Η κατάσταση των ανοιχτών δεδομένων στον τομέα της Δικαιοσύνης στην Ελλάδα: Τα έργα Νομοθεσία και Χωρονομοθεσία

Μανόλης Κουμπαράκης
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Σε συνεργασία με τους Ηλία Χαλκίδη, Ιωσήφ Αγγελίδη, Χρήστο Παπαλουκά, Μπάμπη Νικολάου, Παναγιώτη Σούρσο (Νομοθεσία), Δημήτρη Βαρσάμη, Χρήστο Παπαδόπουλο και Γιάννη Παναγάκη (Χωρονομοθεσία)

ΤΝ στη Δικαιοσύνη
10 Ιουνίου 2021

Motivation

Current state of affairs (Documents)

LEGISLATIVE DRAFTING

DISCUSSION / VOTING

PUBLISHING

Ministries
(legislative committees, social actors consultations, public consultation)

draft law

Hellenic Parliament
(Scientific Service of the House, parliamentary committees, reports, amendments, discussion, voting)

law

National Printing Office
(printing, publishing in the Government Gazette)

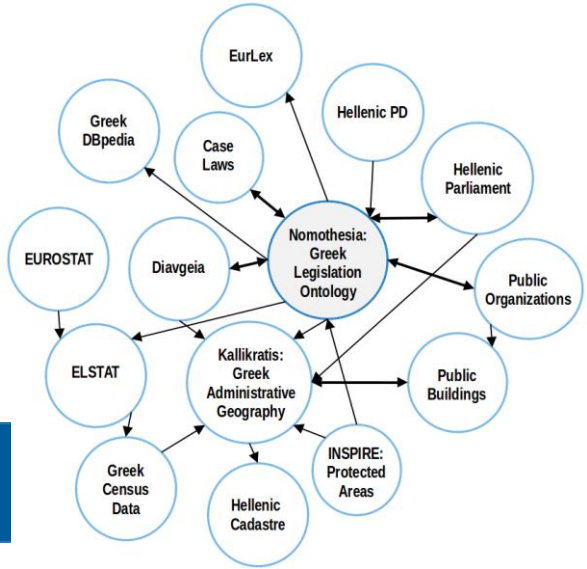
New state of affairs (using Artificial Intelligence)

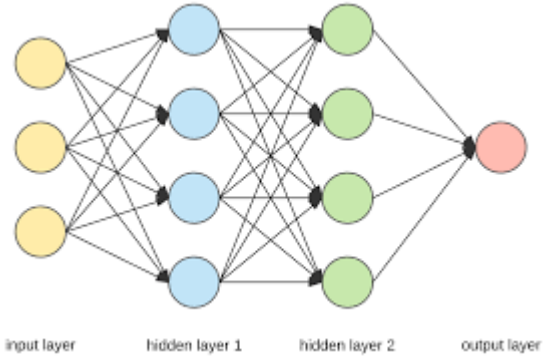
*We envision a **new state of affairs** in which:*

- ***legislation is published** in a **machine-readable format on the Web** as **open linked data** so that it can be used in **software**, and so that it **can be also combined with other open data** to **increase its value** for interested people.*

- ***professionals** (e.g., lawyers, judges) have the same **advanced search capabilities** but also **advanced AI tools** appropriate for their job.*

- ***ordinary citizens** have **advanced search capabilities** at their fingertips on the content of legislation.*





Nomothesia

Types of Greek Legislation

- In this work we are considering the encoding of ***five primary sources (types)*** of Greek legislation:

Constitution
Σύνταγμα

Presidential Decrees
Προεδρικά Διατάγματα

Laws
Νόμοι

Acts of Ministerial Cabinet
Πράξεις Υπουργικού Συμβουλίου

Ministerial Decisions
Υπουργικές Αποφάσεις

- We also consider ***two secondary sources*** of Greek legislation:

