HYPERMODELEX

European Research Council
Established by the European Commission

Co-funded by
the European Union

NLP, AI and LLM WITH LEGALXML

Michele Corazza – Researcher at CIRSIFD-ALMA AI, Department of Legal Studies, University of Bologna, Italy

ECAI 2025 – University of Bologna
Day 1 - October 25, 2025

Large Language Models and NLP

- The recent advancements in Natural Language Processing (NLP) have led to the creation of Large Language Models that can perform a multitude of tasks

- These are very powerful tools, however there is always a non-zero risk that the generated text contains errors (hallucinations)

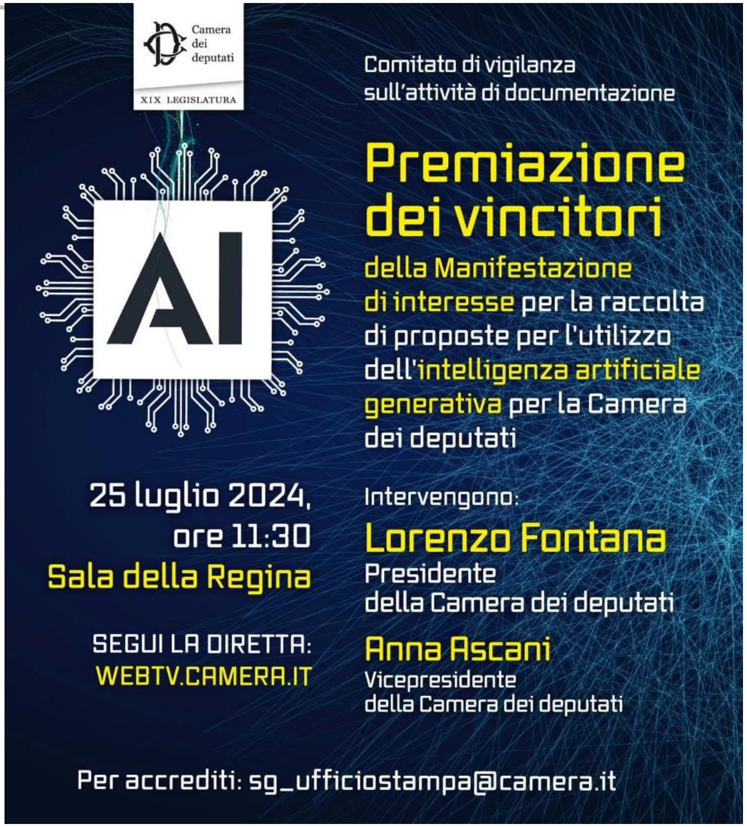
- This is particularly severe for the legal domain, which has some characteristics that complicate any AI/NLP/LLM application.

ENDORSE

THE EUROPEAN DATA CONFERENCE ON REFERENCE DATA AND SEMANTICS

European Commission

USE-CASES



Camera dei deputati
XIX LEGISLATURA

Comitato di vigilanza sull'attività di documentazione

Premiazione dei vincitori della Manifestazione di interesse per la raccolta di proposte per l'utilizzo dell'intelligenza artificiale generativa per la Camera dei deputati

25 luglio 2024,
ore 11:30
Sala della Regina

SEGUI LA DIRETTA:
WEBTV.CAMERA.IT

Intervengono:

Lorenzo Fontana
Presidente
della Camera dei deputati

Anna Ascani
Vicepresidente
della Camera dei deputati

Per accrediti: sg_ufficiostampa@camera.it

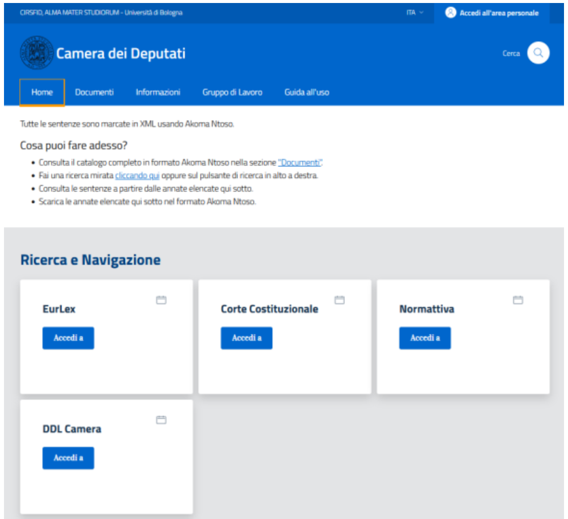
AI
della Manifestazione di interesse per la raccolta di proposte per l'utilizzo dell'intelligenza artificiale gene… per la Camera …putati
Camera dei deputati

Problem

1. **Assist** the legislative drafting offices and the members of parliament in retrieving relevant legislative sources related to the bill that they want to draft (heterogeneous corpora) – **legal ex-ante analysis**
2. Monitor **policies** over time – **ex-post analysis**
3. Detect inconsistencies in legislative bills or **unconstitutionalities** with the Constitutional Court
4. **Generate** new legislative definitions according to the existing corpora with **Agentic AI**

#1 Relevant documents with a lack of information in legislative heterogenous corpora

1. Complete drafting of normative references
2. Retrieve definitions
3. Extract relevant legal documents


CIRSFID, ALMA MATER STUDIORUM - Università di Bologna
ITA
Accedi all'area personale
Cerca
Home
Documenti
Informazioni
Gruppo di Lavoro
Guida all'uso

