Auto Annotation of "Definition" in EU Legislative Text Files Using Symbolic AI

Author`s:

Muhammad Asif
ALMA-AI, University of Bologna, Italy
University of Luxembourg, Luxembourg

Prof. Monica Palmirani
ALMA-AI, University of Bologna, Italy

Joint International Doctoral (Ph.D.) Degree in Law, Science and Technology

Objectives

- To detect the legal definitions in the EU legislation.

- To annotate the legal definitions in the EU legislation using Akoma Ntoso OASIS standard (LegalDocML TC).

- Akoma Ntoso is an XML international standard adopted officially by the EU authorities (AKN4EU) for modeling the legislation, parliamentary documents and case-law.

- The annotation of AKN help the LLM fine-tuning in the legal domain.

Definitions

- A definition is "a phrase signifying a thing's essence" defined by Aristotle.

- It can be defined as, "the degree of distinctness in the outline of an object".

- The definition is conceived as commitments which needs supported arguments.

- Legal Definitions are one of the more important legal instruments for creating new legal concepts (e.g., mobbing, stalking, e-commerce).

- In the LLM era the definitions are fundamental for making fine-tuning in legal domain.

Legal Definitions

A *Definition* is composed of the following four components:

- **Definiendum**: (is the subject of the Definition, called *term*)

- **Definitor**: (is a verb phrase which is used to introduce the Definition - *means*)

- **Definiens**:(phrase which is used to define something in the Definition)

- **Conditions**: (conditions connected with situations, temporal parameters, geographic areas, etc.).

(a) "ship" means any seagoing vessel, whether publicly or privately owned, which is ordinarily engaged in commercial maritime operations. Fishing vessels are not included in this definition,

Definition`s in Legislations

In Legislation, definitions are not needed to define a simple or less technical terms. It only needed for the following reasons in Legislative text.

- To avoid any ambiguity regarding the meaning of the word.
- To explain the meaning of the new or unusual word.
- To make the text shorter by avoiding repetition.
- To define new legal concepts.
- To simplify the legal interpretation.

Types of Definitions in Legislation:

Delimiting definitions   Extending definitions

Narrowing Definitions   Mixed Definitions
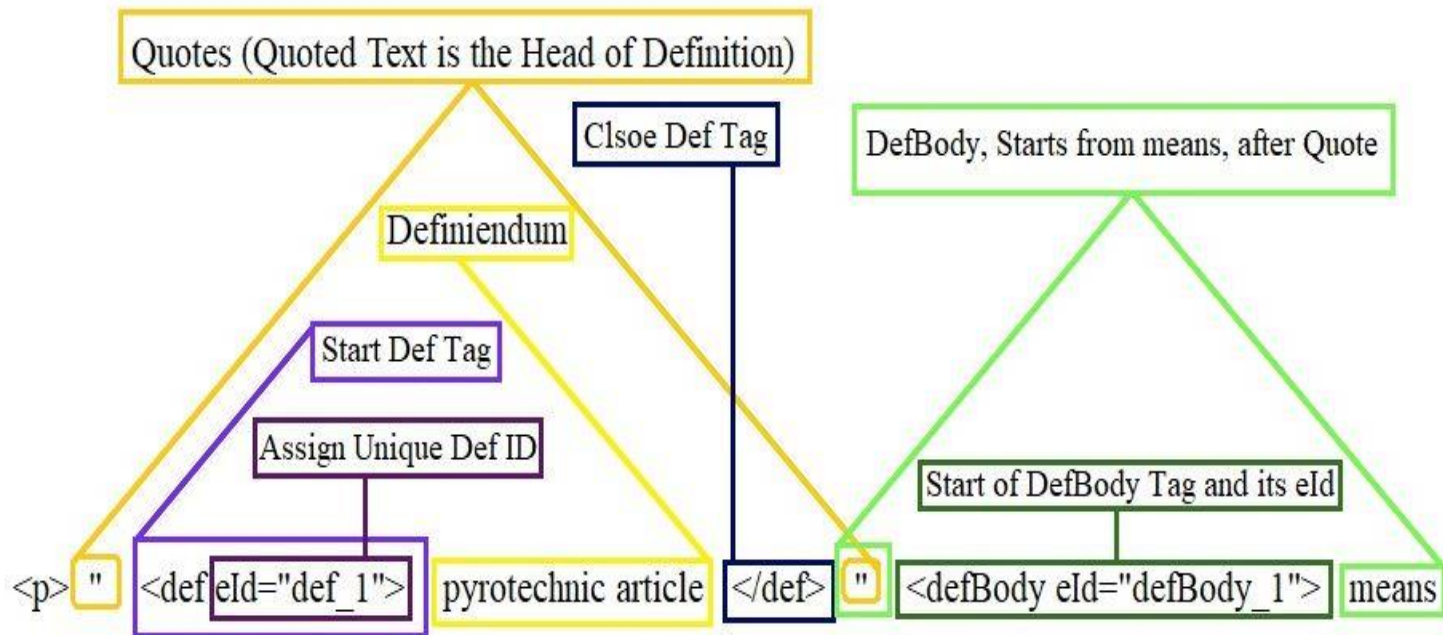
Count-As Definitions