Legislative Acts
Πράξεις Νομοθετικού Περιεχομένου

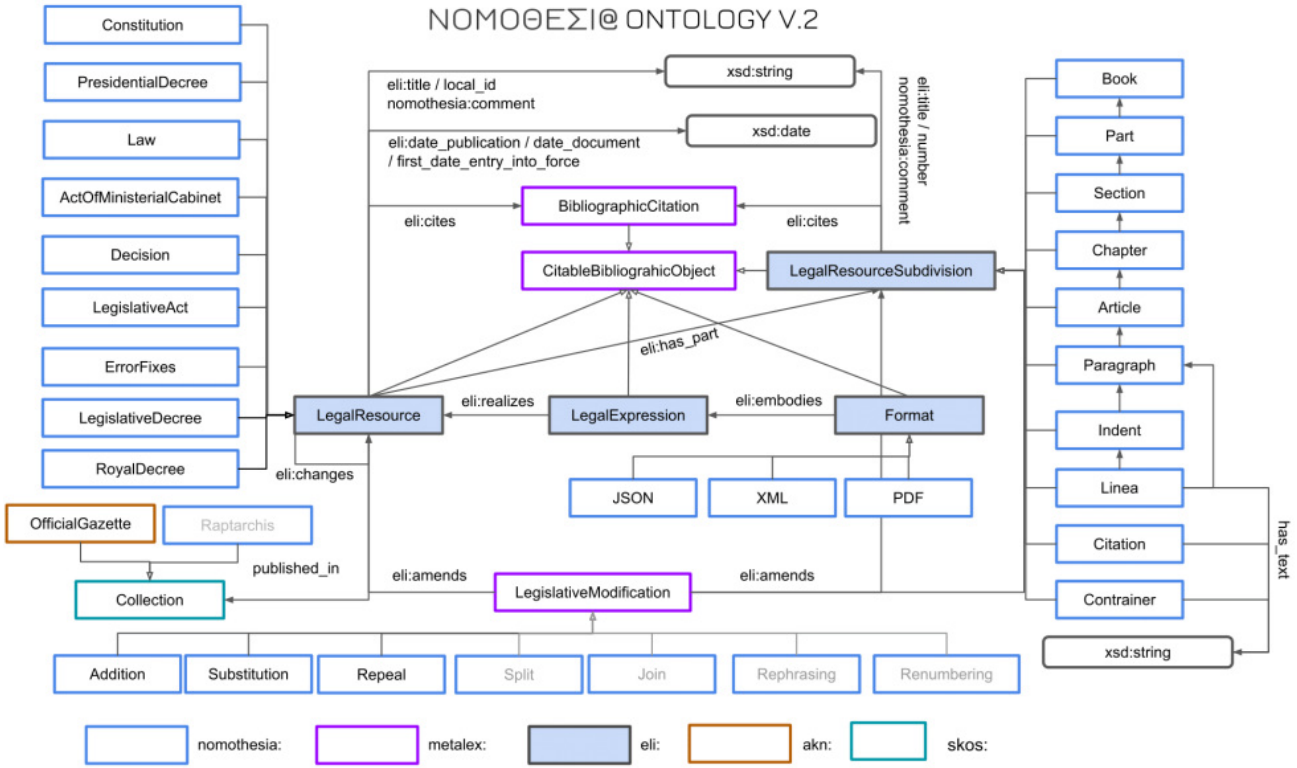
Regulatory Provisions
Κανονιστικές Διατάξεις

- All these legislation types are materialized in ***legal documents*** that adhere to a typical structure organized in a tree hierarchy around the concept of ***fragments (ενότητες)***.

Legislative Modifications

- We have studied three main types of ***legislative modifications***:
    - ***Insertion***
    - ***Substitution***
    - ***Repeal***
- These three kinds of modifications (amendments) produce new ***versions*** of the original, as enacted, legal document.

