



Master Thesis in Data Science

21 September 2021

Law and Order: a Search Engine for Italian Law

Prof. Mauro Conti

Dr. Luca Pajola

Piero Romare
mat. 1205804

Academic Year 2020/2021





Our Goal

Make available to Italian law firms a tool in which to retrieve jurisprudence case.

Legal Data

Working prototype of a search engine for a web application.

**Natural Language
Processing**

Lawyers can do fast and cheap research about cases.

Search Engine

Web Application



Our Goal



Legal Data



Pre-processing



Search Engine



User Interface

Legal Data

**Natural Language
Processing**

Search Engine

Web Application



Our Goal

One | LEGALE

 Law Insider

Legal Data

 Lexica

ROSS 

Natural Language
Processing

DeJure

 casetext

Search Engine

Web Application



Our Goal

Civil Law

written laws bind the decisions of judicial bodies

Legal Data

Judgements & Ordinances

**Natural Language
Processing**

Constitutional Court

Search Engine

Source of Law

Web Application



Our Goal

Legal Data

**Natural Language
Processing**

Search Engine

Web Application

Civil Law

Judgements & Ordinances

is the final decision of the judge that puts
an end to a dispute

is a decision taken by the judge
characterized by flexibility

Constitutional Court

Source of Law



Our Goal

Civil Law

Legal Data

Judgements & Ordinances

**Natural Language
Processing**

Constitutional Court
guarantor of the Constitution

Search Engine

Source of Law

Web Application



Our Goal

Civil Law

Legal Data

Judgements & Ordinances

**Natural Language
Processing**

Constitutional Court

Search Engine

Source of Law

Constitution → Constitutional Laws → EU sources → Ordinary Law → Law and Legislative Decrees → Regional Laws

Web Application



Our Goal

SENTENZA N.

ANNO 2021

REPUBBLICA ITALIANA
IN NOME DEL POPOLO ITALIANO

LA CORTE COSTITUZIONALE

Ritenuto in fatto

...

Considerato in diritto

...

PER QUESTI MOTIVI

LA CORTE COSTITUZIONALE

...