Definition Extraction

- To detect the inconsistencies in the definitions:
  - There is a need to extract the definitions text from the legal documents.
  - After definition extraction the annotation of the definition is performed.

- The detection of definition using Natural Language Processing Techniques i.e., Rule Base Mining

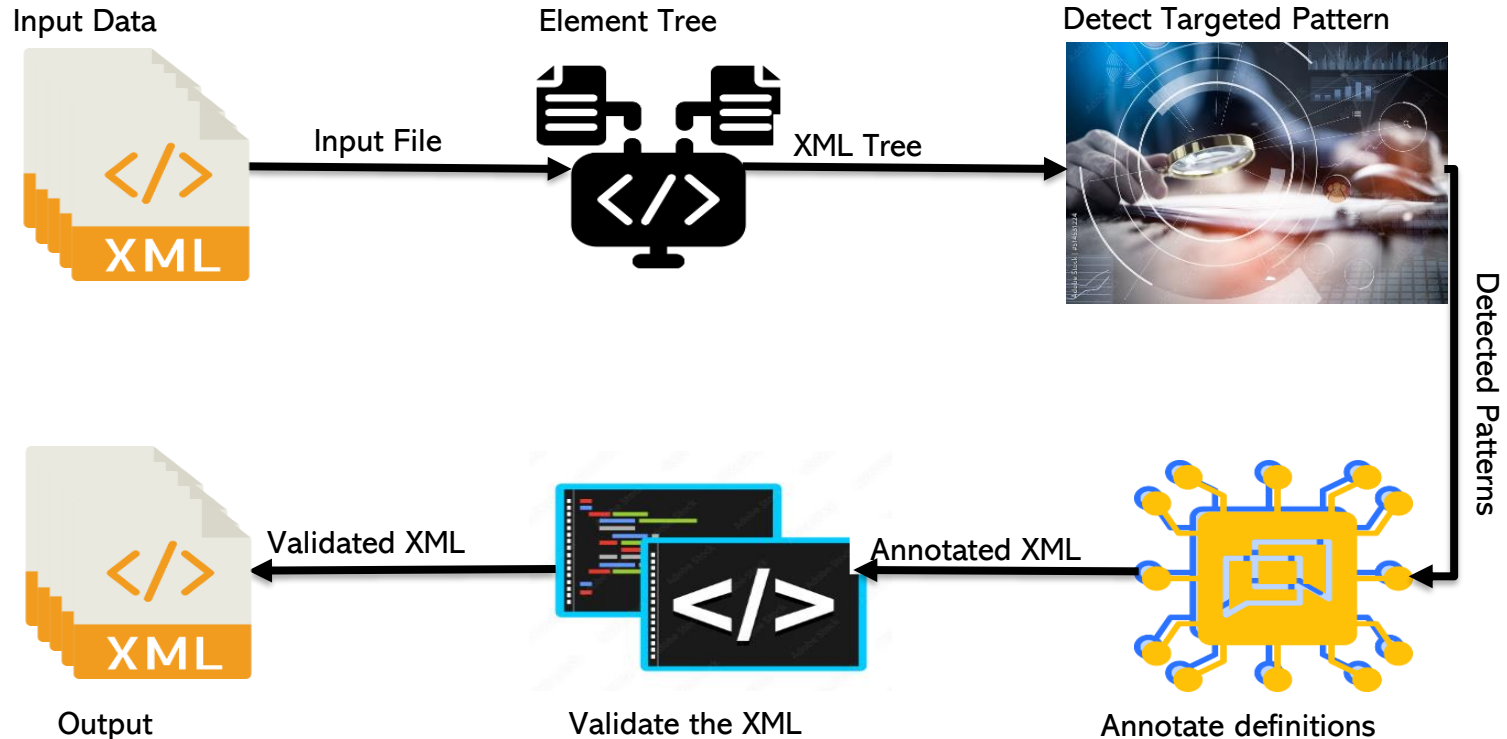- "means" is usually the delimitation verbs (**Definitor)**