The Nomothesia Ontology

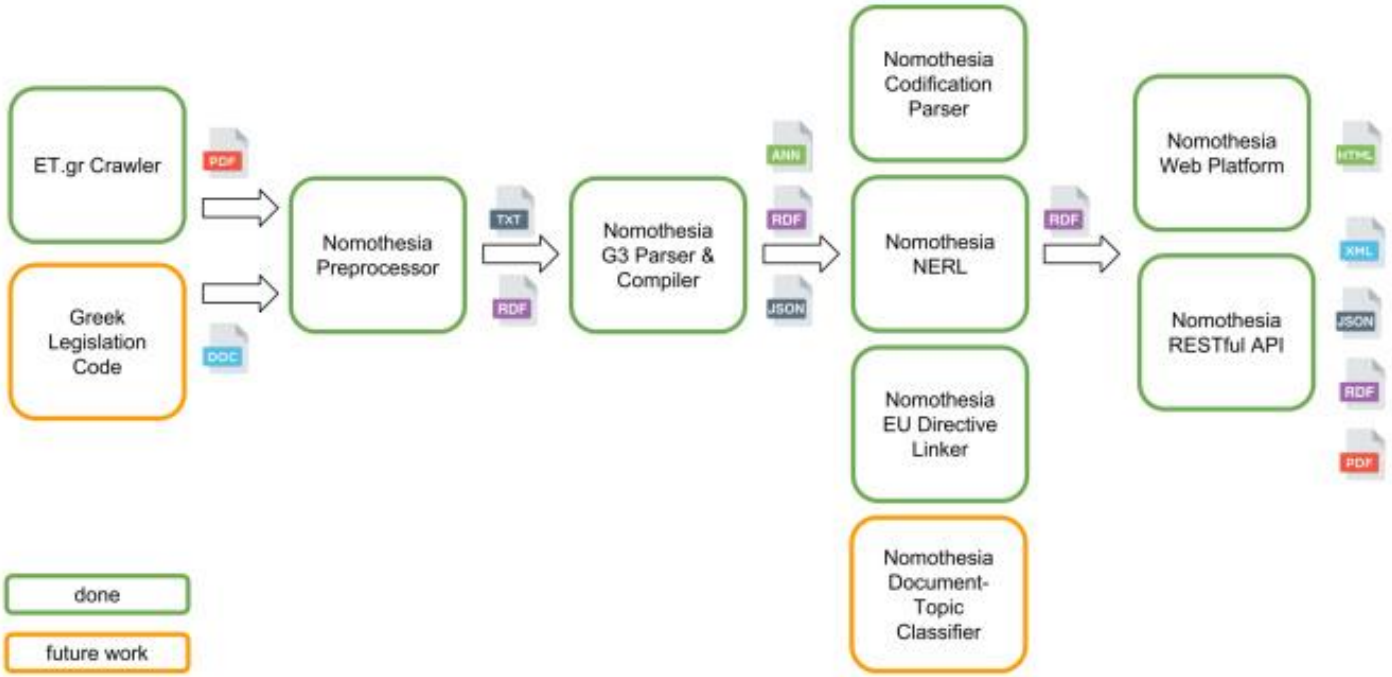


Persistent URIs in Nomothesia (as in ELI)

`http://legislation.di.uoa.gr/eli/{type}/{year}/{id}`

- For example, the URI for Presidential Decree 54/2011 is http://legislation.di.uoa.gr/eli/pd/2011/54
- **Version extensions**:
  - current **/**
  - original (as enacted) **`/enacted`**
  - chronological **`/{YYYY-MM-DD}`** (ISO 8601) (e.g., http://legislation.di.uoa.gr/eli/pd/2011/54/2015-10-21)
- **Manifestation extensions:**
  - HTML **/**
  - XML **`/data.xml`** (e.g., http://legislation.di.uoa.gr/eli/pd/2011/54/data.xml)
  - JSON **`/data.json`** (e.g., http://legislation.di.uoa.gr/eli/pd/2011/54/data.json)
- **Fragment extensions:**
  - **`.../article/{article_number}/paragraph/{para_number}`**
  - e.g., http://legislation.di.uoa.gr/eli/pd/2011/54/article/1/paragraph/2

The Nomothesia Pipeline



Interesting AI Contributions

- We use a **bi-LSTM-CRF network** to extract entities of the following types from FEKs: **persons, geopolitical entities, organizations, geographical landmarks, legislation references and public document references.**
- **Interlinking** with other open data such as Greek Politicians in DBpedia, Greek Administrative Geography and European Legislation.

