Emoveo—Design Document

Objectives: The objective is to allow the user to generate a redacted document that maintains readability, structure, and semantic/pragmatic meaning.
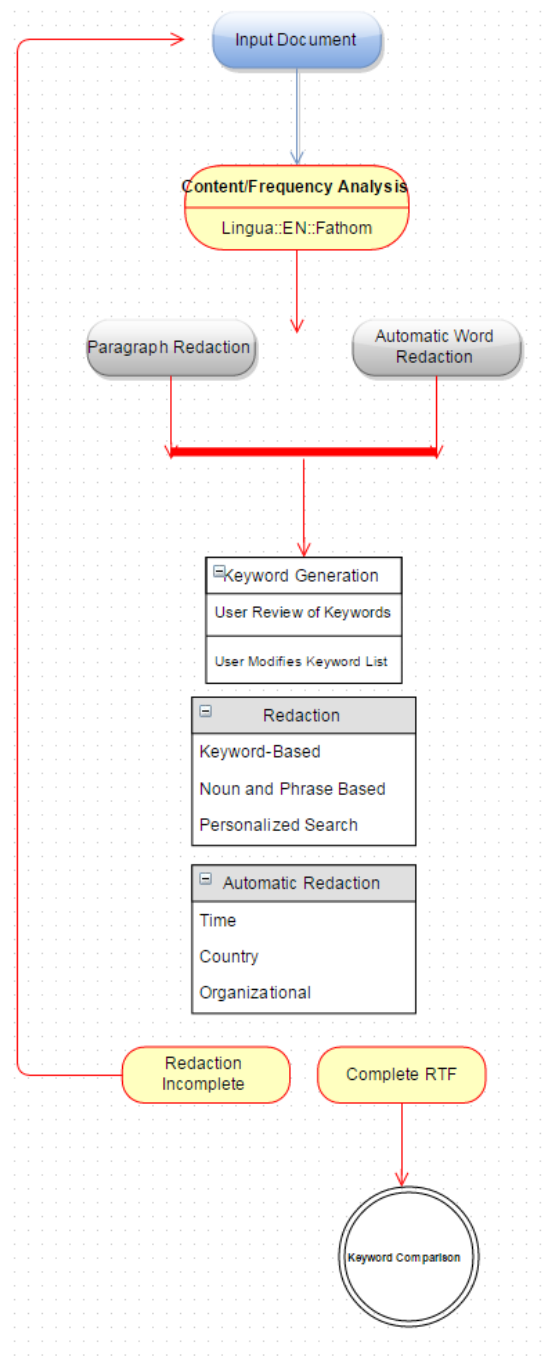
End Users:

- Lawyers: These users have the role to remove unrelated case information from subpoena documents and may have a vested interest in redacting classified documents
- Government/Private Sector Employees: The users may have a vested interest or have the security clearance level which information they create are classified and need to declassify these documents when due for mandatory redaction or normal redaction.
- Media: These users may have an ethical responsibility to redact leaked documents before release to public.
- General Public: The users have a right to request declassified and publicly available redacted information with limitations under federal right to information laws. They may also access publicly available documents which have been redacted.

Use Cases:

- Lawyers
    - o Redact
        - ▪ Actors: Lawyer
        - ▪ Pre-condition: Lawyer has subpoena or classified document
        - ▪ Trigger action: Actors use redaction tool to remove case-irrelevant information from subpoena document or the classified document.
    - o Review
        - ▪ Actors: Lawyer, Clerks, Judges
        - ▪ Pre-condition: Emoveo generated documents
        - ▪ Trigger action: Actors are able to review every step of the method used to create the pre-condition for accuracy and appropriateness.
- Government/Private Sector Employees
    - o Redact
        - ▪ Actors: Government/Private Sector Employees
        - ▪ Pre-condition: Classified, Secret, or Top-Secret document
        - ▪ Trigger action: Actors are able to use redaction tool to selectively remove elements from the pre-condition document.
    - o Review
        - ▪ Actors: Government/Private Sector Employees
        - ▪ Pre-condition: Emoveo generated documents
        - ▪ Trigger action: Actors are able to review every step of the method used to create the pre-condition for accuracy and appropriateness.

- Media
  - Redact
    - Actors: Media Employees
    - Pre-condition: Classified, Secret, or Top-Secret leaked document
    - Trigger action: Actors are able to use redaction tool to selectively remove elements from the pre-condition document.
  - Review
    - Actors: Media Employees
    - Pre-condition: Emoveo generated documents
    - Trigger action: Actors are able to review every step of the method used to create the pre-condition for accuracy and appropriateness prior to public release.
- General Public
  - View Document
    - Actors: General Public
    - Pre-condition: Access a redacted document
    - Trigger action: Can make a request for a redacted and declassified document with limitations stated in Freedom of Information Act (FOIA).

Workflow/Features:

```
                    ┌──────────────────┐
                    │  Input Document  │
                    └──────────────────┘
                             │
                             ▼
                 ╔═══════════════════════╗
                 ║ Content/Frequency Analysis ║
                 ║                        ║
                 ║  Lingua::EN::Fathom    ║
                 ╚═══════════════════════╝
                             │
            ┌────────────────┴────────────────┐
    ┌───────────────────┐          ┌───────────────────┐
    │ Paragraph Redaction│          │  Automatic Word   │
    │                    │          │    Redaction      │
    └───────────────────┘          └───────────────────┘
            │                                │
            └────────────┬───────────────────┘
                         ▼
```

⊟ Keyword Generation
User Review of Keywords
User Modifies Keyword List

⊟ Redaction
Keyword-Based
Noun and Phrase Based
Personalized Search

⊟ Automatic Redaction
Time
Country
Organizational

```
    ┌───────────────┐          ┌───────────────┐
    │   Redaction   │          │  Complete RTF │
    │  Incomplete   │          │               │
    └───────────────┘          └───────────────┘
                                       │
                                       ▼
                                 ╔═══════════╗
                                 ║  Keyword  ║
                                 ║ Comparison║
                                 ╚═══════════╝
```

Requirements:

- There must be internet access.
- The following modules must be installed and up-to-date:
  - Data::Dump
  - Encode
  - File::Slurp
  - HTML::FormatText
  - HTML::TreeBuilder
  - Lingua::Concordance
  - Lingua::EN::Fathom
  - Lingua::EN::StopWords
  - Lingua::EN::Tagger
  - Lingua::Stem::En
  - List::Compare
  - Locale::Country
  - Locale::Language
  - LWP::Simple
  - Path::Class
  - RTF::Writer
  - Text::TermExtract
  - Tkx
  - WWW::Wikipedia
- Read, write, and execute access necessary for emoveo and input documents.
- Must meet memory and CPU requirements to run the perl script:
  - Disk Space: 30 MB free (min)
  - Processor: 1 GHz (min)
  - Memory: 300 MB (min)
- Documents to redact must be in English with proper paragraph formatting.