Experiences Developing an IBM Watson Cognitive Processing Application to Support Q&A of Application Security (Software Assurance) Diagnostics

SEI staff:

Mark Sherman (PI)

Lori Flynn (Tech lead)

Chris Alberts (Assurance SME)

Students:

Christine Baek

Anire Bowman

Skye Toor

Myles Blodnick

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM-0004026

Can DoD Use IBM Watson to Improve Assurance?

- Acquisition programs generate voluminous documentation
- Assurance is based on assembling and reviewing relevant evidence from documents
- Finding appropriate evidence or explanations can be challenging

Q : Can typical developers build IBM Watson applications to support an assurance review?

Key Take Aways

- You do not need a PhD in AI or Natural Language Processing to build IBM Watson applications on BlueMix
- Significant automation will be required for corpus (knowledge database) preparation, potentially larger than application development
- Subject matter expert needed to help craft document structure
- End user involvement needed to guide and improve training
- IBM Watson is one of many tools to bring to bear for cognitive processing applications

Approach: Simulate a Development Process

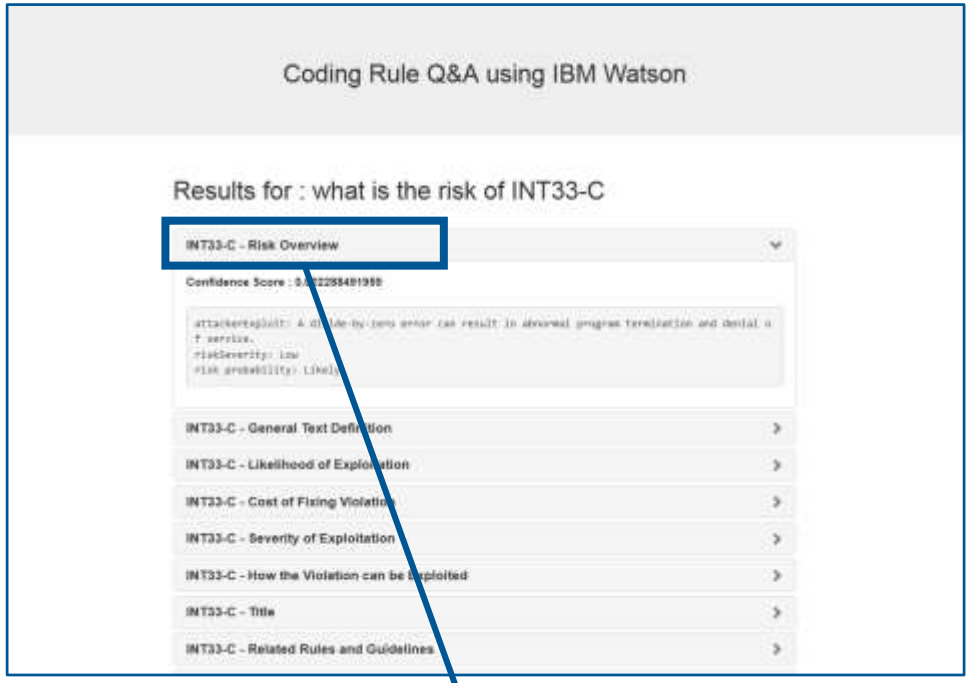
Assemble team of assurance experts
- Determine interesting questions
- Select appropriate documents
- Define training (ground truth)

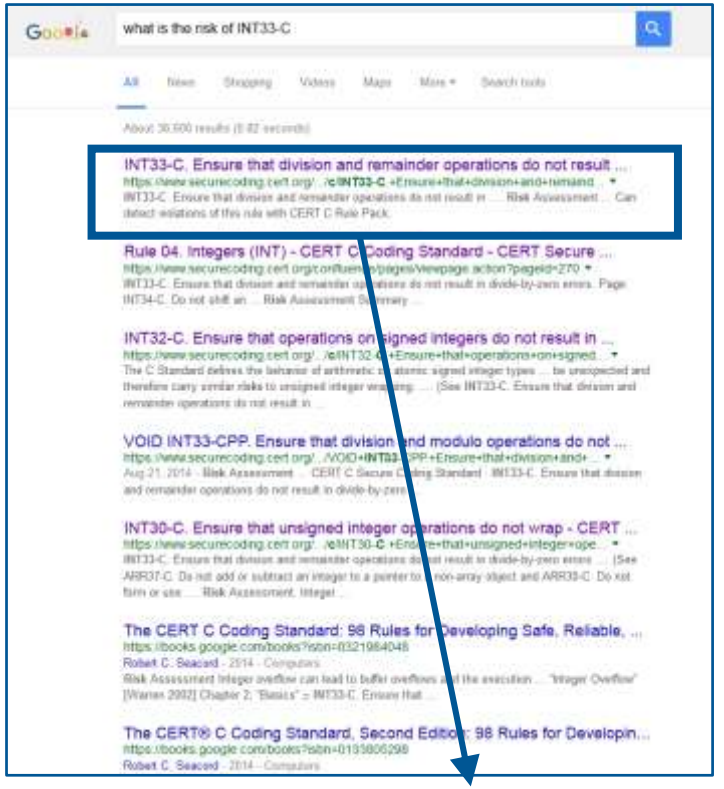
Assemble team of developers
- Experienced Python programmers
- No specific expertise in artificial intelligence or natural language processing

Two phases
- 2 weeks of 3 SMEs specifying corpus
- 11 weeks of 4 student developers building application and corpus

Application Performance

Better **Recall** and **Precision**:
Example: “What is the risk of INT33-C”





INT33-C – Risk Overview

INTC33-C. Ensure that division and remainder operations do not result …
https://www.securecoding.cert.org/…/c/INT33-C. =Ensure+that+dividion+and+remaind…

Lessons Learned From Project

Theory

Practice

Disposition of Materials

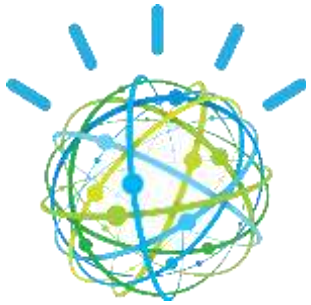
Government use rights apply. IBM Watson software (and any dependencies) must be licensed from IBM.

sparkcognition™

SparkCognition is an IBM Watson business partner (independent software vendor) and has licensed the project materials from CMU for use in their products.

We Want to Thank and Acknowledge Collaborators

SparkSecure team at SparkCognition



IBM Watson team at IBM

Prof. Eric Nyberg, Language Technologies Institute, School of Computer Science, CMU

And our student interns: Christine Baek, Anire Bowman, Skye Toor and Myles Blodnick

Contact Information

Mark Sherman

Technical Director, Cyber Security Foundations

Telephone: +1 412.268.9223

Email: mssherman@sei.cmu.edu