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

AI Enabling Technologies

Grooper and
Watson Content Analytics

June 29, 2021

Overview

What is Grooper?

Software that provides "*Thrilling Automation with Intelligent Document Processing*"*

Use Case: Extraction of data from operator licensing (OL) applications

Forms:

- NRC Form 396 (Certification of Medical Examination by Facility Licensee)
- NRC Form 398 (Personal Qualification Statement – Licensee)

Interfaces:

- Electronic Information Exchange (document ingestion)
- Reactor Program System (authoritative OL data source)

Grooper

NRC Grooper Features

1. Capture Tool

De-skew, Brighten, etc.

2. Image Processing

Optical Character Recognition (OCR)

5. Extraction

Parse and extract data, and write in XML schema

Other features used:

Optical Mark Recognition (OMR) – Recognizes checkmarks

Fuzzy Logic – Dictionary of defined values that can be OCRed or extracted based on a confidence threshold

Grooper

AI Grooper Features

Natural Language Processing and **Machine Learning** finds paragraphs, sentences, or other language elements in documents based on contextual meaning.

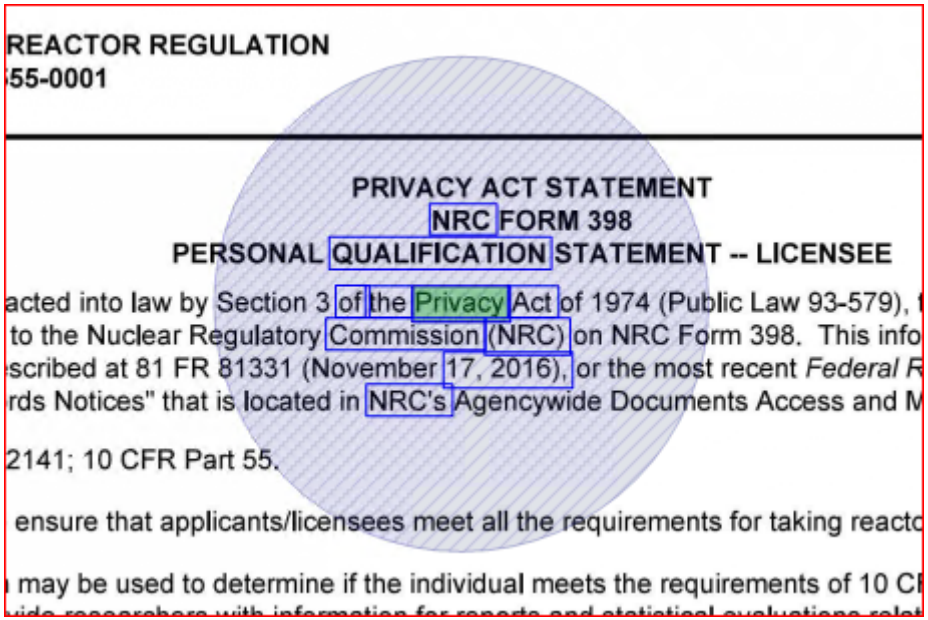
Use Case: Document sensitivity

Method:

- Manually review documents for sensitive keywords and identify true positives (in Grooper client)
- Start to train Grooper to contextually search around area of true positive
- Repeat with several document samples until properly trained

Grooper

AI Grooper Features Cont'd*



Context Scope

Type: *ContextScopeEnum*, ***Default:*** *Zonal*

Determines the scope of context feature extraction. Can be one of the following values:

- **Zonal** - Context features will be extracted from one or more zones, specified relative to the data value.
- **Flow** - Context will include a limited number matching features before and/or after the data value in the text flow.
- **Self** - Context will include all matching features which occur inside of or overlap with the data value.
- **Nearest** - Context will include a limited number of features which are closest to the data value.

Overview

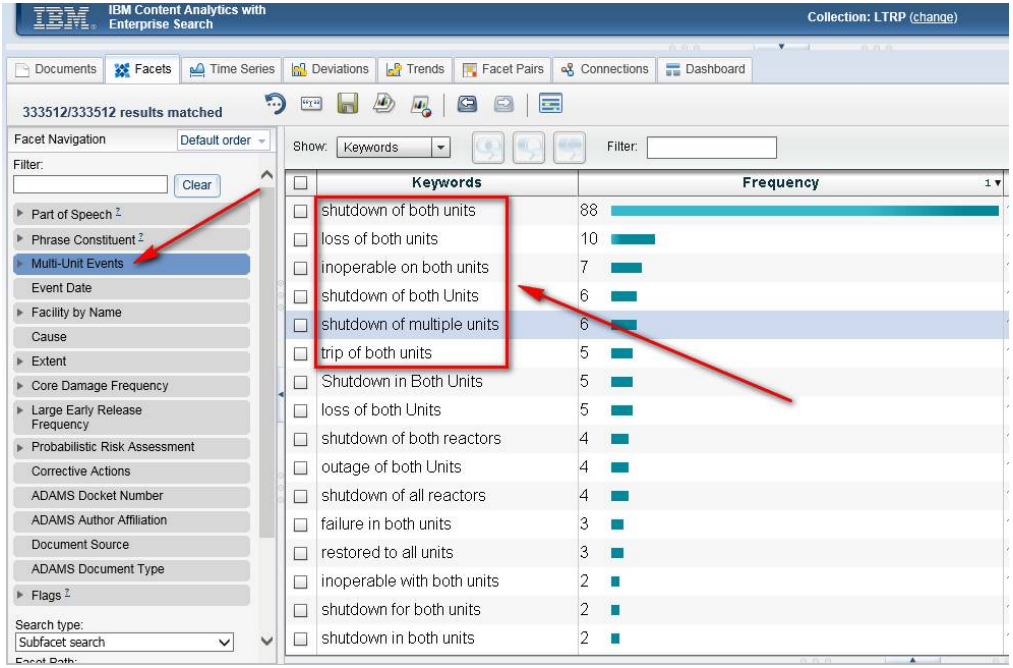
What is Watson Content Analytics?

Software that extrapolates business information from large collections of documents and uses natural language processing to uncover meaningful business insights.

Use Case: RES - Identify Event Reports that included an outage of two or more units

NLP Method:

- Define noun/verb combinations and NLP automatically contrives derivations of those combinations



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

References

References (Indicated by an *)

BIS, Inc. (2020-2021). AI-Powered Data Integration. Retrieved from https://www.bisok.com/.