Knowledge Engineering Tools – Ready to Support Risk-informed Decision Making? [A Status Report]

N. Siu, M. Tobin, P. Appignani, K. Coyne, and G. Young
U.S. Nuclear Regulatory Commission

S. Raimist
ECM Universe

Presented at
ANS International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA 2015: A Forum for Learning)
Sun Valley, ID, USA; April 26-30, 2015

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

KE Tools for RIDM: why are we interested?*

- Needs
  - Large (and increasing) amounts of information
  - Changing organizational demands
    - More (and more demanding) RI-applications
    - Lessons learned from major events (imagination…)
    - Changing demographics, consumption of information
- Promise of new technologies
  - IBM Watson demo (2011)
  - More practical tools (e.g., content analytics)
- Need for SME involvement

*See Siu et al., "Knowledge engineering tools – an opportunity for risk-informed decision making?" *Proc. PSA 2013*, Columbia, SC, 2013 (ML13212A238)

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

FY2015 scoping study

Is additional agency effort recommended?

- Under auspices of NRC Long-Term Research Program
- Scope: content analytics and formal methods
- Content analytics
  - Sources sought notice (2013)
  - Joint effort with NRC Office of Information Services
    - IBM Content Analytics 2.2
    - OIS also supporting other NRC applications

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

Content analytics use cases

No.	Description	Notes
1	**Search for multi-unit events**	Supports characterization of past events involving multiple units at a site. This characterization could support the ongoing development of a site-wide PRA model.
2	**Search for common cause failure events**	Identify and characterize past CCF events. The results of this activity could support the conduct of an expert elicitation for the likelihood of events potentially relevant to the analysis of Interfacing Systems Loss of Coolant Accidents (ISLOCAs).
3	**Characterization of current licensee PRA results**	Support decision maker understanding of current risk levels and contributors. This activity addresses a common question raised by managers and external stakeholders.

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

Content analytics database ("corpus")

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.
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 per 10 CFR 50.72
LERs	Documents notifying the NRC of events 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
ACRS letter reports	1985-present
ACRS Meeting Transcripts	1999-present (subcommittee as well as full committee)

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

Examples of participant activities

- PRA SME
  - Support software engineer understanding of use case, re-definition/re-scoping as necessary
  - Provide vocabulary lists (possibly structured) and example manual solutions
  - Review draft tool results (e.g., to reduce false positives), help determine 80/20 point
- Software engineer
  - Educate SME regarding what tool can/cannot do
  - Develop “annotator” to identify text patterns

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

Project status

- Use Case 1 (multi-unit events)
  - Initial analysis of corpus (~50 hits, ~30-40% false positives)
  - Currently refining search terms
- Use Case 3 (current CDFs)
  - Initial pass explored exponential notation
  - Currently building on manual search cases (focus on fire PRA)
  - Current aim: demonstrate support of user in developing desired answer

