service when the other unit had an event, and a number of events that seem to be plant specific (e.g., a specific plant layout caused the event to affect the other unit). The LOOP events were the most common single event, but many of them came from a single blackout (the East Coast blackout of 2003).

APPENDIX C – POTENTIAL IMPROVEMENTS FOR CUSTOMIZED ICA 2.2 TOOL

The following potential improvements, which are not expected to require custom programming, should improve the usefulness of the customized ICA 2.2 tool developed for this project.

General

- Ensure ADAMS accession numbers are included as corpus document properties
- Ensure document dates are included as corpus document properties
- Make most keywords case-insensitive, ignore hyphenation for most cases, ensure searches capture terms with the same root (e.g., "cause," "caused," "causing"). (This change should improve the usefulness of search results, and will also reduce user burden by eliminating the need to select multiple keyword boxes.)

Use Case 1

- Ensure event dates are included as corpus document properties
- Ensure plant docket numbers (provided in the subfacet "Multi-Docket LERs") correspond to actual docket numbers
- Update the subfacet "Multi-Unit Failure Phrases" to capture a larger fraction of the ASP events shown in Table 11
- Move keywords referring to reporting systems (e.g., Nuclear Plant Reliability Data System) from the subfacet "Plant Systems" to a different facet
- Update the facet "Cause" to reflect a more PRA-oriented view of "cause." (This will likely be related to Use Case 2 improvements identified below.)
- Ensure that early, ASP-related NUREG/CR reports (NUREG/CR-2497, NUREG/CR-3591, and all volumes of NUREG/CR-4674) are included in the corpus

Use Case 2

- Rename and reorganize facets and subfacets (e.g., to address plant operating mode, radioactive material source, hazards, initiating events, PRA level, risk metrics, significant contributors)
- Develop keywords to address non-NPFA 805 LARs

Use Case E

- Refine keywords in subfacet "Other Hazards Phrases" to emphasize storm surge (and eliminate hits for surge tanks, surge lines, and surge capacitors). Also applies to Use Case 2.
- Refine keywords for facet "Facility by Name" to ensure hits refer to the Virgil C. Summer plant (and not to the season). Also applies to Use Cases 1 and 2.

APPENDIX D – EXPLORATION PROCESSES FOR USE CASE E

This appendix provides the full queries used in Use Case E. These queries can be copied and pasted into the query window of ICA 2.2 (see Figure 4) to reproduce the results of the various searches.

