

DEFINITIONS EXTRACTIONS FROM CODE OF FEDERAL REGULATIONS

Mohamma M. AL Asswad, Deepthi Rajagopalan, Neha Kulkarni– Department of Computer Science, Affiliation: LII, Cornell Law School



Overview

CFR is the codification of all regulations in the USA.

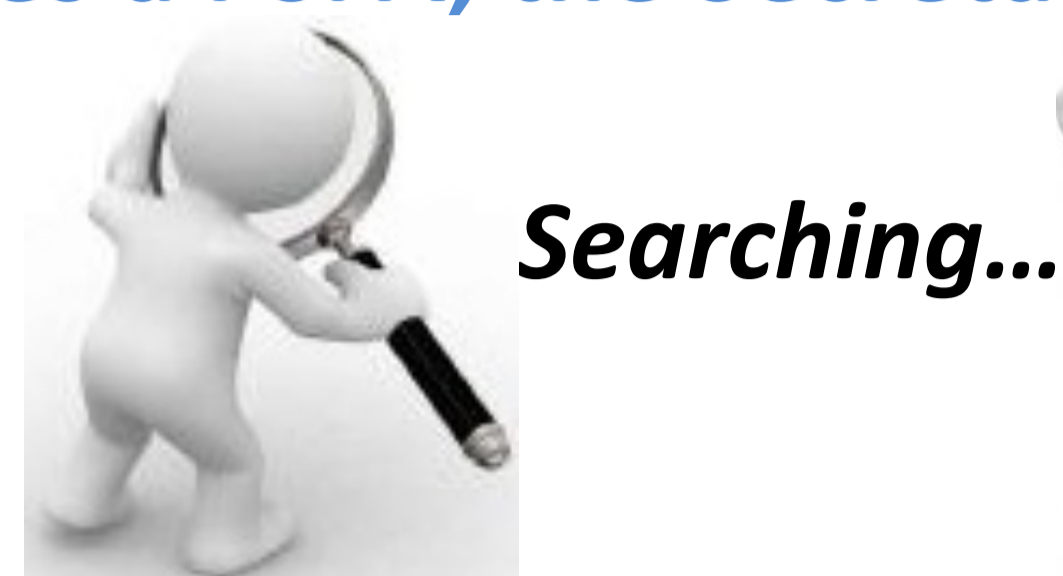
There are 50 titles in CFR categorized according to subject matter.

Our project focuses on extracting definitions from Title 7 (Agriculture) but testing on other titles revealed that our approach is generic.

Problem: A person reading the following text in Title 42 of CFR.

If the Secretary immediately reassumes a PSFA, the Secretary must provide the Self-Governance Tribe

Which secretary?
Can I find its definition?
Okay. Let me search in CFR...



Searching...



Found it!!!

Secretary means the Secretary of Health and Human Services to whom the authority involved may have been delegated



OOPS... Another definition for Secretary. Which one is applicable in this context???

Secretary. The Secretary of Agriculture of the United States or any other officer or employee of the U.S. Department of Agriculture

Solution

The Definition Extraction Team Says:

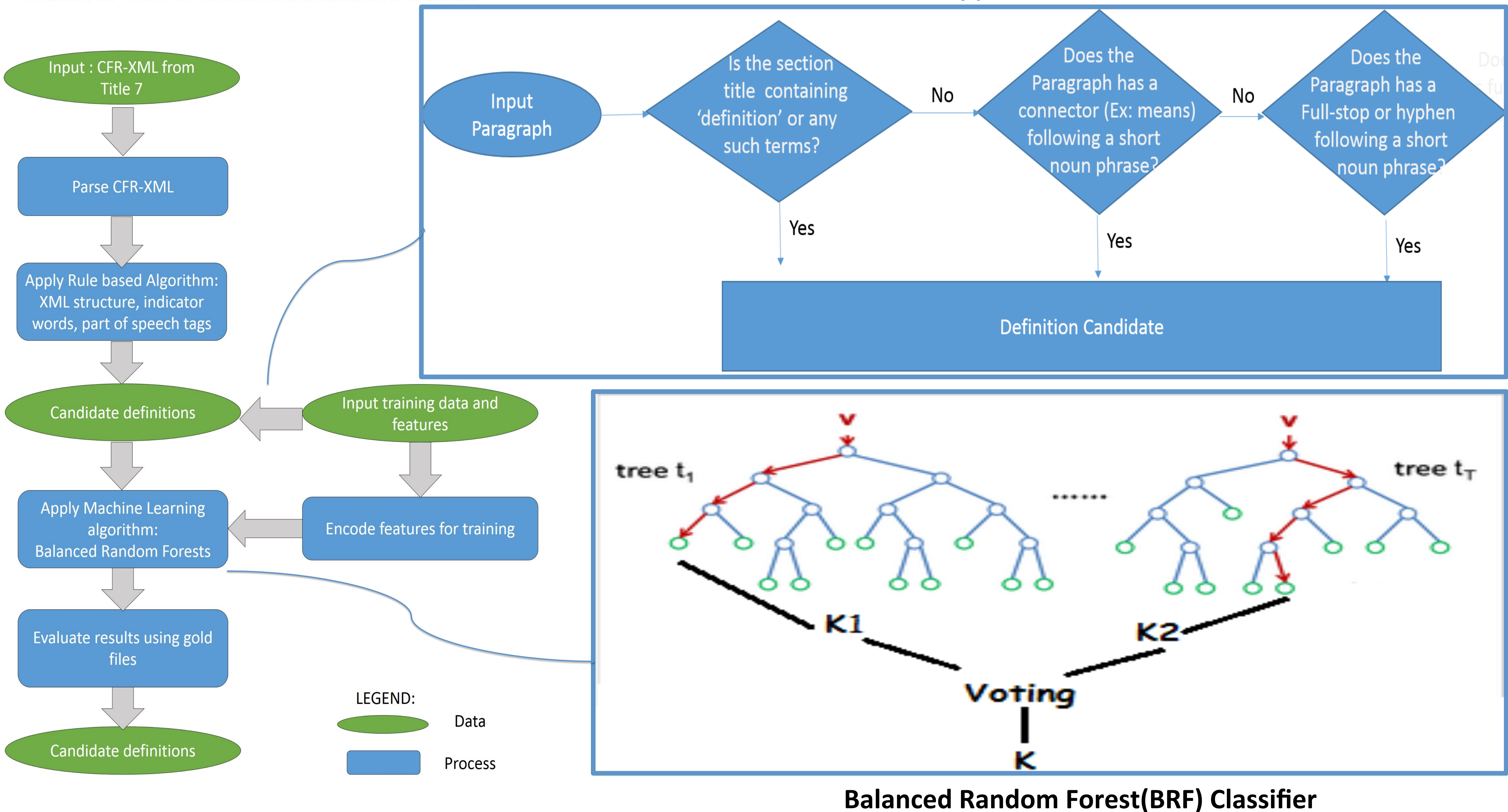
Don't worry. We will identify and extract the definitions from CFR corpus, and link them to the defined terms according to their scope



The team decides on the approach:

- 1) Do XML Parsing to extract paragraphs that can be potential definitions. Include their scope.
- 2) Apply well-defined grammar rules on the extracted paragraphs to narrow down the definition candidates
- 3) Train a Machine Learning Classifier(Balanced Random Forest) to classify the candidates as definitions and non-definitions

DESIGN OF THE DEFINITION EXTRACTION APPROACH



Results and Futurework

	Rulebased	Rule based + ML
Precision	69%	84%
Recall	96%	75%
F-Measure	80.5%	79%
Accuracy	90%	92%

- Use classifier to generate connectors
- Use POS Tagging for classifier
- Extend to other Titles and legal corpora.
- Publish research paper (In Progress)

Where is the "Boom"? First-ever project to provide an elegant working solution for Definition Extraction from CFR. Other projects in the same domain but for different corpora such as US Code are being built using our efficient framework. Our experimental approach that combines grammar and ML has led to an increased accuracy compared to existing algorithms relying on only one of these. In addition, using BRF owing to the imbalanced dataset has improved our results further.