Camera dei Deputati

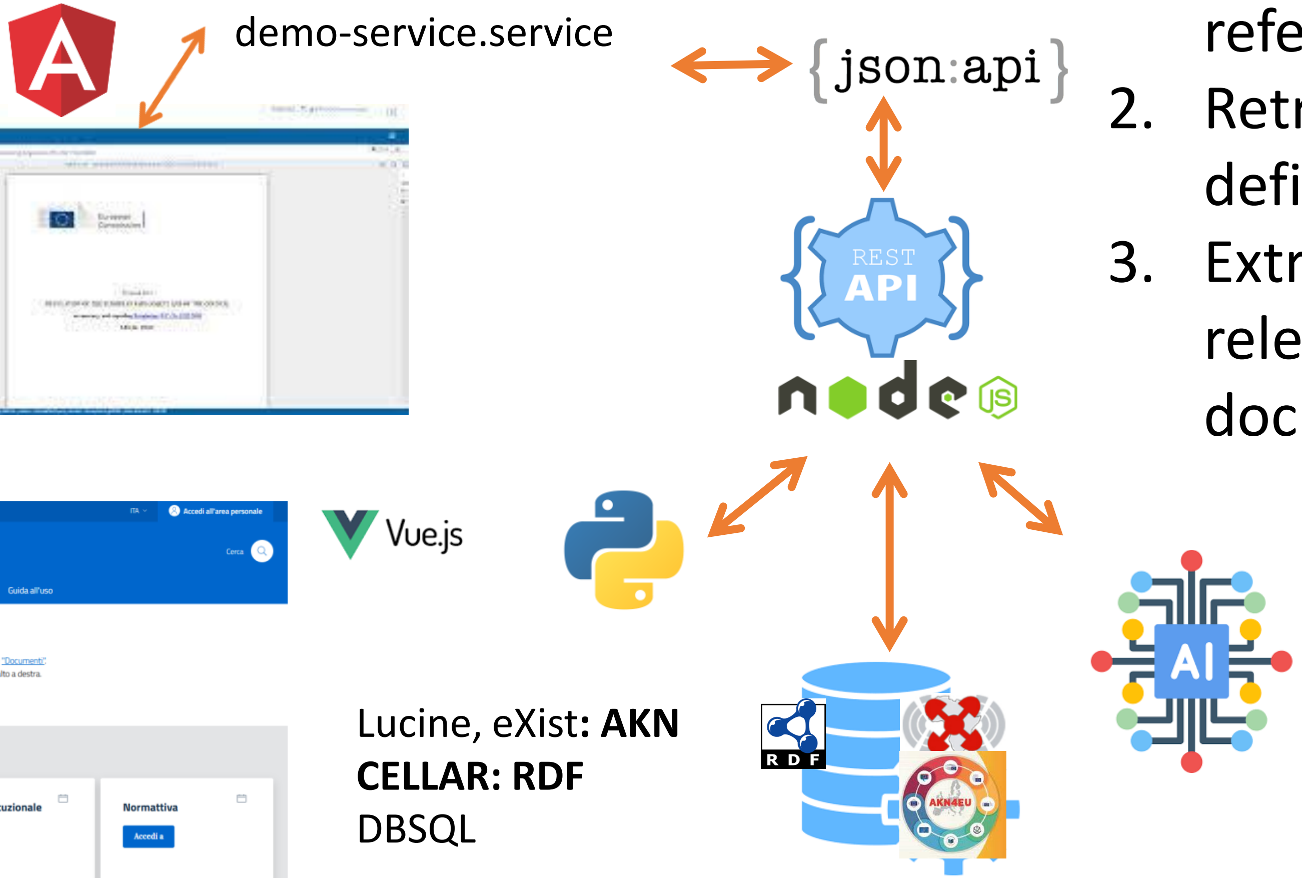
Tutte le sentenze sono marcate in XML usando Akoma Ntoso.

Cosa puoi fare adesso?

- Consulta il catalogo completo in formato Akoma Ntoso nella sezione "Documenti".
- Fai una ricerca mirata cliccando qui oppure sul pulsante di ricerca in alto a destra.
- Consulta le sentenze a partire dalle annate elencate qui sotto.
- Scarica le annate elencate qui sotto nel formato Akoma Ntoso.

Ricerca e Navigazione

EurLex Accedi a	Corte Costituzionale Accedi a	Normattiva Accedi a
DDL Camera Accedi a		

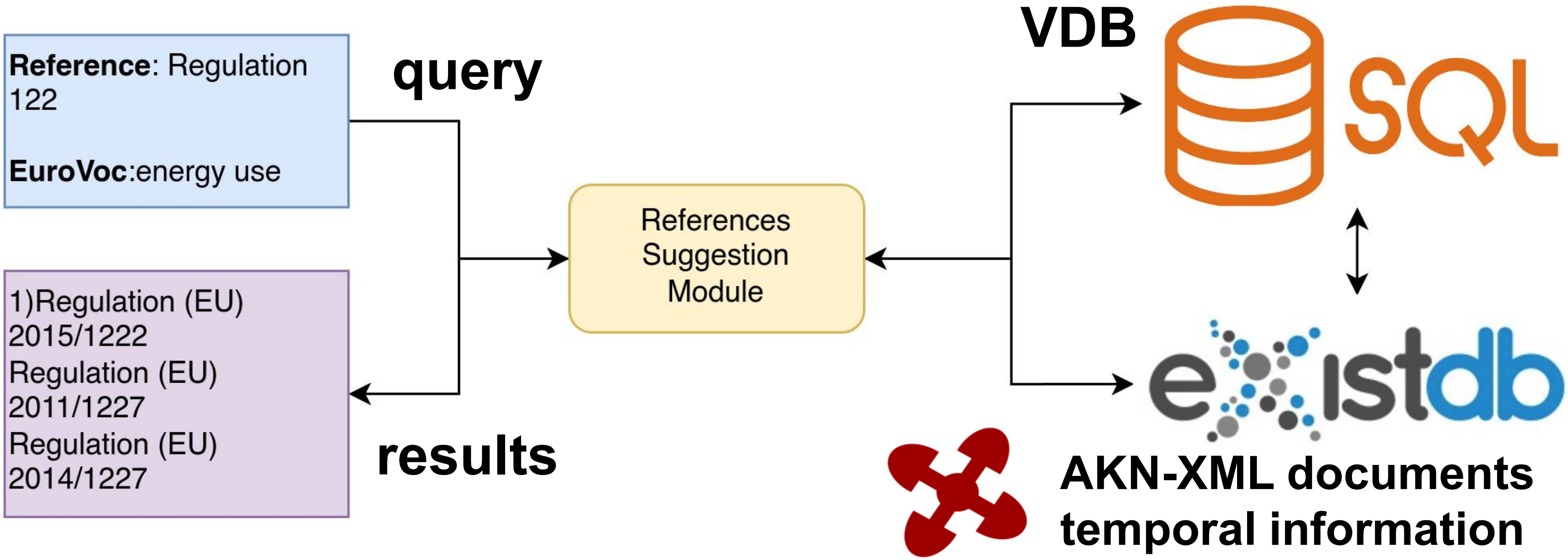


demo-service.service

Lucine, eXist: **AKN**
CELLAR: RDF
DBSQL

1.1 Normative references suggestion

Given an **incorrect** or **partial text**, to get the list of relevant references, syntactically correct, in force, refer to documents that have a similar topic.