Legal Data

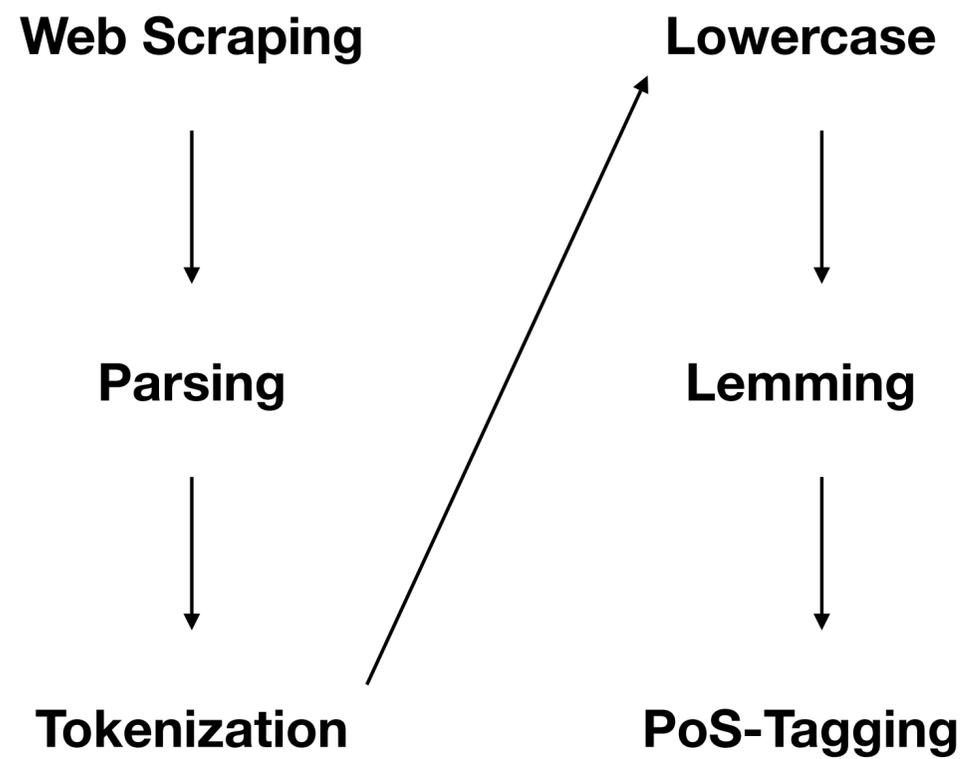
Natural Language
Processing

Search Engine

Web Application



Our Goal



Legal Data

