

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
OCR'd 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*

REACTOR REGULATION
55-0001

PRIVACY ACT STATEMENT
NRC FORM 398
PERSONAL QUALIFICATION STATEMENT -- LICENSEE

acted into law by Section 3 of the Privacy Act of 1974 (Public Law 93-579), and added to the Nuclear Regulatory Commission (NRC) on NRC Form 398. This information is described at 81 FR 81331 (November 17, 2016), or the most recent "Federal Register Notices" that is located in NRC's Agencywide Documents Access and Management System (ADAMS) under accession number NRC-2016-012141; 10 CFR Part 55.

ensure that applicants/licensees meet all the requirements for taking reactor licenses. This information may be used to determine if the individual meets the requirements of 10 CFR Part 55.11. This information is provided to researchers with information for reports and statistical evaluations related to reactor operations.

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 of 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

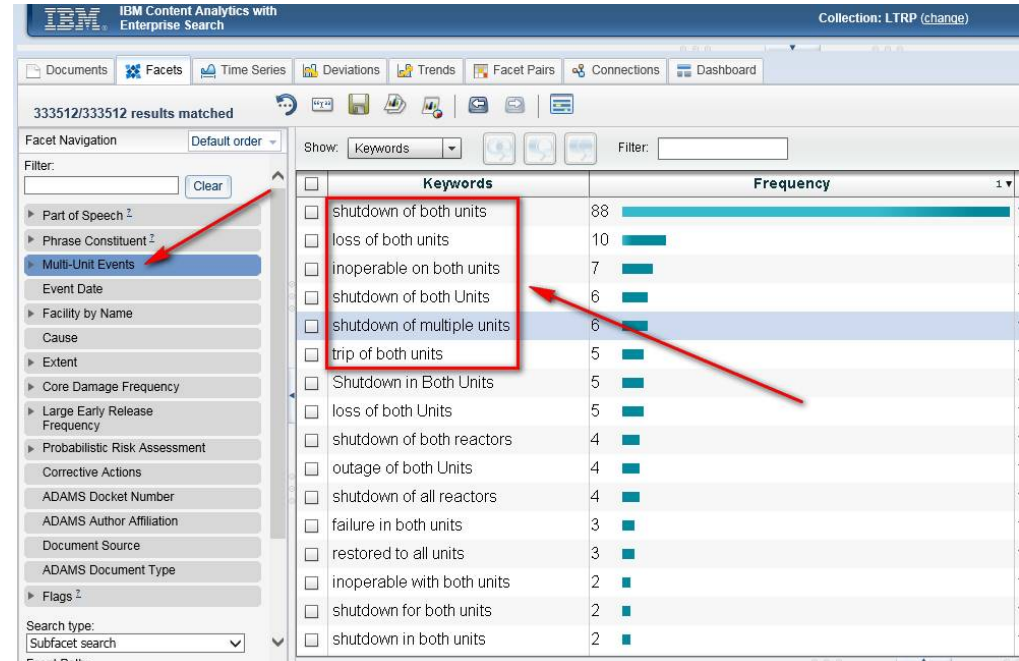
What is Watson Content Analytics?

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



References

References (Indicated by an *)

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