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

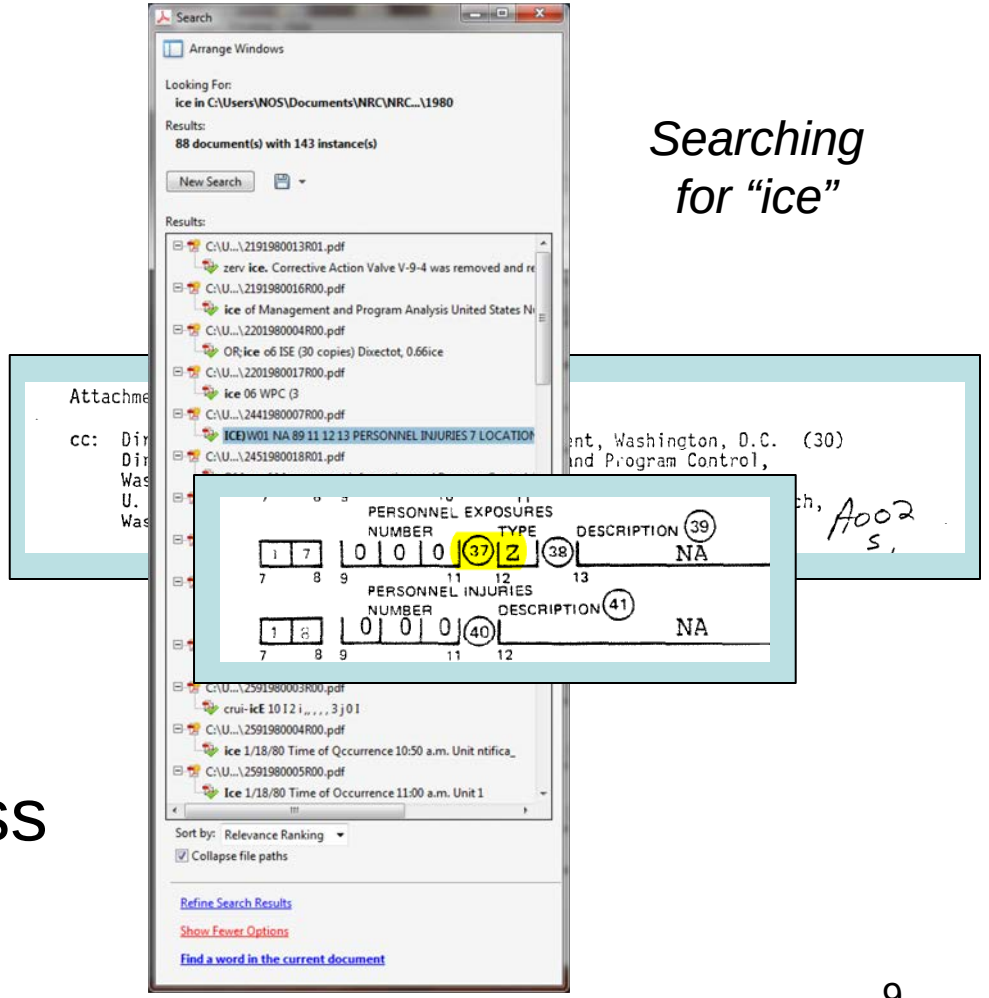
Study challenges

- Technology
  - Data reliability
  - “Out-of-the-box” capabilities vs. custom programming
  - Effective search patterns
- Multidisciplinary communication
  - Terminology
  - Problem-solving approaches
- Resources
- Study
  - Drawing broader lessons from use cases
  - Rapidly improving “conventional” tools

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

Technology challenge – data reliability

- Original document
  - Typewriting
  - Spelling errors
- Document metadata
  - Within document
  - Document system
- Scanning/OCR
- Inconsistencies across multiple documents