**Natural Language
Processing**

Search Engine

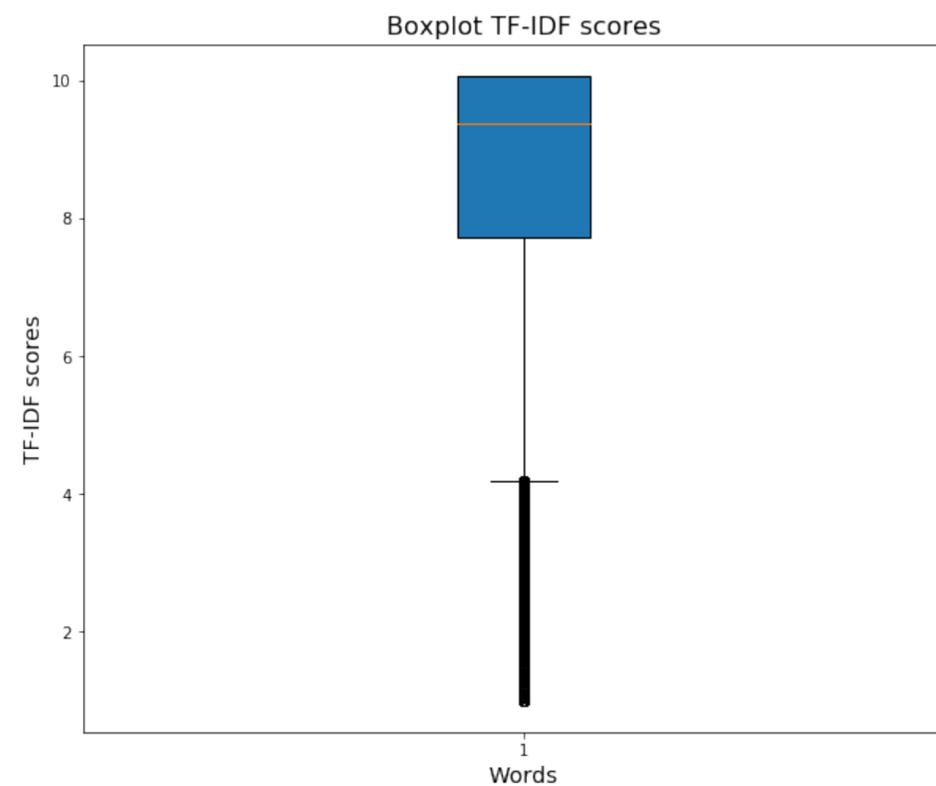
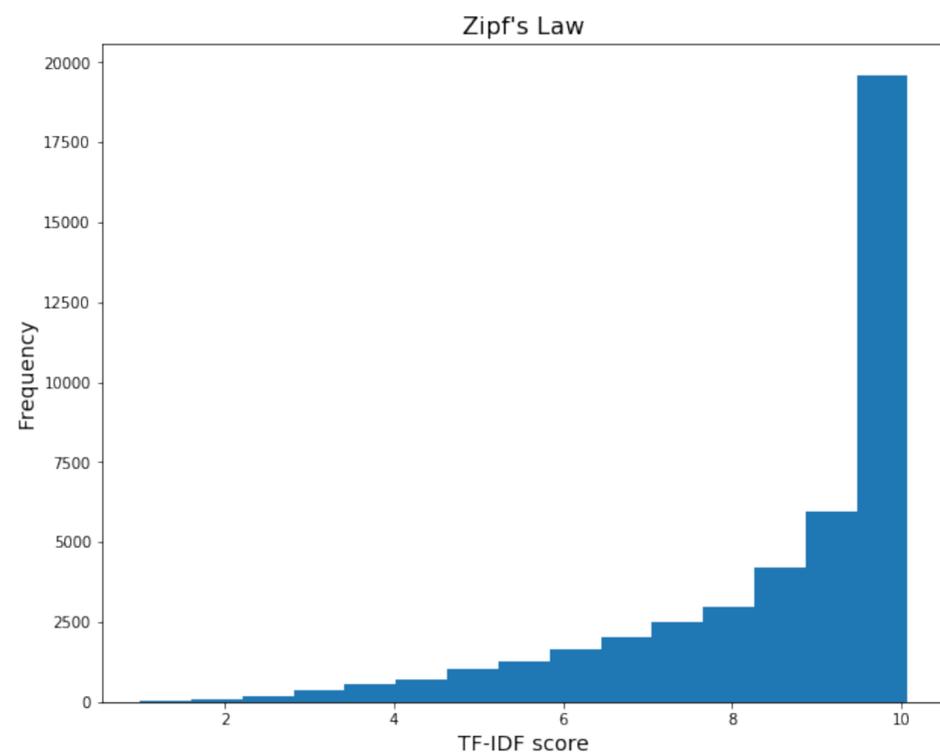
Web Application