Table D-1. Exploration Process for External Event Occurrences – Initial Phase

Step	Incremental Query[a]	Hits
0	"*:*"	333,512
1	AND subfacet::/"Core Damage Frequency"/"Other Hazards Phrases"	69,551
2	AND (keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" OR keyword::/"Document Source"/"Licensee Event Reports - ADAMS")	9,017
3	AND trip*	3,220
4	AND (keyword::/"Extent"/"Plant Systems"/"Reactor Protection System" OR keyword::/"Extent"/"Plant Systems"/"Reactor coolant system" OR keyword::/"Extent"/"Plant Systems"/"Engineered Safety Feature" OR keyword::/"Extent"/"Plant Systems"/"Reactor Pressure Vessel" OR keyword::/"Extent"/"Plant Systems"/"Condensate and feedwater" OR keyword::/"Extent"/"Plant Systems"/"Safety Injection" OR keyword::/"Extent"/"Plant Systems"/"Residual Heat Removal" OR keyword::/"Extent"/"Plant Systems"/"Emergency Diesel Generator" OR keyword::/"Extent"/"Plant Systems"/"Auxiliary feedwater system" OR keyword::/"Extent"/"Plant Systems"/"Emergency Core Cooling System" OR keyword::/"Extent"/"Plant Systems"/"High-pressure Safety Injection" OR keyword::/"Extent"/"Plant Systems"/"Service water system" OR keyword::/"Extent"/"Plant Systems"/"Charging System" OR keyword::/"Extent"/"Plant Systems"/"Reactor Core Isolation Cooling" OR keyword::/"Extent"/"Plant Systems"/"High-pressure Core Spray" OR keyword::/"Extent"/"Plant Systems"/"Pressurized Water Reactor" OR keyword::/"Extent"/"Plant Systems"/"Boiling Water Reactor" OR keyword::/"Extent"/"Plant Systems"/"High-pressure Coolant Injection" OR keyword::/"Extent"/"Plant Systems"/"Radiation monitoring" OR keyword::/"Extent"/"Plant Systems"/"Performance Indicator" OR keyword::/"Extent"/"Plant Systems"/"Emergency Feedwater" OR keyword::/"Extent"/"Plant Systems"/"Low-pressure Coolant Injection" OR keyword::/"Extent"/"Plant Systems"/"Heating, Ventilating, and Air Conditioning" OR keyword::/"Extent"/"Plant Systems"/"Component cooling water system" OR keyword::/"Extent"/"Plant Systems"/"Electrical distribution" OR keyword::/"Extent"/"Plant Systems"/"Main steam system" OR keyword::/"Extent"/"Plant Systems"/"Emergency safety functions actuation signals" OR keyword::/"Extent"/"Plant Systems"/"Fire Protection System" OR keyword::/"Extent"/"Plant Systems"/"Chemical and volume control system" OR keyword::/"Extent"/"Plant Systems"/"Auxiliary steam system" OR keyword::/"Extent"/"Plant Systems"/"Emergency Power System" OR keyword::/"Extent"/"Plant Systems"/"Rod control system" OR keyword::/"Extent"/"Plant Systems"/"Rod position indication" OR keyword::/"Extent"/"Plant Systems"/"Safety System Failure" OR keyword::/"Extent"/"Plant Systems"/"Steam generator water level control" OR keyword::/"Extent"/"Plant Systems"/"Safety System Actuation" OR keyword::/"Extent"/"Plant Systems"/"Low-pressure Safety Injection" OR keyword::/"Extent"/"Plant Systems"/"Reactor trip signals" OR keyword::/"Extent"/"Plant Systems"/"Low-pressure Core Spray" OR keyword::/"Extent"/"Plant Systems"/"Containment sprays" OR keyword::/"Extent"/"Plant Systems"/"Fuel pool cooling system" OR keyword::/"Extent"/"Plant Systems"/"Pressurizer pressure control system" OR keyword::/"Extent"/"Plant Systems"/"Steam dump control system" OR keyword::/"Extent"/"Plant Systems"/"Control Room Emergency Filtration System" OR keyword::/"Extent"/"Plant Systems"/"Electrohydraulic control system" OR keyword::/"Extent"/"Plant Systems"/"Power distribution limits" OR keyword::/"Extent"/"Plant Systems"/"Control Room Emergency Pressurization System" OR keyword::/"Extent"/"Plant Systems"/"Ex-core nuclear instrumentation" OR keyword::/"Extent"/"Plant Systems"/"Instrument and service air system" OR keyword::/"Extent"/"Plant Systems"/"In-core instrumentation" OR keyword::/"Extent"/"Plant Systems"/"Control Room Area Ventilation System" OR keyword::/"Extent"/"Plant Systems"/"Main turbine and auxiliaries")	3,066
5	Review keyword frequency counts for - Subfacet "Other Hazard Phrases" - Subfacet "SAMA Phrases"	See notes b and c

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and previous queries.
[b]See Figure 16. Shows that "surge" is the most frequently occurring keyword. The references to "external electrical loads" also appear interesting.
[c]See Figure 17. Potentially unexpected results (considering that LERs are the source) include the frequent references to loss of coolant accidents (LOCA) and to CDF.