Reference: Regulation 122

EuroVoc:energy use

query

References Suggestion Module

VDB

SQL

existdb

1)Regulation (EU) 2015/1222
Regulation (EU) 2011/1227
Regulation (EU) 2014/1227

results

AKN-XML documents
temporal information

1.1 Normative references suggestion

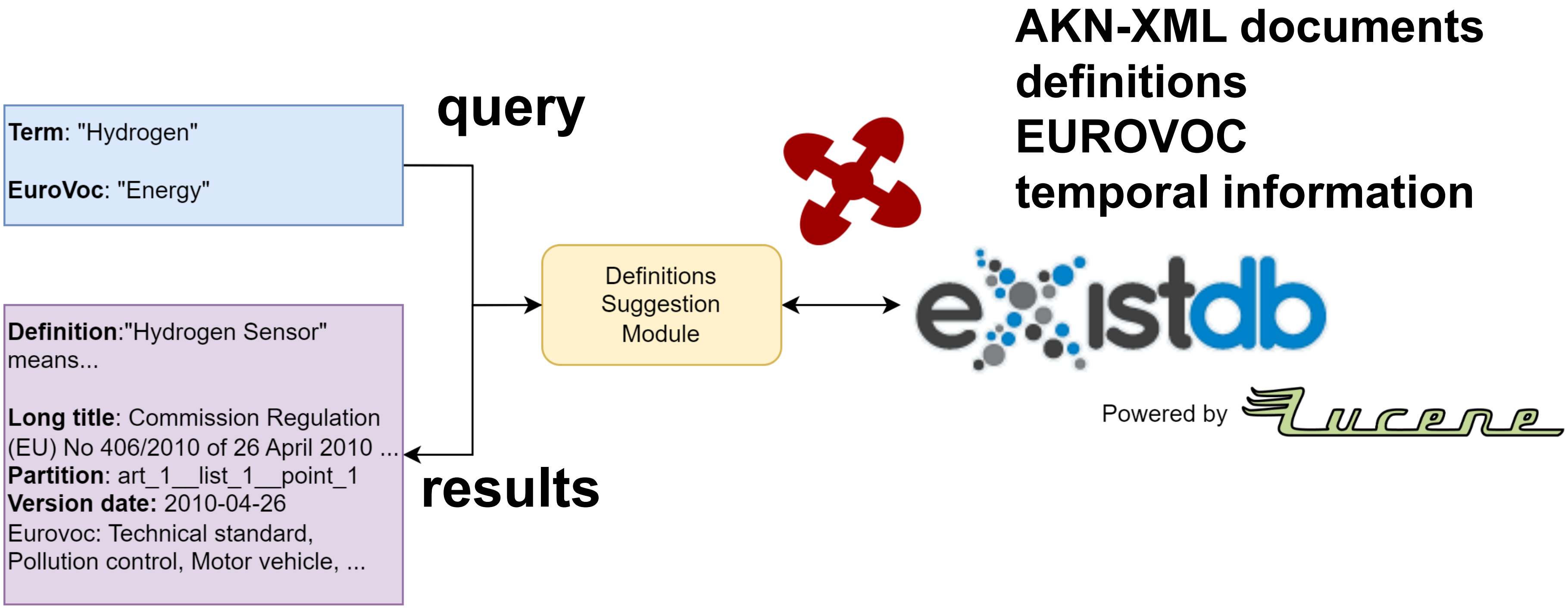
The module works as follows:

1. The year and number of the document are extracted from the query using heuristics
2. First, a query is used to find documents with the same number and/or year in the query (Directive 31 → Directive 2010/31/EU)
3. The model applies **Levensthein distance** to find documents that have a distance of at most 1 for both the year and document number
4. Averaged FastText embeddings are then used to encode the EuroVoc terms in the query and compare them with documents, obtaining a **ranking**

1.2 Legal definitions suggestion

Use prior existing definitions to suggest relevant ones, consolidated and updated, to the user using the topic of the bill (EuroVoc).

- Hydrogen definition in the Energy Bill updated today
- Hydrogen definition in the Food Bill in force in 2026



AKN-XML documents
definitions
EUROVOC
temporal information

Term: "Hydrogen"

EuroVoc: "Energy"

query

Definitions Suggestion Module

Definition:"Hydrogen Sensor" means...

Long title: Commission Regulation (EU) No 406/2010 of 26 April 2010 ...
Partition: art_1__list_1__point_1
Version date: 2010-04-26
Eurovoc: Technical standard, Pollution control, Motor vehicle, ...

results

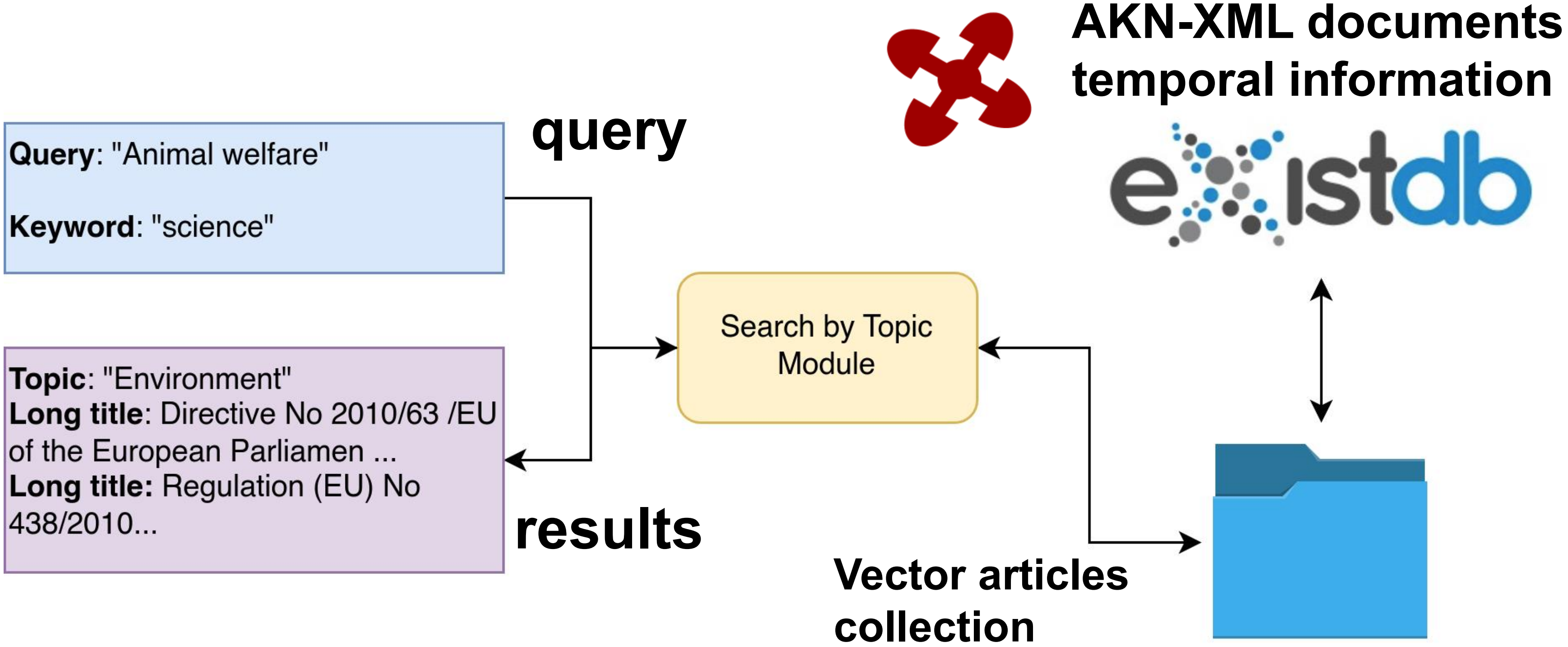
1.2 Legal definitions suggestion

The module uses the Similarity class from Lucene (a refined version of the Vector Space Model (VSM) based on TF-IDF weights) to compare a query and existing definitions:

1. The query EuroVoc and the EuroVoc from the document are compared
2. The query is compared with both the definendum and definiens of the definition (<<"definendum" means "definiens")
3. The results are filtered using user-specified manual constraints to obtain relevant results

1.3 Search by topic

- Given a keyword of EuroVoc and the topic of bill, it returns a list of the 10 relevant documents divided in two relevant topics top level EuroVoc terms.