Querying the Greek legal knowledge graph

- Queries ***examining the metadata*** of a legislative document (e.g., “Give me the titles of all PDs in the last year.”).
- Queries ***examining the structure*** of a legislative document (e.g., “Give me the titles and first two articles of all PDs in the last year.”).
- Queries ***examining the changes*** of a legislative document (e.g., “Give me all the PDs that have modified PD 2011/54.”)
- Queries ***computing various statistics*** e.g., on changes of laws (e.g., “Give me the 5 most frequently modified legal documents in 2008-2013.”).
- Queries ***linking legislation with other open data*** (e.g., “Give me laws that deal with reforestation of areas burned by fires during the summer fires of 2007.”).
- ***Investigative queries*** e.g., by a journalist.

Some statistics

- We provide all legal documents of issues A and D of the FEK for the years 1990-2019, the penal and civil code of Greece and all European directives and treaties extracted from EUR-Lex in Greek.
- 12.000 legal documents.
- 195.000 references to entities.
- 5 million RDF triples in the Greek legal knowledge graph.

Demonstration

http://legislation.di.uoa.gr

Evaluation and Feedback

- Nomothesia is a **research prototype** and it is still not well-known to the public.

  - In April of 2016, Nomothesia was presented and received an award in the **1st IT4GOV Contest**, which was organized by the Greek Ministry of Administrative Reform & Electronic Governance. Among the contest jury members were 3 ministers and people from the industry, who judged Nomothesia positively.

- Following this contest, we have been in contact with officials from the **Hellenic Parliament,** the **Ministry of Administrative Reform & Electronic Governance** and the **Ministry of Justice** which have expressed their interest in Nomothesia.

- We presented Nomothesia to the ELI group in Brussels in October 2017.

- We presented Nomothesia to the Athens Bar Association in 2018.

Choronomothesia

Introduction

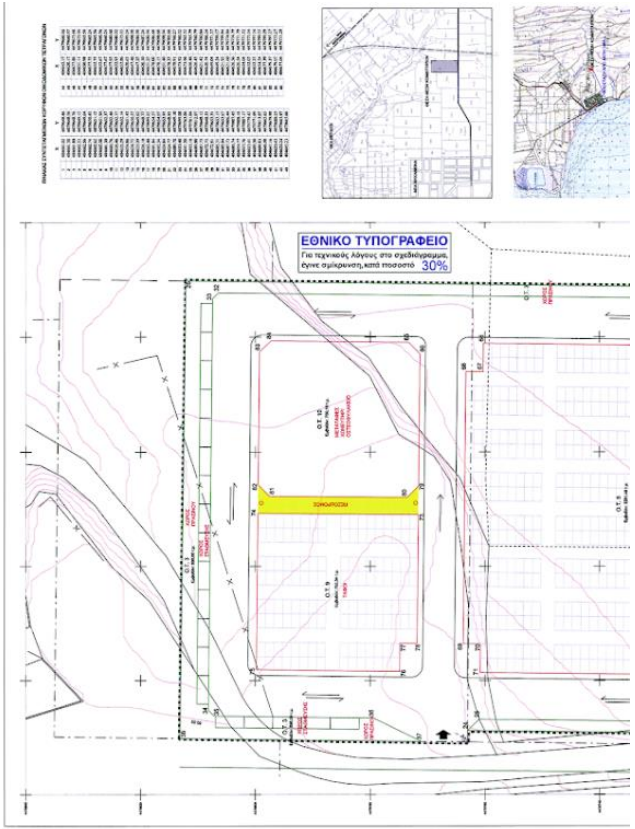
- Choronomothesia is a project funded by the programme **Ερευνώ – Δημιουργώ – Καινοτομώ**.
- It is a joint project with geoinformation company Geoapikonisis (https://www.geoapikonisis.gr/).
- The goal of Choronomothesia is to extract geospatial information present in the Government Gazette (issues D and AAP).

