CLAIMS

What is claimed is:

1. A computer-implemented method of creating object models for regulations, the method comprising:

accessing, by a computer processor, a set of regulatory information corresponding to a regulation;

segmenting, by the computer processor, the set of regulatory information into a set of structured texts and a set of metadata;

generating, by the computer processor, an object model for the regulation, the object model comprising the set of structured texts and the set of metadata;

performing, by the computer processor, a linguistic analysis on the object model to detect a set of sentences within the set of structured texts;

ranking, by the computer processor, the set of sentences according to a relevancy of the set of sentences to the regulation;

extracting, by the computer processor, a portion of the set of sentences that are relevant to the regulation according to the ranking of the set of sentences;

generating, by the computer processor from the portion of the set of sentences that was extracted, a summary of the regulation using a machine learning algorithm that yields a summarization model from a corpus of regulatory training data; and

enriching, by the computer processor, the object model for the regulation with the summary of the regulation.

2. The computer-implemented method of claim 1, further comprising:

training, by the computer processor using the summary of the regulation, at least one of a classification model or an entity recognition model;

determining, by the computer processor using the classification model, a set of topics for any portion of the object model for the regulation; and

enriching, by the computer processor, the object model for the regulation with the set of topics for the portion of the object model.

3. The computer-implemented method of claim 2, further comprising:

determining, by the computer processor for each topic in the set of topics assigned to the portion of the object model, a probability that the topic is applicable to that portion.

4. The computer-implemented method of claim 3, further comprising:

outputting at least a portion of the object model having at least one topic of the set of topics with a probability that at least meets a specified threshold.

5. The computer-implemented method of claim 2, further comprising:

analyzing, by the computer processor, a portion of the set of structured texts within the object model using an entity recognition analysis to determine a set of entity names meaningful to the regulation; and

enriching, by the computer processor, the object model with the set of entity names.

6. The computer-implemented method of claim 5, further comprising:

outputting at least a portion of the set of structured texts within the object model corresponding to a set of keywords or phrases; and

enriching, by the computer processor, the object model with the set of keywords or phrases.

7. The computer-implemented method of claim 1, wherein performing the linguistic analysis on the object model comprises one of:

detecting the set of sentences and generating a set of token n-grams for each sentence in the set of sentences; or

detecting the set of sentences from a set of n-grams tokenized from the set of structured texts.

8. The computer-implemented method of claim 1, wherein segmenting the set of regulatory information comprises:

segmenting the set of regulatory information into one or more of header, footer, title, body,

sections, sub-sections, paragraphs, lists, sub-lists, citations, or references, wherein the set of metadata comprises a set of topic labels.

9. The computer-implemented method of claim 1, further comprising:

examining, by the computer processor, the set of regulatory information to determine that a set of topic labels is not present;

generating, by the computer processor, a set of applicable topic labels for the set of regulatory information; and

storing the set of applicable topic labels with the set of metadata.

10. The computer-implemented method of claim 1, further comprising:

storing the object model for the regulation along with at least one of a historical version or a consolidated version of the regulation.

11. A system for dynamically creating object models for regulations, the system comprising:

a memory storing instructions; and

a processor interfaced with the memory and configured to execute the instructions to cause the processor to:

access a set of regulatory information corresponding to a regulation,

segment the set of regulatory information into a set of structured texts and a set of metadata,

generate an object model for the regulation, the object model comprising the set of structured texts and the set of metadata,

perform a linguistic analysis on the object model to detect a set of sentences within the set of structured texts,

rank the set of sentences according to a relevancy of the set of sentences to the regulation,

extract a portion of the set of sentences that are relevant to the regulation according to the ranking of the set of sentences,

generate, from the portion of the set of sentences that was extracted, a summary of the regulation using a machine learning algorithm that yields a summarization model from a corpus of regulatory training data, and

enrich the object model for the regulation with the summary of the regulation.

12. The system of claim 11, wherein the processor is further configured to:

train, using the summary of the regulation, at least one of a classification model or an entity recognition model,

determine, using the classification model, a set of topics for any portion of the object model for the regulation, and

enrich the object model for the regulation with the set of topics for the portion of the object model.

13. The system of claim 12, wherein the processor is further configured to:

determine, for each topic in the set of topics assigned to the portion of the object model, a probability that the topic is applicable to that portion.

14. The system of claim 13, wherein the processor is further configured to:

output at least a portion of the object model having at least one topic of the set of topics with a probability that at least meets a specified threshold.

15. The system of claim 12, wherein the processor is further configured to:

analyze a portion of the set of structured texts within the object model using an entity recognition analysis to determine a set of entity names meaningful to the regulation, and

enrich the object model with the set of entity names.

16. The system of claim 15, wherein the processor is further configured to:

output at least a portion of the set of structured texts within the object model corresponding to a set of keywords phrases, and

enrich the object model with the set of keywords or phrases.

17. The system of claim 11, wherein to perform the linguistic analysis on the object model, the processor is configured to one of:

detect the set of sentences and generate a set of token n-grams for each sentence in the set of sentences; or

detect the set of sentences from a set of n-grams tokenized from the set of structured texts.

18. The system of claim 11, wherein to segment the set of regulatory information, the processor is configured to:

segment the set of regulatory information into one or more of header, footer, title, body, sections, sub-sections, paragraphs, lists, sub-lists, citations, or references, wherein the set of metadata comprises a set of topic labels.

19. The system of claim 11, wherein the processor is further configured to:

examine the set of regulatory information to determine that a set of topic labels is not present,

generate a set of applicable topic labels for the set of regulatory information, and

store the set of applicable topic labels with the set of metadata.

20. The system of claim 11, wherein the processor is further configured to:

store the object model for the regulation along with at least one of a historical version or a consolidated version of the regulation.