Table D-2. Exploration Process for External Event Occurrences – Exploratory Phase 1

Step	Incremental Query[a]/Description	Hits
6A	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"surge" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Surge")[b]	660[c]
6B	AND (keyword::/"Core Damage Frequency"/"SAMA Phrases"/"LOCA" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"loss-of-coolant accident" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"ISLOCA" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"Loss-of-Coolant Accident" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"loca" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"LOSS-OF-COOLANT ACCIDENT" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"Loss-of-coolant accident" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"Loca")[d]	476[e]
6C	AND (keyword::/"Core Damage Frequency"/"SAMA Phrases"/"total CDF" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"estimated core damage frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"estimated Core Damage Frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"CDF contribution" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"calculated core damage frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"core damage frequency contributions" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"total core damage frequency" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"(CDF) contribution" OR keyword::/"Core Damage Frequency"/"SAMA Phrases"/"total Core Damage Frequency")[f]	20[g]
6D	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"External Electrical Load event" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external operating events" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external electrical load/turbine trip is an analyzed event" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"External Load event" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external electrical load accident" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"external electrical load and/or turbine trip event")[h]	42[i]

[a]Query strings are provided as increments to the query from Step 4 (Table D-1).

[b]Returns hits for LERs referring to surge.

[c]Review of the contextual text for the first several LERs indicates the references are to surge tanks, surge lines, and surge capacitors – further review appears uninteresting with respect to external events.

[d]Returns hits for LERs referring to LOCA.

[e]The first several references are to detailed ASP analyses (whose results are presented in various ASP SECY papers). This search process provides an unexpected path to find such analyses.

[f]Returns hits for LERs referring to CDF.

[g]Document reviews for the relatively small number of LERs show that the CDF references are generally indications that an analysis was done – the quantitative results are not presented.

[h]Returns hits for LERs referring to external electrical loads and similar terms.

[i]Contextual text review indicates that the references are to the implications of analyses, and appear uninteresting with respect to external events.

Table D-3. Exploration Process for External Event Occurrences – Exploratory Phase 2

Step	Incremental Query[a]/Description	Hits
6E1	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"seismic" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"flooding" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"industrial" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Seismic" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"wind" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Surge" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"flood" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"earthquake" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Industrial" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornado" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Flood" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Hurricane" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Earthquake" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Flooding" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"floods" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Tornado" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"hurricane" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"earthquakes" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Wind" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"SEISMIC" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"high wind" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"other external" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Outside Design Basis EVENT" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornados" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"FLOOD" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornado that causes substantial plant equipment damage and a loss" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"seismic event or tornado caused a loss" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"tornado event is reduced")[b]	1,346
6E2	AND keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory"[c]	985
6E3	Review - Keyword frequency counts for subfacet "Other Hazard Phrases"[d] - Time Series chart[e] - Trends Plots[f]	See notes d, e, f

[a]Query strings are provided as increments to the query from Step 4 (Table D-1).Based on the results of previous searches, intent is to not look for LERs containing uninteresting keywords in the Other Hazards Phrases subfacet (e.g., surges). (Note that search does not exclude such LERs.)

[b]Returns hits for documents referring to selected external hazards of interest (e.g., earthquakes, floods, high wind events).

[c]Excludes LERs not obtained from INL.

[d]See Figure 18. The frequent references to seismic and flooding events are not surprising.

[e]See Figure 19. The timer series plot indicates a major change at 1988

[f]See Figure 20. Separate trend charts are provided for keywords that differ only in capitalization.

Table D-4. Exploration Process for External Event Analyses – Initial Phase

Step	Incremental Query[a]	Hits
0	"*:*"	333,512
1	AND subfacet::/"Core Damage Frequency"/"Other Hazards Phrases"	69,551
2	AND -(keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" OR keyword::/"Document Source"/"Inspection Reports - DVD" OR keyword::/"Document Source"/"Licensee Event Reports - ADAMS")[b]	58,006
3	Review keyword frequency counts for - Facet "Document Source" - Subfacet "Probabilistic Risk Assessment"/"Levels"/"Level – III"	See notes c and d

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and previous queries.

[b]Excludes LERs and Inspection Reports.

[c]LARs are predominant source. Also indicates 4 references to SECY papers; interesting to follow-up on.

[d]Indicates over 300 documents referring to Level 3 PRA; interesting to follow-up on.