AKN-XML documents
temporal information

query

Query: "Animal welfare"

Keyword: "science"

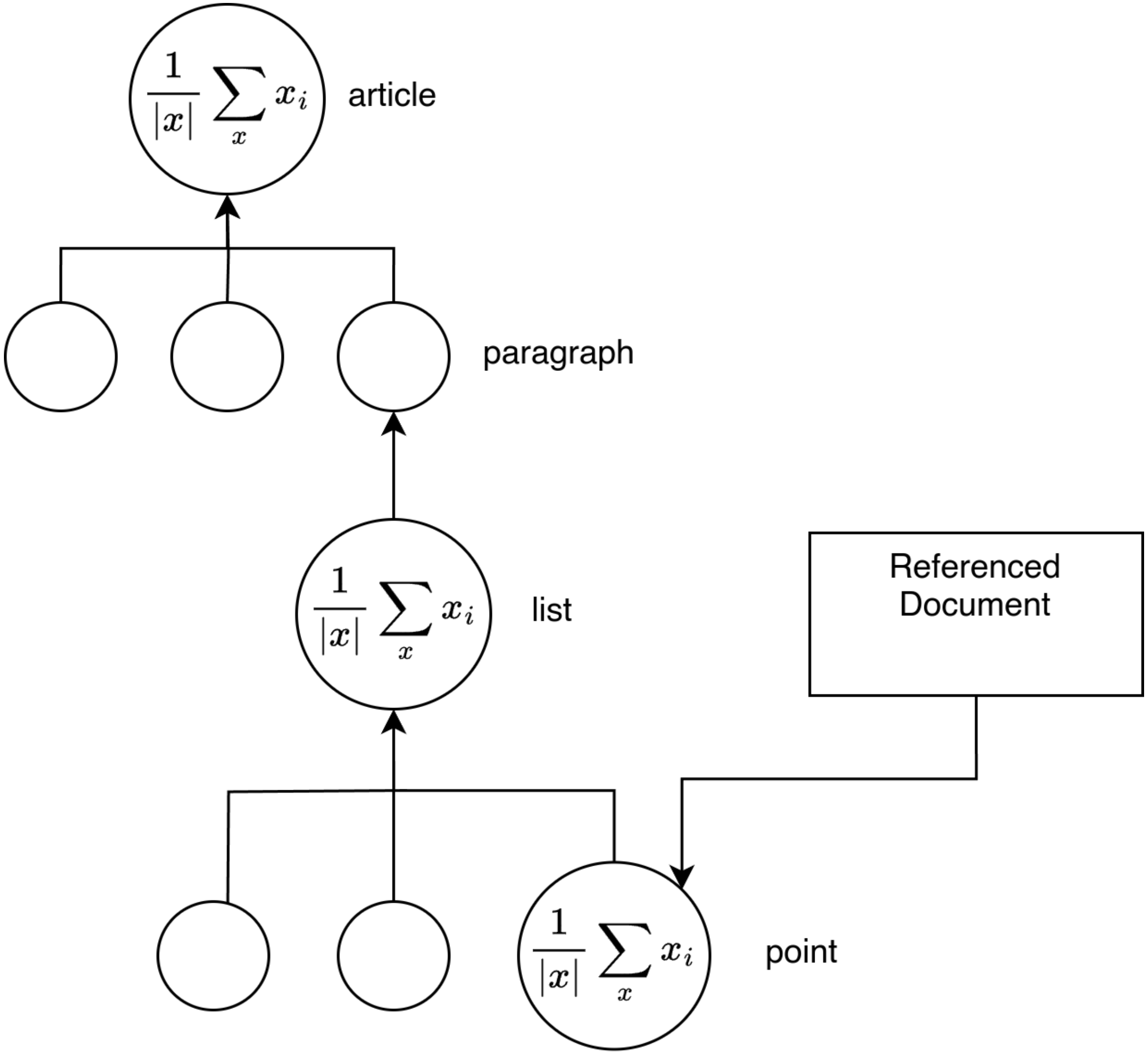
Search by Topic Module

Topic: "Environment"
Long title: Directive No 2010/63 /EU of the European Parliamen ...
Long title: Regulation (EU) No 438/2010...

results

Vector articles collection

1.4 Embedding method


$\frac{1}{|x|}\sum_x x_i$
article
paragraph
$\frac{1}{|x|}\sum_x x_i$
list
Referenced Document
$\frac{1}{|x|}\sum_x x_i$
point

#1 - Evaluation

Hybrid AI Enhancing European Drafting Legislation for a Better Regulation Michele Corazza, Generoso Longo, Monica Palmirani, Leonardo Zilli, Salvatore Sapienza and Emanuele Di Sante, ICEDEG, 18-20 June 2025, Bern

