

Magna Carta

Mining

[GitHub Repository](#)

PREAMBLE

Every person that is located in the European Union's territory has and is owed human rights, which they can assert against the State, companies, the police, and your fellow Hertie lecturers or class mates. However, people are unaware of what their rights are, how rights evolve, and what rights even mean. The HUDOC European Union Court of Human Rights database provides access to all the relevant European Union human rights case law. However, this database fails at making this law accessible to the people it serves.

Magna Carta Mining fills the gap left by HUDOC and makes the legal information accessible to the European Union's population by providing the relevant information in a manner that is readily understandable (data visualisation), providing the user with the the ability to identify trends in human rights issues, and with the ability to download documents en masse. The data visualisations pertain to rights based issues that are generalised or specific to the user.

Specific information can be selected by the user who will be provided with a multitude of infographics which they can manipulate. Examples of human rights information that the user can identify include, but is not limited to:

- the change in human rights issues over time;
- the number of cases decided by European Member States in any given year; or
- the most topical issue for any or all European Member States.

1 § Modules

1. Web scraping

Initially our program will access the comprehensive database of HUDOC's European Union Court of Human Rights legal documents (<https://hudoc.echr.coe.int>). Specifically, we built a web scraper to download all substantive judgements from this database. Additionally, we will extract all metadata associated with the legal documents to enable further exploration and analysis.

2. Data cleaning & wrangling

In order to be able to work with the scraped data we will construct a database, comprising all relevant information regarding all judgements. During the data cleaning & wrangling stage we will use the spaCy library for NLP-preprocessing to conveniently organise information for further analysis.

3. Data analysis

In the analysis stage of the project we will do basic to advanced NLP analysis of the documents to provide the underlying information of the data visualisation. This analysis is what transforms the HUDOC database to a database where meaningful insight can be gained, i.e. a transformation of an inaccessible database to an accessible database.

4. Information Visualisation

The user will be presented with interactive infographics, the output of the previous modules, that they can manipulate. Here, we will leverage the plotly library to produce the plots and the dash library to produce a navigable dashboard for the user.

2 § Benefits to Community

As the most authoritative and free database on Human Rights judgements delivered in the European Union, the HUDOC database is largely inaccessible because, amongst other things, it:

- does not provide the ability to mass download judgments;
- does not provide any form of textual analysis of a single document or multiple documents; or
- does not provide any trend analysis of legal issues.

Magna Carta Mining will increase the accessibility of this complex and incomprehensible database to individuals that do not have a legal background. In short, Magna Carta Mining will deliver human rights to the world.

3 § Deliverables

1. First Draft: November 8, 2021

- a. Presentation of clean data
 - i. Completed web-scraping / data collection
 - ii. Completed data cleaning
- b. Initial data wrangling
- c. Preliminary findings from NLP text-mining
- d. Preliminary visualizations (predefined / user input)
- e. Documentation for completed modules

2. Final Draft: November 29, 2021

- a. Fully developed modules with functioning code
- b. Junit testing docs
- c. Complete documentation

3. Presentation: November 29, 2021

- a. Slides
- b. Video

4 § Related Work

There has been related academic work focused on [leveraging machine learning to predict the European Union Court of Human Rights' judgments](#) and [Argument Mining](#). Magna Carta Mining will add to these by making research and analysis of the European Union Court of Human Rights' judgements more accessible, bridging gaps between research, human rights law, and the general public.

Signatories

Julian Kath
Kabir Sandrolini

Katalin Bayer
Viragj Akuthota