Stopwords: Generic - Specific - Misspelling

$$TF_{ij} = \frac{n_{ij}}{|d_j|}, IDF_i = \log \frac{|D|}{|\{d : i \in D\}|}$$

*Lower bound = quartile1 - (1.5 * IQR)*



Legal Data

Natural Language
Processing

Search Engine

Web Application



Our Goal

Search Engine

Vertical Search Engine

- **Volume**
- **Document Size**
- **Structure**
- **Heterogeneity of document types**
- **Legal Hierarchy**
- **Temporal Aspects**
- **Citations**
- **Terminology**
- **Audience**
- **Personal Data**

Legal Data

**Natural Language
Processing**

Search Engine

Web Application



Our Goal

Search Engine

Legal Data

Indexing & Storing Data

**Natural Language
Processing**

Search Engine

Web Application



Our Goal

Search Engine

Legal Data

Indexing & Storing Data

**Natural Language
Processing**

Querying & Retrieval Functionality

Search Engine

Web Application



Our Goal

Search Engine

Legal Data

Indexing & Storing Data

**Natural Language
Processing**

Querying & Retrieval Functionality

Search Engine

Ranking

Web Application



Keyword Extractors



Simple Statistical

Graph-based

TF-IDF

Rake

Yake



Text-Rank

K-Core

K-Truss

Sgrank

Scake

Legal Data

Natural Language
Processing

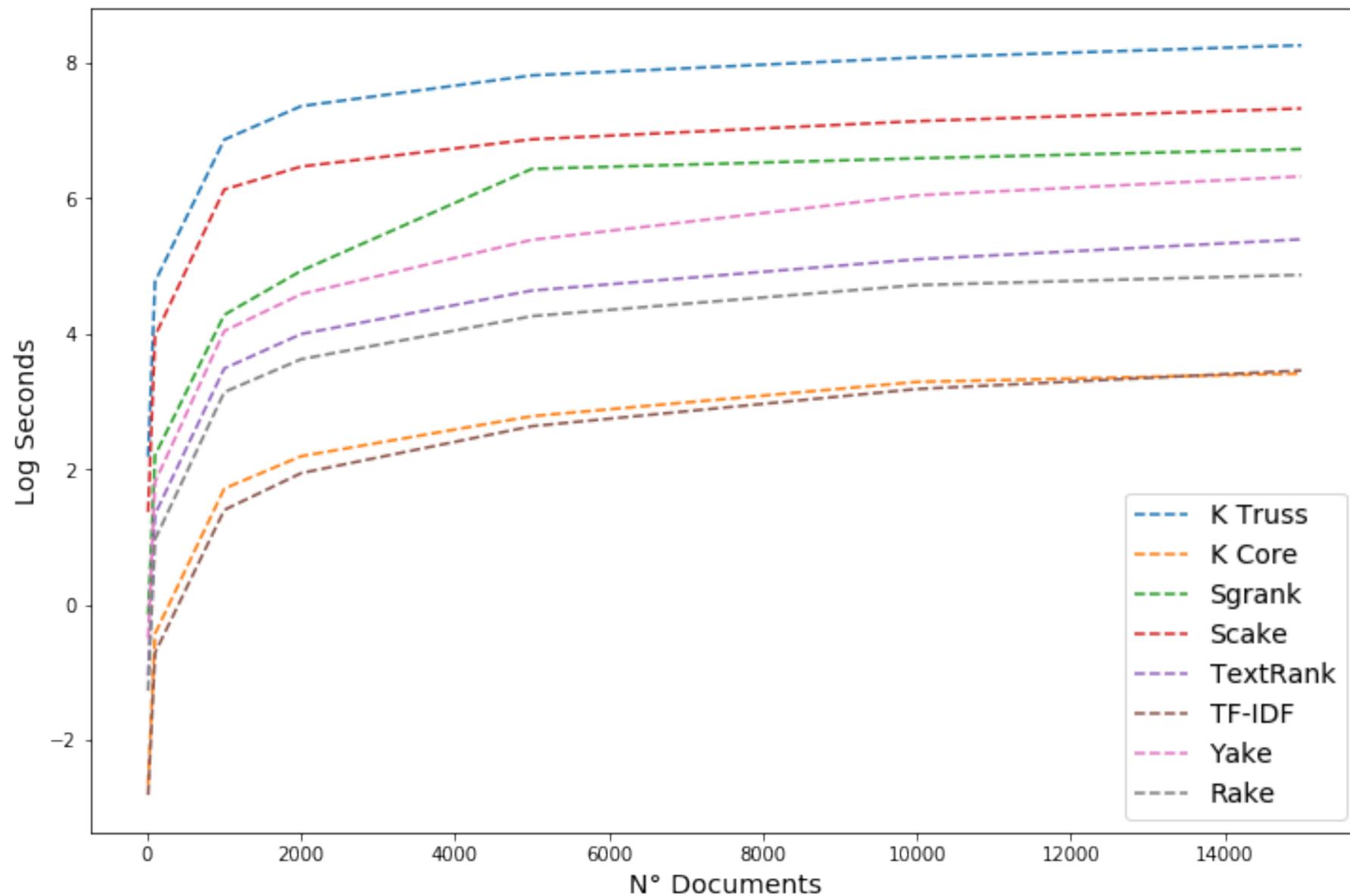
Search Engine

Web Application



Our Goal

Keyword Extractors



Legal Data

**Natural Language
Processing**

Search Engine

Web Application



Embedding

Word2Vec

- **Input layer:**

containing as many neurons as there are words in the corpus vocabulary;

- **Hidden layer:**

the number of neurons contained here defines the dimensionality of the resulting word embeddings;

- **Output layer:**

a layer containing the same number of neurons as the input layer.