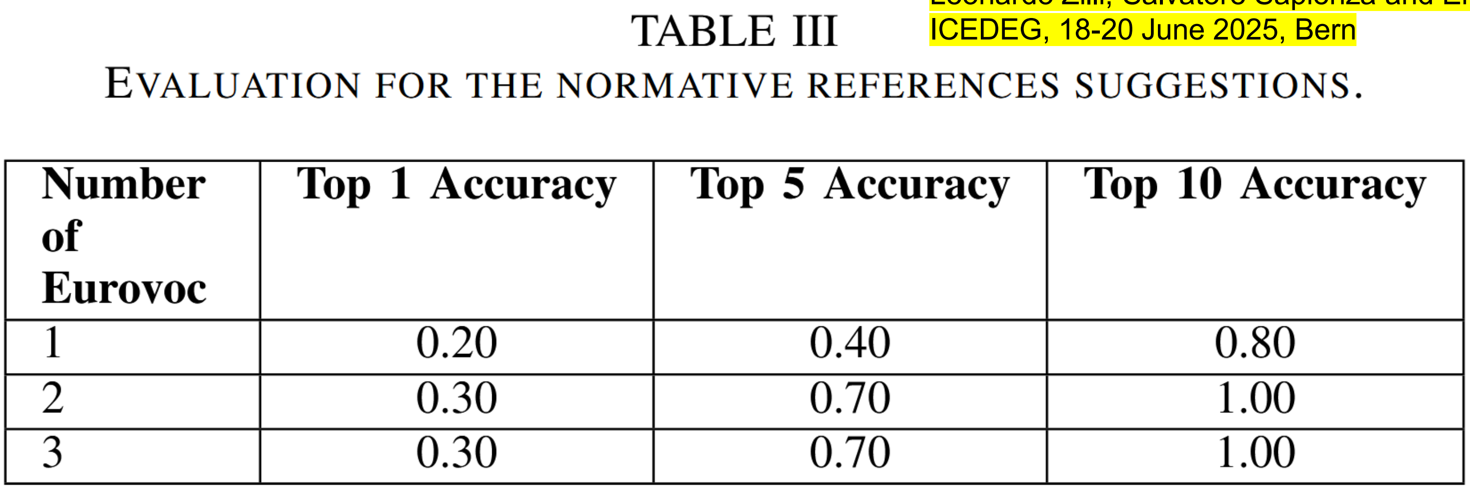


TABLE III

EVALUATION FOR THE NORMATIVE REFERENCES SUGGESTIONS.

Number of Eurovoc	Top 1 Accuracy	Top 5 Accuracy	Top 10 Accuracy
1	0.20	0.40	0.80
2	0.30	0.70	1.00
3	0.30	0.70	1.00

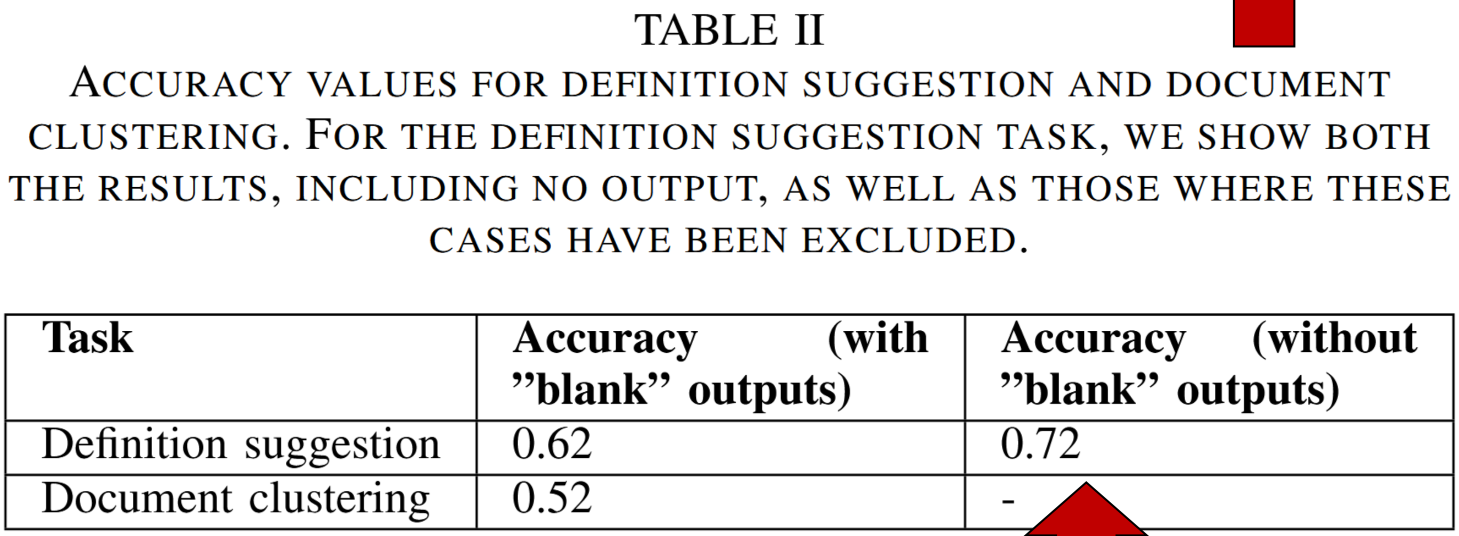


TABLE II

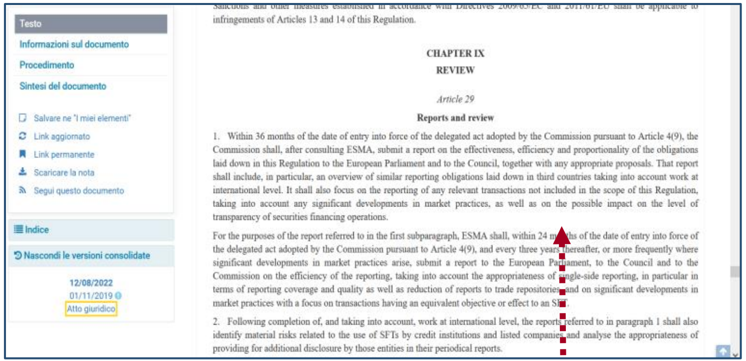
ACCURACY VALUES FOR DEFINITION SUGGESTION AND DOCUMENT CLUSTERING. FOR THE DEFINITION SUGGESTION TASK, WE SHOW BOTH THE RESULTS, INCLUDING NO OUTPUT, AS WELL AS THOSE WHERE THESE CASES HAVE BEEN EXCLUDED.

Task	Accuracy (with ”blank” outputs)	Accuracy (without ”blank” outputs)
Definition suggestion	0.62	0.72
Document clustering	0.52	-

#2 SORTIS: Monitoring and Measuring the Policy

Digital Ready Policy Making

AKOMA NTOSO
Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontolo



Sanctions and other measures established in accordance with Directives 2009/65/EC and 2011/61/EU shall be applicable to infringements of Articles 13 and 14 of this Regulation.