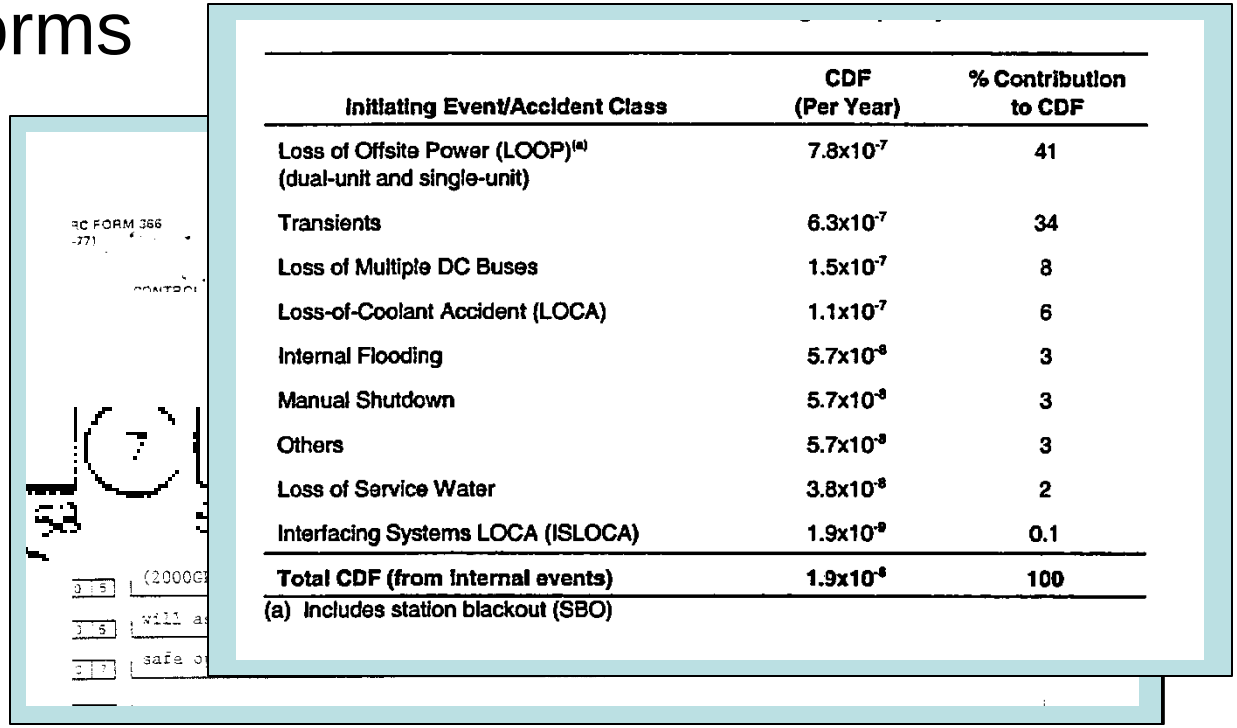


Searching for "ice"

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

Technology challenge – contextual indicators for text

- Graphical forms
- Tables



Initiating Event/Accident Class	CDF (Per Year)	% Contribution to CDF
Loss of Offsite Power (LOOP)[a] (dual-unit and single-unit)	7.8×10^{-7}	41
Transients	6.3×10^{-7}	34
Loss of Multiple DC Buses	1.5×10^{-7}	8
Loss-of-Coolant Accident (LOCA)	1.1×10^{-7}	6
Internal Flooding	5.7×10^{-8}	3
Manual Shutdown	5.7×10^{-8}	3
Others	5.7×10^{-8}	3
Loss of Service Water	3.8×10^{-8}	2
Interfacing Systems LOCA (ISLOCA)	1.9×10^{-9}	0.1
Total CDF (from internal events)	1.9×10^{-6}	100

(a) Includes station blackout (SBO)

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

Technology challenge – effective patterns

relationship to evaluated SAMAs; rationale for why the core damage frequency (CDF) for fire events would be substantially lower than reported in the IPEEE; results of a revised screening based on consideration of the potential impact of external events and uncertainties; more

The baseline CDF for the purpose of the SAMA evaluation is approximately 1.9×10^{-6} per year, and the baseline large early release frequency (LERF) is approximately 3×10^{-7} per year. The

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

Project challenge – need to know

- SME: how does the tool work?
  - Native inquisitiveness
  - More efficient support
- Software engineer: what do the terms mean?
  - Term uniqueness (and ability to discriminate among potential hits)
  - Basis for SME's search-limiting heuristics

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

Observations and questions

- Tool (when completed) will likely help user explore data (and perhaps “discover” insights).
  - Can effectively reduce search set – completeness needs to be evaluated.
  - “Right” role of user?
  - Reduced learning benefits from too much help?
- Tool relies on manually-developed rules (i.e., it does not automatically develop rules by reasoning from examples).
  - Effort justified? (One-time/infrequent searches vs. routine)
  - Learning value from supporting rule development?

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

Planned next steps

- Complete mult-unit and CDF use cases
- Initiate CCF use case
- Initiate formal methods review

Planned project completion: September, 2015