Legal Data

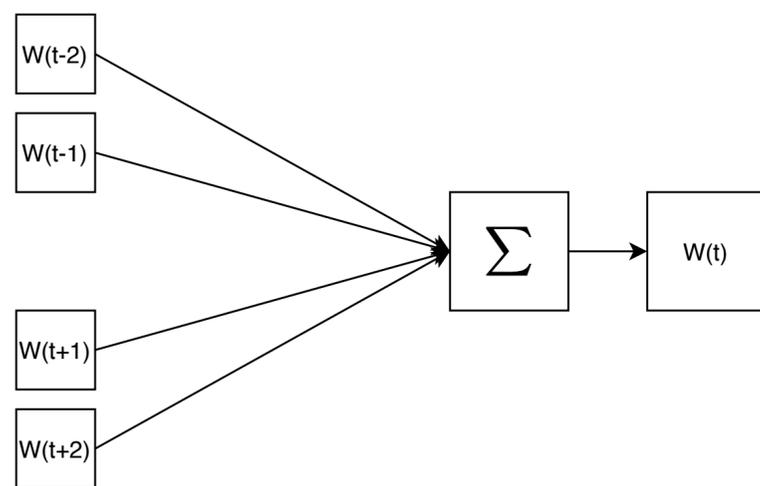
Natural Language
Processing

Search Engine

Web Application

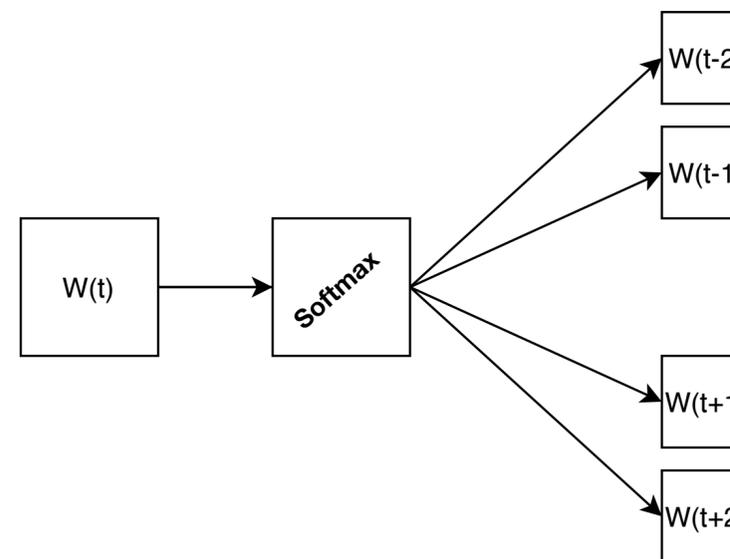


Embedding



**Continuous
Bag-of-Words**

Skip-Gram



Legal Data

Natural Language
Processing

Search Engine

Web Application



Web Application

- **No installation is required;**
- **Lower technical requirements;**
- **Updates are applied centrally;**
- **There are no incompatibility issues;**
- **Access everywhere with web browser;**
- **Subscription-based web applications.**

Human-Computer Interaction goals

- **Effectiveness**
- **Efficiency**
- **Utility**
- **Safety**
- **Memorability**
- **Learnability**

Legal Data

**Natural Language
Processing**

Search Engine

Web Application



Who We Are

I obtained a bachelor's degree in Cognitive Psychology in 2018 and, maybe, a master's degree in Data Science in 2021 with a thesis on a search engine for the judgments of the Constitutional Court. I am co-inventor of the international patent Method of authenticating a user through the analysis of changes in the external ear canal. I started my research by publishing the first article in 2020. My interests include cybersafety, usability of security, and human-computer interaction.



Our Mission

We propose an Italian Search-Engine for law practice. The goal of this thesis is to make available to Italian law firms a tool in which to retrieve jurisprudence cases. We address the application of methods of Legal Analytics (LA) of case law and empirical legal data, in particular judgements, using text mining. We implement the state-of-art of Natural Language Processing to the legal domain. We apply computational models to legal reasoning and behaviour, to extract legal knowledge, infer relationships, and encourage data-driven insights. We build a working prototype of search engine for a web application that requires minimal engineering and programming to bring the system at a level closer to the market.



Challenges & Future Works

Citations are do not respect any patterns.

Evaluation A/B Testing & Experts in the field.

Improve HCI and correlated section.