CHAPTER IX
REVIEW

Article 29

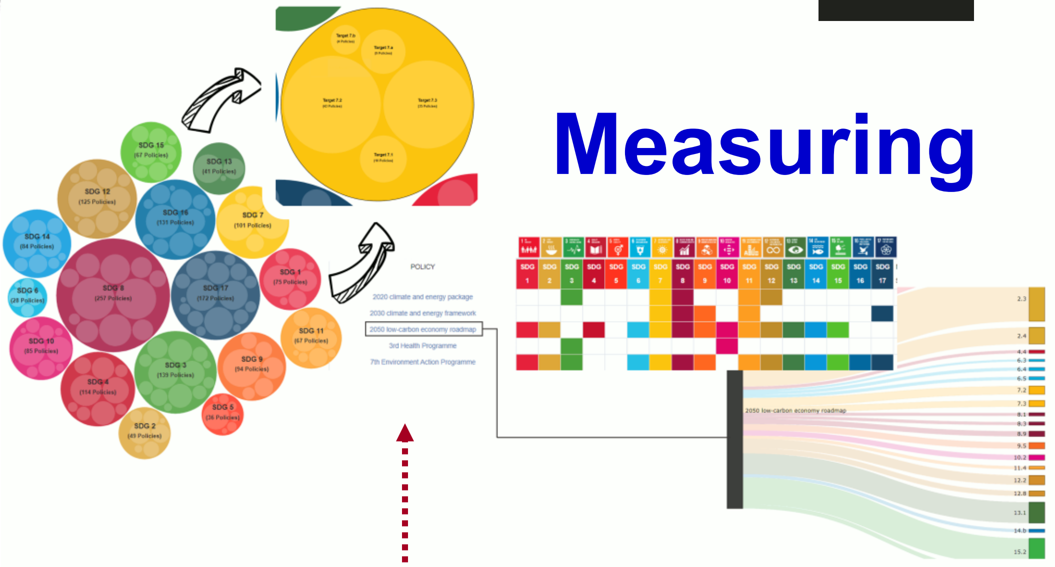
Reports and review

1. Within 36 months of the date of entry into force of the delegated act adopted by the Commission pursuant to Article 4(9), the Commission shall, after consulting ESMA, submit a report on the effectiveness, efficiency and proportionality of the obligations laid down in this Regulation to the European Parliament and to the Council, together with any appropriate proposals. That report shall include, in particular, an overview of similar reporting obligations laid down in third countries taking into account work at international level. It shall also focus on the reporting of any relevant transactions not included in the scope of this Regulation, taking into account any significant developments in market practices, as well as on the possible impact on the level of transparency of securities financing operations.

For the purposes of the report referred to in the first subparagraph, ESMA shall, within 24 months of the date of entry into force of the delegated act adopted by the Commission pursuant to Article 4(9), and every three years thereafter, or more frequently where significant developments in market practices arise, submit a report to the European Parliament, to the Council and to the Commission on the efficiency of the reporting, taking into account the appropriateness of single-side reporting, in particular in terms of reporting coverage and quality as well as reduction of reports to trade repositories and on significant developments in market practices with a focus on transactions having an equivalent objective or effect to an SFT.

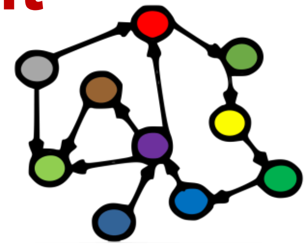
2. Following completion of, and taking into account, work at international level, the reports referred to in paragraph 1 shall also identify material risks related to the use of SFTs by credit institutions and listed companies and analyse the appropriateness of providing for additional disclosure by those entities in their periodical reports.

Measuring



POLICY	SDG 1	SDG 2	SDG 3	SDG 4	SDG 5	SDG 6	SDG 7	SDG 8	SDG 9	SDG 10	SDG 11	SDG 12	SDG 13	SDG 14	SDG 15	SDG 16	SDG 17
2020 climate and energy package			■				■	■			■	■					
2030 climate and energy framework							■	■	■		■						■
2050 low-carbon economy roadmap	■	■		■		■	■	■	■	■	■	■	■	■	■		
3rd Health Programme			■							■							
7th Environment Action Programme	■	■	■			■	■	■	■		■	■	■	■	■	■	■

Reporting Requirement

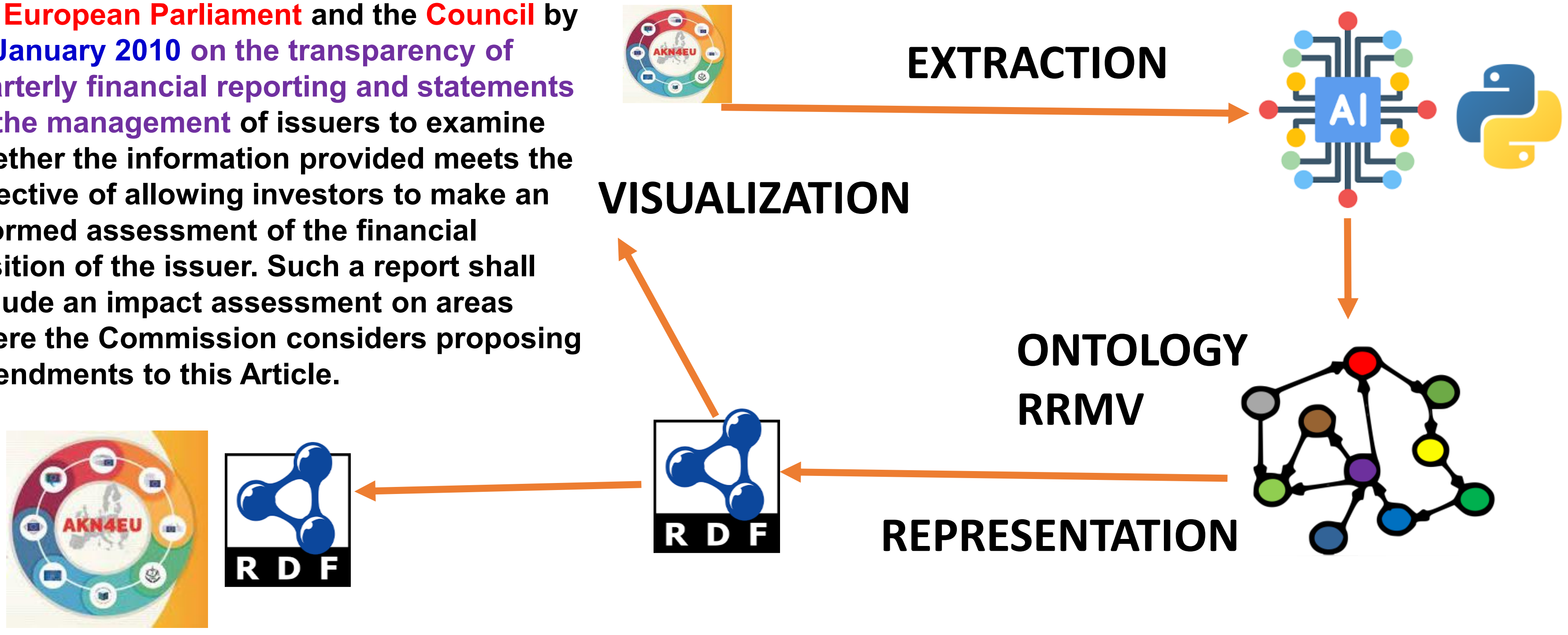


RRMV-RDF

Monitoring

#2 - Extraction and Representation of «Reporting Requirements»

3. The Commission shall provide a report to the European Parliament and the Council by 20 January 2010 on the transparency of quarterly financial reporting and statements by the management of issuers to examine whether the information provided meets the objective of allowing investors to make an informed assessment of the financial position of the issuer. Such a report shall include an impact assessment on areas where the Commission considers proposing amendments to this Article.