Table D-5. Exploration Process for External Event Analyses – Exploratory Phase

Step	Incremental Query[a]/Description	Hits
4A	AND keyword::/"Document Source"/"Commission SECY Paper - ADAMS"[b]	4[c]
4B1	AND (keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"storm surge" OR keyword::/"Core Damage Frequency"/"Other Hazards Phrases"/"Storm Surge")[d]	580
4B2	Review keyword frequency counts for subfacet “Core Damage Frequency”/”805 Fire Phrases”	See note e
4C	AND CDF[f]	2,226[g]
4D1	AND subfacet::/"Probabilistic Risk Assessment"/"Levels"/"Level - III"[h]	327[i]
4D2	AND keyword::/"Document Source"/"License Amendment Requests - ADAMS"[j]	122[k]

[a]Query strings are provided as increments to the query from Step 2 (Table D-4).

[b]Limits search to SECY papers.

[c]Review and download indicates that none are interesting from an external events analysis perspective. Interesting that the search does not identify ASP SECY papers but this is not pursued.

[d]As a matter of curiosity (not prompted by the Step 3 results), a follow-up to the previous exploration of surges (but ensuring that the references are to storm surges).

[e]Keyword frequencies indicates a surprising number of hits referring to fire protection. Follow-up document review indicates that many are associated with the FSAR for South Texas – not surprising in hindsight.

[f]Returns hits for documents referring to CDF.

[g]Document count is too high for review; further narrowing of search judged unnecessary for purposes of exploratory study.

[h]Returns hits for documents referring to Level 3 PRA.

[i]Contextual text review and document downloads indicate a number of reports relevant to full-scope Level 3 PRAs, including the PRA Procedures Guide for the Kalinin plant (NUREG/CR-6572), the LaSalle PRA (NUREG/CR-4832), and the Surry low power and shutdown PRA (NUREG/CR-6144). It also reveals a less-well known full-scope Level 3 PRA: a “generic Level 3 PRA” done for Crystal River in 2000. This suggests a follow-up search to find other industry Level 3 PRAs.

[j]Narrows the search to LARs since the reference to Crystal River comes from a LAR.

[k]A limited review of document titles, contextual text, and downloads for hits indicates potential Level 3 PRAs for DC Cook, the US EPR, Levey, TMI-1, and Harris, and perhaps others. This provides a search process for staff interested in finding instances of industry full-scope Level 3 PRAs.

Table D-6. Exploration Process for Ice Storms in Southeast

Step	Incremental Query[a]	Hits
0	“*:*”	333,512
1A	AND (keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" OR keyword::/"Document Source"/"Licensee Event Reports - ADAMS" OR keyword::/"Document Source"/"Inspection Reports - DVD")[b]	67,586
1B	AND “ice storm*”	28
1C	AND (trip* OR scram*)	24[c]
1D	AND (keyword::/"Facility by Name"/"Summer" OR keyword::/"Facility by Name"/"Browns Ferry")	9[d]
2A[e]	*:* AND (keyword::/"Document Source"/"Licensee Event Reports - Idaho National Laboratory" OR keyword::/"Document Source"/"Licensee Event Reports - ADAMS")[f]	63,714
2B	AND “ice storm*”	10[g]
3A[h]	*:* AND "Vogtle" AND "ice storm*"	26[i]

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and previous queries.

[b]Limits search to LERs and Inspection Reports.

[c]Keyword frequencies show only two Region II plants affected: Summer and Browns Ferry.

[d]Resulting hits (determined by download and review):

- 2 LERs on a 3/1/80 ice storm at Browns Ferry (2961980007R01 and 1961982007R01)
- 2000, 2004, 2005, and 2006 Inspection reports for Summer referring to an ice storm (an actual event on 1/26/04 and an apparently hypothetical event); indicates loss of >25% of Early Warning Siren System
- 2003 inspection report for Harris mentioning emergency preparedness failures due to an ice storm (no details)
- 2005 inspection report for Robinson referring to an ice storm on 12/26/2004
- 2003 inspection report for Fitzpatrick referring to a severe ice storm April 3-7 (led to loss of emergency sirens, offsite power lines)

The results indicate an effect relevant to Level 3 PRA (the loss of emergency sirens). They also indicate events identified in inspection reports but not reported in LERs.