Legal Data

Natural Language
Processing

Search Engine

Web Application



Our Goal

Conclusions

In this project, we started by obtaining documents, as judgements and ordinances, and we concluded with the implementation of a working prototype of three different search engines techniques we have explored: inverted index, keyword extractors, and word and document embeddings.

Legal Data

**Natural Language
Processing**

Search Engine

Web Application



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Our Goal

Legal Data

**Natural Language
Processing**

Search Engine

Web Application

Thanks



UNIVERSITÀ
DEGLI STUDI
DI PADOVA





Inverted Index

Term	Document #1	Document #2
best	X	
carbonara		X
delicious		X
pasta	X	X
pesto	X	
recipe	X	X
the	X	
with	X	

Legal Data

Natural Language
Processing

Search Engine

Web Application



Our Goal

Web Application

Legal Data

Backend in Flask (Python)

**Natural Language
Processing**

Fronted in Bootstrap5 HTML, JS & CSS

Search Engine

Web Application



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

`app.route(/home)`

No methods needed

Law and Order: a Search Engine for Italian Law

Home Search Engine Results

Who We Are

I obtained a bachelor's degree in Cognitive Psychology in 2018 and, maybe, a master's degree in Data Science in 2021 with a thesis on a search engine for the judgments of the Constitutional Court. I am co-inventor of the international patent Method of authenticating a user through the analysis of changes in the external ear canal. I started my research by publishing the first article in 2020. My interests include cybersafety, usability of security, and human-computer interaction and usability of security.

Our Mission

we propose an Italian Search-Engine for law practice. The goal of this thesis is to make available to Italian law firms a tool in which to retrieve jurisprudence cases. We address the application of methods of Legal Analytics (LA) of case law and empirical legal data, in particular judgements, using text mining. We implement the state-of-art of Natural Language Processing to the legal domain. We apply computational models to legal reasoning and behaviour, to extract legal knowledge, infer relationships, and encourage data-driven insights. We build a working prototype of search engine for a web application that requires minimal engineering and programming to bring the system at a level closer to the market.