- The information extraction pipeline provides a ready-to-use PostGIS database where the information extracted from the legal documents is stored; this information can be used for further application development.

- One of our developments in this project has been **image processing techniques for extracting geospatial information from images found in PDFs.**

Choronomothesia

- We use **convolutional neural networks** trained for the object detection task of locating tables containing spatial coordinates within the PDF files and fine-tuned using our own dataset.
- In the example shown, we are interested in **detecting the sideways upper-left table and extract its coordinates.**
- Our model achieves a classification error of 0.0694 and a bounding box error of 0.2653 on unseen data (the available annotated dataset is relatively small, ~300 images).



Choronomothesia

- After extracting the tables, we use OCR to locate and extract the coordinates.
- We first extract the cells using simple OpenCV methods.
- We then use an OCR engine (tesseract) to get the coordinate number within the extracted cell of the table.

X	Y
390613	4503765
390630	4503701
390637	4503669
390640	4503624
390645	4503567
390646	4503484
390648	4503456
390646	4503408
390649	4503361
390651	4503330
390640	4503303
390610	4503244
390619	4503226
390630	4503182
390653	4503159
390694	4503131
390775	4503080
390857	4503032
390934	4502979
391015	4502925
391079	4502886
391108	4502863
391134	4502829
391165	4502778
391214	4502699
391239	4502678
391299	4502647
391306	4502643
391341	4502637
391526	4502641
391666	4502697
391749	4502734
391839	4502734
391926	4502672
391941	4502651
391943	4502649
391988	4502593
391992	4502572
391994	4502569
392004	4502503
392003	4502485
392006	4502452
392005	4502344

X	Y
392008	4502296
392076	4502227
392078	4502227
392316	4502253
392382	4502243
392429	4502240
392479	4502240
392530	4502236
392618	4502202
392663	4502173
392729	4502117
392776	4502060
392806	4502025
392845	4501998
392865	4501985
392925	4501950
392969	4501923
393024	4501891
393188	4501813
393309	4501750
393313	4501748
393399	4501722
393506	4501717
393547	4501710
393670	4501651
393782	4501610
393957	4501527
394027	4501488
394040	4501469
394056	4501453
394074	4501421
394102	4501382
394127	4501344
394148	4501311
394173	4501288
394207	4501263
394273	4501221
394344	4501195
394389	4501182
394452	4501162
394508	4501143
394581	4501116
394636	4501076

Other Related Work

- **Legal text classification** for the “Permanent Greek Legislation Code - **Raptarchis**” dataset.
- This dataset contains Greek legislation since the creation of the Greek state in 1834 until 2015.
- It includes laws, decrees, regulations and decisions with their respective amendments such as replacements, modifications and deletions, while its only source of information is the Official Government Gazette.
- Raptarchis is publicly offered through the portal e-Themis (http://e-themis.gov.gr).
- We implemented and evaluated **neural network techniques and large language models (M-BERT, Greek BERT) for the problem of classification.**

Other Related Work (cont’d)

- **Legal language models**: developed the **Greek Legal BERT** which is built using the resources of Nomothesia, Raptarchis and Greek part of EurLex.

Conclusions and Future Work

Conclusions and Future Work

- **AI has a lot more to give** to the development of intelligent systems for legislation.
- **Recommendation to the Greek Government:** AI technologies should be used in all stages of producing legislation.
- We can imagine the **European legal knowledge graph** interconnecting all legislations of European countries.

- An interesting open problem for us is answering **natural language questions** over legal knowledge graphs such as Nomothesia.

- Another interesting problem is developing Nomothesia **chatbots (legal digital assistants).**

Thank you!
Questions ?

http://www.di.uoa.gr/~koubarak
http://ai.di.uoa.gr
koubarak@di.uoa.gr
@mkoubarakis