EXTRACTION

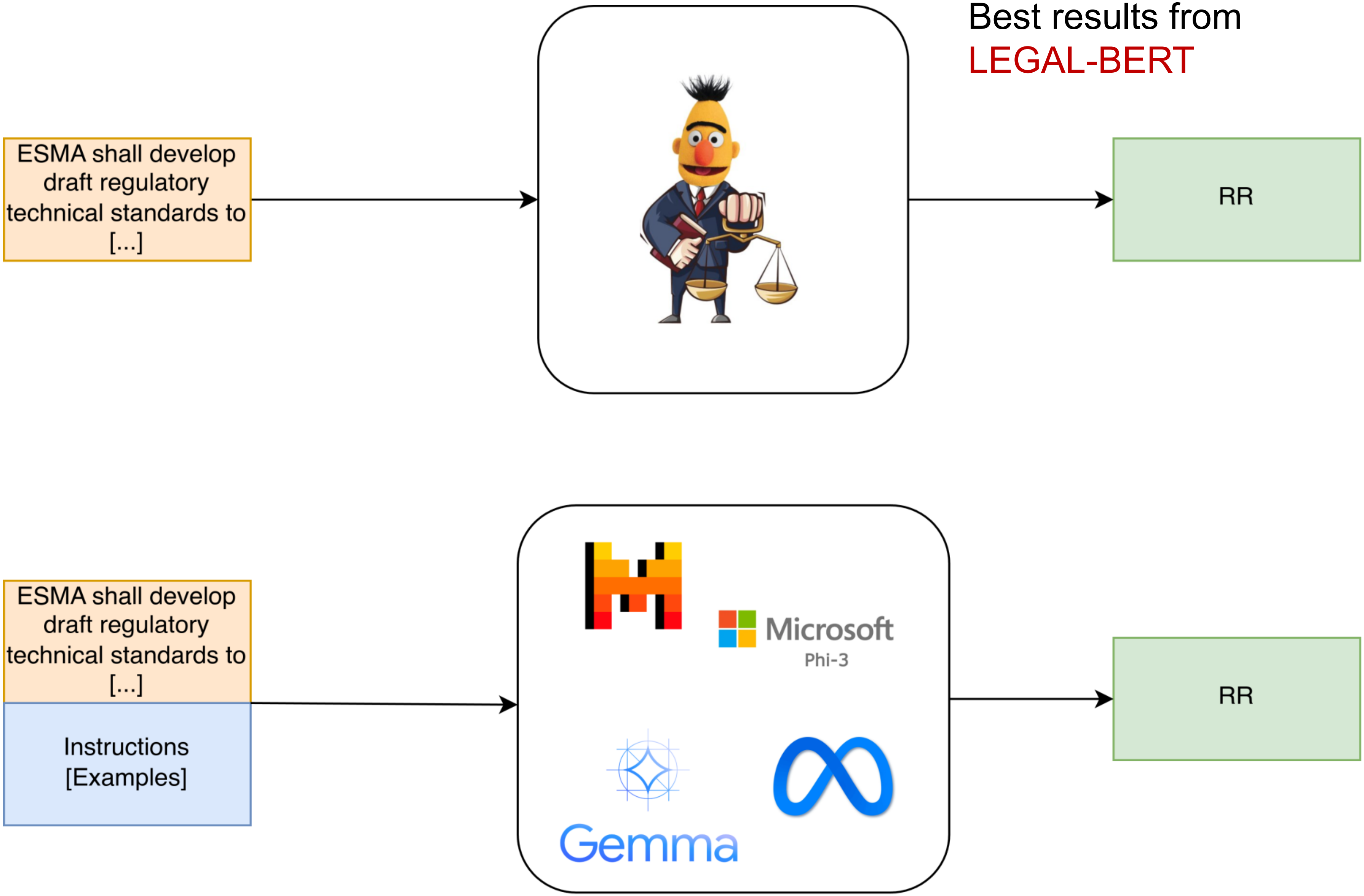
VISUALIZATION

ONTOLOGY RRMV

REPRESENTATION

INTEGRATION IN AKN and in KG

#2 - RR binary classification



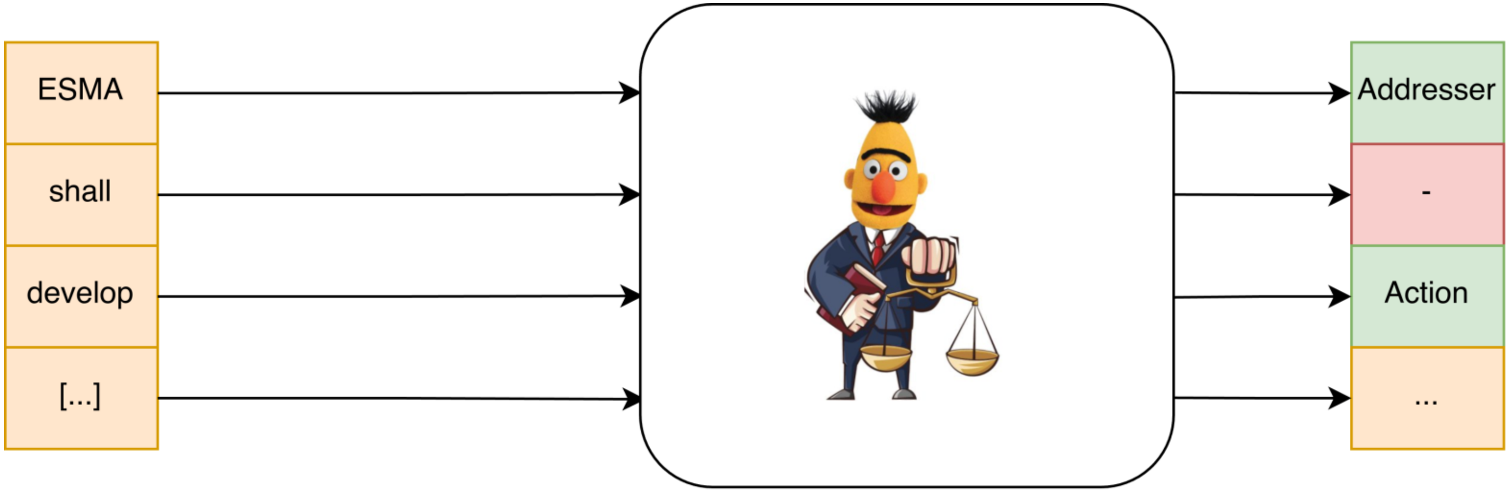
ESMA shall develop draft regulatory technical standards to [...]

Best results from LEGAL-BERT

ESMA shall develop draft regulatory technical standards to [...]

#2 - Entity Extraction

Best result from LEGAL-BERT, followed closely by GEMMA 10-shot and Llama3.3 10-shot RDF



ESMA

shall

develop

[...]