Definition`s Annotation

Quotes (Quoted Text is the Head of Definition)

Clsoe Def Tag

DefBody, Starts from means, after Quote

Definiendum

Start Def Tag

Assign Unique Def ID

Start of DefBody Tag and its eId

`<p>` `"` `<def eId="def_1">` pyrotechnic article `</def>` `"` `<defBody eId="defBody_1">` means

any article containing explosive substances or an explosive mixture of substances designed to produce heat, light, sound, gas or smoke or a combination of such effects through self-sustained exothermic chemical reactions; `</defBody>` — DefBody Closing Tag (which closes at ;)

How to Perform Annotation

Methodology

Input Data

Input File

Element Tree

XML Tree

Detect Targeted Pattern

Detected Patterns

Validated XML

Annotated XML

Output

Validate the XML

Annotate definitions

Dataset: EU Legislation for Agri-Food in the span of time from 2010 to 2021 - 15.000 documents.

Algorithm for Annotation

```
//all definitions and tags that will be generated and stored in the file
will also be stored in a linked list so that all information can be stored
in tag i.e. analysis at the end of fille

Create TagsList
If (fileAsPerRules(file, rules) = True)
   While (!endOfFile())
      If (findArticleTag() = True) ;
         enter_into_article_tag();
         If (findDefinitionTag() = True) ;                            //¹
            enter_into_definition_tag();
            Term = findQuatedText () ;                                //²
            If (Term != NULL)
               DefBody_text = find_defBody_text_after_means()      //³
               If (DefBody_text!= NULL)
                  uniqueEID=generateUniqueEID ()                      //⁴
                  AddDefTag (uniqueEID,Term)                          //⁵
                  AddDefBodyTag (uniqueEID,Def_body_text)            //⁶
                  Count_sub_list=0
                  If (sub_list_exist()=true)
                     While (! endOfList_Tag())                        //⁷
                        LocalListId=generateLocalListId(UniqueEID)//⁸
                        Count_sub_list=count_sub_list+1
                  TagsList[i].add(uniqueEID, Term,count_sub_lists)
AddMetadataToAnalysisTag();                                          //⁹
```

//¹ this function will return true if a tag titled "Definition" or "Definitions" is found or contains in text
//² this function will search and return a Term / string that is in double quotes, this is a loop that will continue its search until it finds next article
//³ this function finds the word means and then returns all the text next with it
//⁴ e.g. 1
//⁵ <def eId="def_1"> Term going to be defined </def>
//⁶ <defBody eId=defBody_1> def body text </defBpdy>
//⁷ this is search </List> and will terminate this loop
//⁸ add <defBody eId=defBody_1-1> def body text </defBody> the list of defBody>
//⁹ add all TagsList items in analysis tag with proper formatting (analysis source="#cirsfid", definitions source="#unibo", definition refersTo="#term", definitionHead href="#def_1" refersTo="#term", definitionBody href="#defBody_1" and definitionBody href="#defBody_1-1") if list contains no data then print "File not following rules so cannot be annotated".