Our Goal

Legal Data

Natural Language
Processing

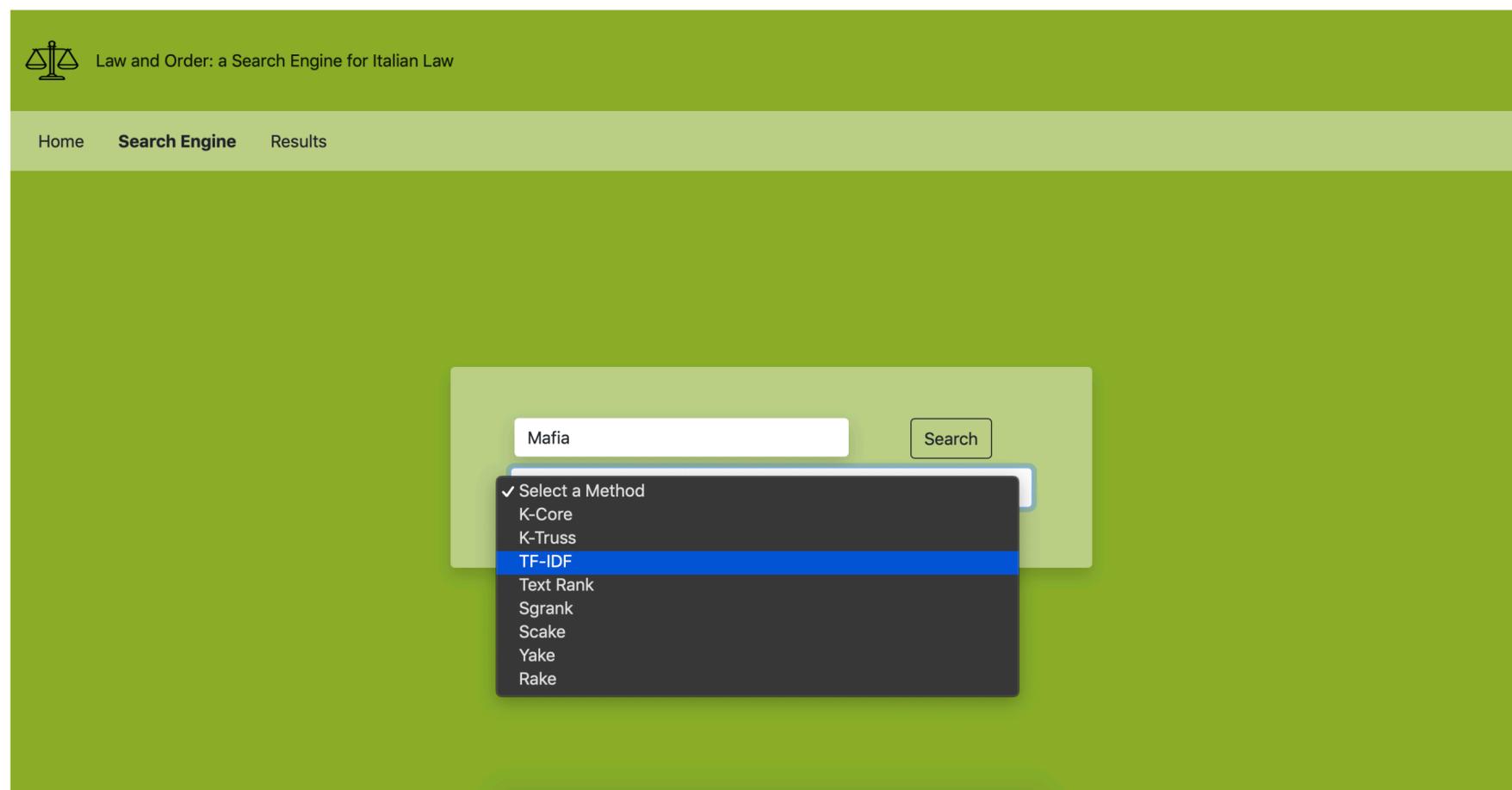
Search Engine

Web Application



`app.route(/search_engine)`

GET method



Legal Data

Natural Language
Processing

Search Engine

Web Application



app.route(/results)

POST method

Law and Order: a Search Engine for Italian Law

Home Search Engine Results

SENTENZA N. 144 ID: 13971 09/07/2015

SENTENZA N. 144 ANNO 2015 REPUBBLICA ITALIANA IN NOME DEL POPOLO ITALIANO LA CORTE COSTITUZIONALE composta dai signori: Presidente: Marta CARTABIA; Giudici : Giuseppe FRIGO, Paolo GROSSI, Giorgio LATTANZI, Aldo CAROSI, Mario Rosario MORELLI, Giancarlo CORAGGIO, Giuliano AMATO, Silvana SCIARRA, Daria de PRETIS, Nicolò ZANON, ha pronunciato la seguente SENTENZA nel giudizio per conflitto di attribuzione tra poteri dello Stato sorto a seguito della deliberazione del Senato della Repubblica ...

[Expand All](#)

— Keywords: ['senato', 'senato', 'xvi', 'legislatura', 'opinioni', 'senato', 'senato', 'monocratica', 'senato', 'senato']

ORDINANZA N. 317 ID: 4295 17/12/2013

ORDINANZA N. 317 ANNO 2013 REPUBBLICA ITALIANA IN NOME DEL POPOLO ITALIANO LA CORTE COSTITUZIONALE composta dai signori: Presidente: Gaetano SILVESTRI; Giudici : Luigi MAZZELLA, Sabino CASSESE, Giuseppe TESAURO, Paolo Maria NAPOLITANO, Giuseppe FRIGO, Alessandro CRISCUOLO, Paolo GROSSI, Giorgio LATTANZI, Aldo CAROSI, Marta CARTABIA, Sergio MATTARELLA, Mario Rosario MORELLI, Giancarlo CORAGGIO, Giuliano AMATO, ha pronunciato la seguente ORDINANZA nel giudizio per conflitto di attribuzion ...