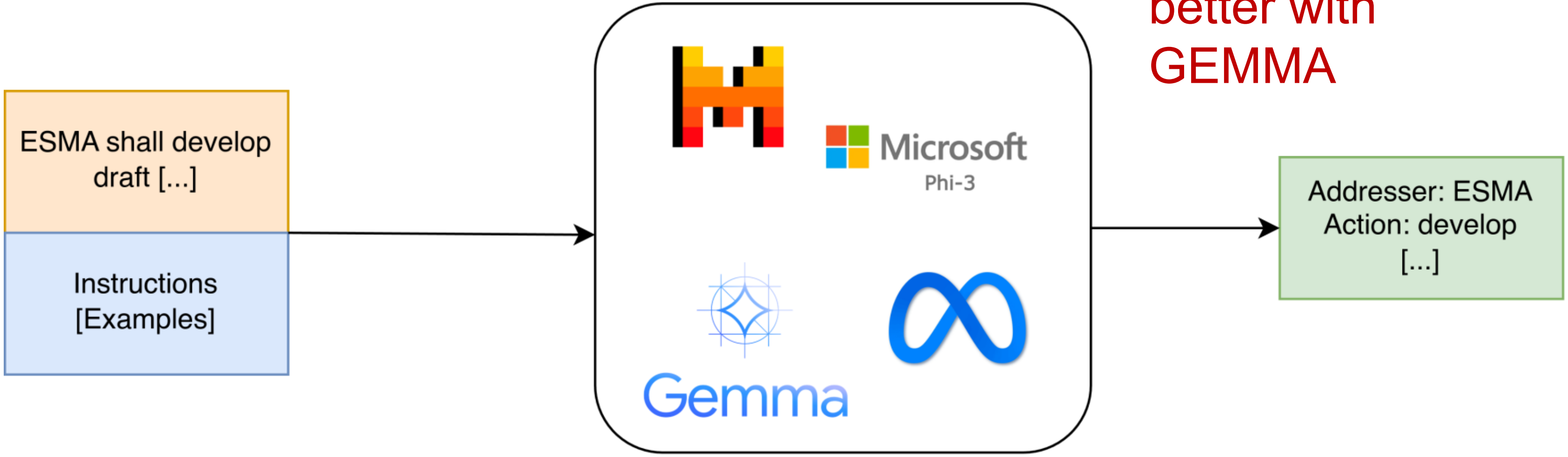
Addresser

-

Action

...

PeriodOfTime better with GEMMA

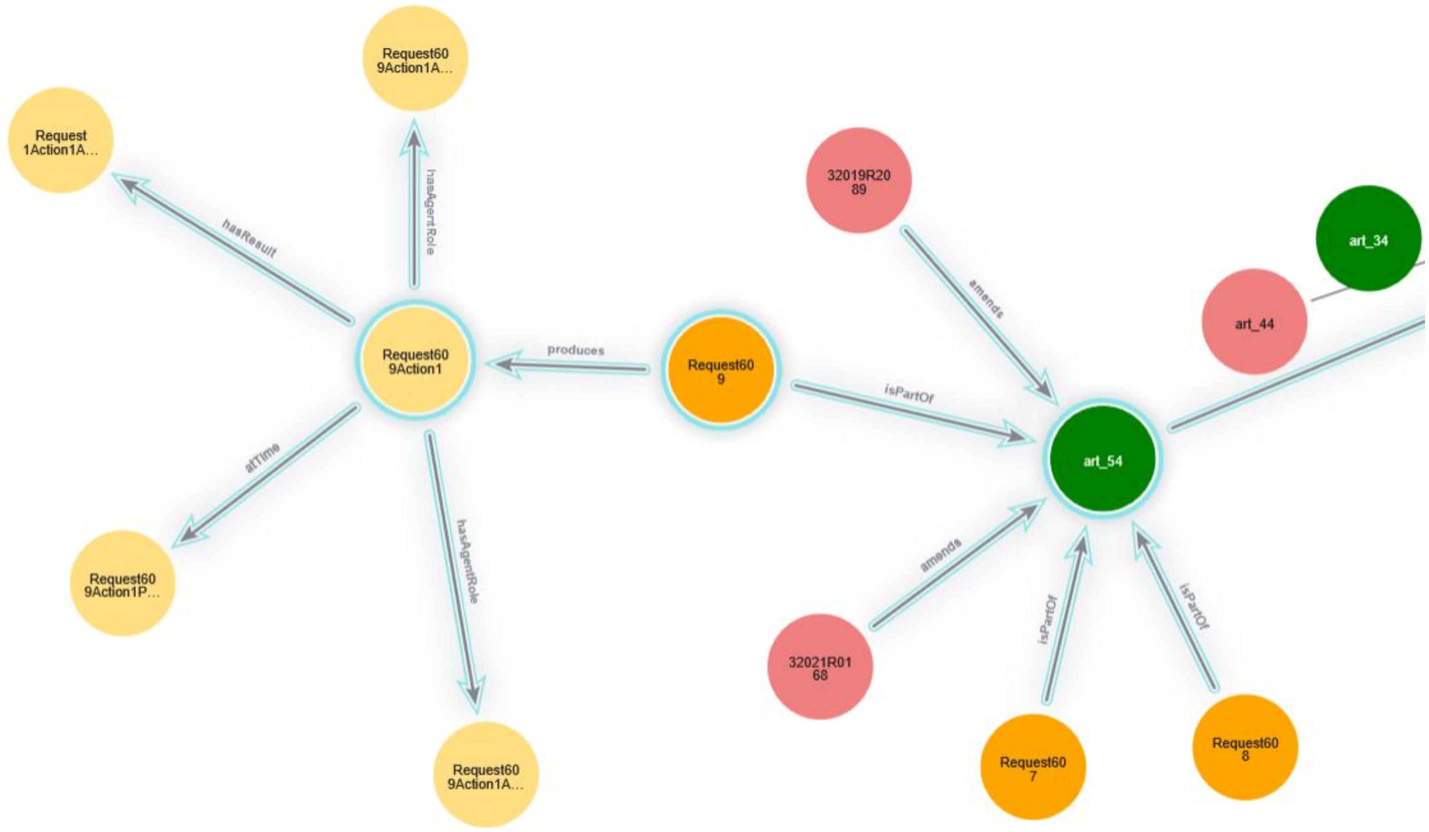


ESMA shall develop draft [...]

Instructions [Examples]

Addresser: ESMA
Action: develop
[...]

#2 - Art. 54 is modified, it contains RR – Neo4J



Conclusions

- In order to mitigate the risks of the application of LLMs to the legal domain, we can apply a hybrid approach
- The Akoma Ntoso standard allows us to leverage its rich information and to use them with sophisticated NLP/AI models (structure of the documents, definitions, references, jurisdiction, etc)
- By enriching statistical models with symbolic information, we reduce the risk of hallucinations and ensure that the models consider the relevant (i.e. in force) documents while also considering the temporal aspects of the law.