The Pseudocode for Annotation of Definitions

Find the Targeted Pattern

In AKN, first find the heading tag (τ_H) having targeted keywords, "definition" or "definitions" ($k_{Definition}$ or $k_{Definitions}$). Enter into τ_H and find the paragraph tag (τ_ρ). After finding τ_ρ enter body of τ_ρ and find the keyword "means" (k_{means}) being followed by some quoted text. Store this quoted text ($Q_{Definiendum}$). Annotate $Q_{Definiendum}$ as per rules defined in equation 1, and then replace it with new quoted text ($Q'_{Definiendum}$).

$$Q'_{Definiendum} = <\textbf{def } \textbf{eId} = \textbf{def\_i} > Q_{Definiendum} </\textbf{def}> \tag{1}$$

Next, store the rest of the text in the body of τ_ρ i.e., T_{Next}. Annotate T_{Next} as per the rules defined in equation 2 and then replace T_{Next} with T'_{Next}.

$$T'_{Next} = <\textbf{defBody } \textbf{eId} = \textbf{"defBody}_{-i} > T_{Next} </\textbf{defBody}> \tag{2}$$

Where *i* is the unique numeric identifier that is assigned in this annotation, and it is the same in both equations.

Akoma Ntoso XML file Before Annotation

```xml
<num>Article 2</num>
<heading>Definitions</heading>
<list eId="art_2__list_1">
  <intro eId="art_2__list_1__intro">
    <p>In addition to the definitions set out in Article 2 of Directive 2010/30/EU, the
    following definitions shall apply for the purposes of this Regulation:</p>
  </intro>
  <point eId="art_2__list_1__point_1">
    <num>(1)</num>
    <intro eId="art_2__list_1__point_1__intro">
      <p>"water heater" means a device that:</p>
    </intro>
    <list eId="art_2__list_1__point_1__list_1">
      <point eId="art_2__list_1__point_a__list_1__point_a">
        <num>(a)</num>
        <content eId="art_2__list_1__point_a__list_1__point_a__content">
          <p>is connected to an external supply of drinking or sanitary water;</p>
        </content>
      </point>
      <point eId="art_2__list_1__point_1__list_1__point_b">
        <num>(b)</num>
        <content eId="art_2__list_1__point_1__list_1__point_b__content">
          <p>generates and transfers heat to deliver drinking or sanitary hot water at
          given temperature levels, quantities and flow rates during given intervals;
          and</p>
        </content>
      </point>
      <point eId="art_2__list_1__point_1__list_1__point_c">

        <num>(c)</num>
        <content eId="art_2__list_1__point_1__list_1__point_c__content">
          <p>is equipped with one or more heat generators;</p>
        </content>
      </point>
    </list>
  </point>
```

Content Before Annotation

Akoma Ntoso XML file Before Annotation

```xml
<num>Article 2</num>
<heading>Definitions</heading>
<list eId="art_2__list_1">
    <intro eId="art_2__list_1__intro">
        <p>In addition to the definitions set out in Article 2 of Directive 2010
        /30/EU, the following definitions shall apply for the purposes of this
        Regulation:</p>
    </intro>
    <point eId="art_2__list_1__point_1" defines="#def_1">
        <num>(1)</num>
        <intro eId="art_2__list_1__point_1__intro">
            <p> "<def eId="def_1">water heater</def>" <defBody eId="defBody_1">
            means a device that:</defBody> </p>
        </intro>
        <list eId="art_2__list_1__point_1__list_1">
            <point eId="art_2__list_1__point_a__list_1__point_a">
                <num>(a)</num>
                <content eId="art_2__list_1__point_a__list_1__point_a__content">
                    <p> <defBody eId="defBody_1-1">is connected to an external
                    supply of drinking or sanitary water;</defBody> </p>
                </content>
            </point>
            <point eId="art_2__list_1__point_1__list_1__point_b">
                <num>(b)</num>
                <content eId="art_2__list_1__point_1__list_1__point_b__content">
                    <p> <defBody eId="defBody_1-2">generates and transfers heat
                    to deliver drinking or sanitary hot water at given
                    temperature levels, quantities and flow rates during given
                    intervals; and</defBody> </p>
                </content>
            </point>
            <point eId="art_2__list_1__point_1__list_1__point_c">
                <num>(c)</num>
                <content eId="art_2__list_1__point_1__list_1__point_c__content">
                    <p> <defBody eId="defBody_1-3">is equipped with one or more
                    heat generators;</defBody> </p>
                </content>
            </point>
        </list>
    </point>
```