[e]Start of a new search (not restricted to events involving trips).

[f]Limits search to LERs.

[g]The smaller number (as compared with the result of Step 1C) is likely due to the exclusion of inspection reports and (perhaps) duplications between the INL and ADAMS LER entries. Review of hits indicates search captures the two Browns Ferry LERs but not the Summer LER. This search and various follow-up searches also confirm that events mentioned in inspection reports are not documented in LERs.

[h]Start of a new search aimed at Plant Vogtle, not limited to LERs or inspection reports.

[i]Most hits are irrelevant. For example, one document refers to Vogtle as an example of a plant with a neighboring industrial complex, and discusses ice storms in a separate portion of the document. The most relevant hits are for the Vogtle 3 and 4 Early Site Permit and the Vogtle 1 and 2 FSAR (which discuss ice storms as potential hazards). The search indicates there is little additional information to support an analysis of ice storms at Vogtle.

Table D-7. Exploration Process for Inspection Report Contents – Initial Phase

Step	Incremental Query[a]	Hits
0	“*:*”	333,512
1	AND keyword::/"Document Source"/"Inspection Reports - DVD"	3,872
2	Review keyword frequency counts for different facets: - “Multi-Unit Events” - “Facility by Name” - “Cause” - “Extent” - “Core Damage Frequency” - “Probabilistic Risk Assessment” - “Corrective Actions”	See note b
3A	AND subfacet::/"Probabilistic Risk Assessment"[c]	14[d]
3B1	AND keyword::/"Extent"/"Plant Systems"/"Service water system"[e]	1,849
3B2	Review Facets, Time Series, Trends, Facet Pairs, and Connections views	See note f

[a]Query strings are provided as increments – for each step, the actual query is the combination (AND) of the indicated query and previous queries.

[b]Some interesting observations (e.g., there are explicit references to PRA in a few inspection reports) and some likely spurious (e.g., the high count for “Summer” probably is due to the season, rather than the Virgil C. Summer plant). The meaning of high correlation scores is unclear, given that the search has not yet considered pairs of facets.

[c]Limits results to inspection reports containing at least one of the keywords in the Probabilistic Risk Assessment subfacet. Search is performed because illustrations of the various regulatory uses of PRA are of interest.

[d]Review of contextual text and downloaded documents shows that the hits include inspection reports that:

- identify a number of important manual actions (as identified by the plant PRA)
- refer to Level 2 PRAs for the plants addressed (Palisades and Hope Creek)
- indicate that PRA was a resource in determining the importance of a finding
- refer to a “qualitative probabilistic risk assessment”

[e]Search addresses service water because it’s an interesting system (not because of keyword frequencies). The query is incremental to Step 1.

[f]Observations from different views include:

- Facets View: Inspection Reports commonly reference EDGs (not surprisingly).
- Time Series View: there are no apparent trends (in the number of inspection reports referencing the service water system) for follow-up.
- Trends View: A few keywords (e.g., “In-core instrumentation”) show different patterns from the Time Series view and might be worth follow-up.
- Facet Pairs View:
  - A cross-comparison of the “Extent”/“Plant Systems” subfacet with the “Core Damage Frequency”/“Other Hazards Phrases” subfacet shows, not surprisingly, multiple entries including service water and flooding.
  - A cross-comparison of the “Facility by Name” facet with the “Core Damage Frequency”/“Other Hazards Phrases” subfacet shows multiple hits involve the Cooper plant in combination with seismic and flooding hazards. This may be an interesting topic for follow-up.
- Connections View: This view does not support the above implications of the Facet Pairs view (the possibility of a strong connection between service water and flooding or between the Cooper plant and flooding). It also indicates a strong connection between the Turkey Point plant and hurricanes (not a surprise) and a connection between the Salem plant and “industrial” (which may be worth follow-up).