The Sample Annotated Definitions

Metadata of Annotation: N-M relationships

```xml
<definitions source="#unibo">
    <definition refersTo="#waterHeater">
        <definitionHead href="#def_1" refersTo="#waterHeater"/>
        <definitionBody href="#defBody_1"/>
        <definitionBody href="#defBody_1-1"/>
        <definitionBody href="#defBody_1-2"/>
        <definitionBody href="#defBody_1-3"/>
    </definition>
    <definition refersTo="#heatGenerator">
        <definitionHead href="#def_2" refersTo="#heatGenerator"/>
        <definitionBody href="#defBody_2"/>
        <definitionBody href="#defBody_2-1"/>
        <definitionBody href="#defBody_2-2"/>
        <definitionBody href="#defBody_2-3"/>
    </definition>
    <definition refersTo="#ratedHeatOutput">
        <definitionHead href="#def_3" refersTo="#ratedHeatOutput"/>
        <definitionBody href="#defBody_3"/>
    </definition>
    <definition refersTo="#standardRatingConditions">
        <definitionHead href="#def_4" refersTo="#standardRatingConditions"/>
        <definitionBody href="#defBody_4"/>
    </definition>
    <definition refersTo="#biomass">
        <definitionHead href="#def_5" refersTo="#biomass"/>
        <definitionBody href="#defBody_5"/>
    </definition>
    <definition refersTo="#biomassFuel">
```

Definition structured in multiple points in a list

Definition structured in multiple articles

Acknowledgments

This study is funded by PON grants of the Italian Government and also the ERC HyperModeLex.

HYPERMODELEX

erc
European Research Council
Established by the European Commission

Co-funded by
the European Union

Joint International Doctoral (Ph.D.) Degree in Law, Science and Technology

PON
RICERCA
E INNOVAZIONE

References

- D. T. H. Hutabarat, M. A. Efendi, M. Fatwa Str, and N. Prayoga, "Analyzing the Relationship Between Law and Technology," Policy, Law, Notary and Regulatory Issues (Polri), vol. 1, no. 2, pp. 99–110, 2022, doi: 10.55047/polri.v1i2.161.

- Löwe, Can Laurens. "Aristotle and John Buridan on." Oxford Studies in Medieval Philosophy Volume 6 77 (2018): 189.

- McGrath, Stephen Keith. "Defining Key Initial Terms." In Speaking Management, pp. 17-22. Springer, Singapore, 2021.

- Sovrano, Francesco, Monica Palmirani, and Fabio Vitali. "Deep learning based multi-label text classification of UNGA resolutions." In Proceedings of the 13th international conference on theory and practice of electronic governance, pp. 686-695. 2020.

- Palmirani, Monica, Francesco Sovrano, Davide Liga, Salvatore Sapienza, and Fabio Vitali. "Hybrid AI Framework for Legal Analysis of the EU Legislation Corrigenda." In Legal Knowledge and Information Systems, pp. 68-75. IOS Press, 2021.

- Palmirani, Monica, and Fabio Vitali. "Akoma Ntoso an open document standard for Parliaments." (2014).

- L. Weissweiler, V. Hofmann, M. J. Sabet, and H. Schütze, "CaMEL: Case Marker Extraction without Labels," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.10010

- E. Ferneda, H. A. do Prado, A. H. Batista, and M. S. Pinheiro, "Extracting definitions from brazilian legal texts," in International Conference on Computational Science and Its Applications, 2012, pp. 631–646.

Many Thanks for your attention

http://www.last-jd.eu/
Muhammad.asif19@unibo.it