[Expand All](#)

— Keywords: ['senato', 'opinioni', 'confl', 'pot', 'amm', 'senato', 'xvi', 'legislatura', 'senato', 'risarcitoria']

Legal Data

Natural Language
Processing

Search Engine

Web Application



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Our Goal

app.route(/144)

POST method

 Law and Order: a Search Engine for Italian Law

Home Search Engine **Results**

SENTENZA N. 144 **09/07/2015**
— ['senato', 'senato', 'xvi', 'legislatura', 'opinioni', 'senato', 'senato', 'monocratica', 'senato', 'senato', 'xvi', 'legislatura', 'risarcitoria', 'senatore', 'opinioni', 'membro', 'insindacabili', 'affermazioni', 'lesive', 'immagine']

SENTENZA N. 144 ANNO 2015 REPUBBLICA ITALIANA IN NOME DEL POPOLO ITALIANO LA CORTE COSTITUZIONALE composta dai signori: Presidente: Marta CARTABIA; Giudici : Giuseppe FRIGO, Paolo GROSSI, Giorgio LATTANZI, Aldo CAROSI, Mario Rosario MORELLI, Giancarlo CORAGGIO, Giuliano AMATO, Silvana SCIARRA, Daria de PRETIS, Nicolò ZANON, ha pronunciato la seguente SENTENZA nel giudizio per conflitto di attribuzione tra poteri dello Stato sorto a seguito della deliberazione del Senato della Repubblica del 20 dicembre 2012 (Atti Senato, XVI legislatura, Doc. IV-ter, n. 30-A), relativa alla insindacabilità, ai sensi dell'art. 68, primo comma, della Costituzione, delle opinioni espresse dall'on. Maurizio Gasparri nei confronti di Marco Travaglio, promosso dal Tribunale ordinario di Roma, prima sezione civile, con ricorso notificato il 14 febbraio 2014, depositato in cancelleria l'11 marzo 2014 ed iscritto al n. 10 del registro conflitti tra poteri dello Stato 2013, fase di merito. Visti l'atto di costituzione del Senato della Repubblica, nonché l'atto di intervento di Marco Travaglio; udito nell'udienza pubblica del 9 giugno 2015 il Giudice relatore Giuseppe Frigo; uditi gli avvocati Francesco Saverio Marini per il Senato della Repubblica e Paola Rizzo per Marco Travaglio. Ritenuto in fatto 1.- Con ordinanza-ricorso depositata il 2 agosto 2013, il Tribunale ordinario di Roma, in composizione monocratica, ha promosso conflitto di attribuzione tra poteri dello Stato in ordine alla deliberazione del Senato della Repubblica del 20 dicembre 2012 (Atti Senato, XVI legislatura, Doc. IV-ter, n. 30-A), con cui l'Assemblea ha affermato che le dichiarazioni in relazione alle quali, nel giudizio civile pendente davanti a detto giudice, è stata avanzata domanda risarcitoria da parte di Marco Travaglio nei confronti del senatore Maurizio Gasparri, concernono opinioni espresse da un membro del Parlamento nell'esercizio delle sue funzioni e sono pertanto insindacabili ai sensi dell'art. 68, primo comma, della Costituzione. Il giudice ricorrente premette che Marco Travaglio ha proposto domanda di risarcimento dei danni derivanti dalle affermazioni, lesive della propria immagine e diffamatorie, rilasciate dal convenuto senatore nel corso dei programmi televisivi RAI "Rainews24" (25 settembre 2009) e "Porta a porta" (15 ottobre 2009), nonché in successive pubbliche dichiarazioni riprodotte dalle agenzie di stampa (28 settembre, 1° 9 e 17 ottobre 2009): in particolare, secondo l'attore, il senatore Gasparri avrebbe aspramente criticato la moralità di Travaglio, accusandolo di essere andato in vacanza in Sicilia a

Judges

composta dai signori: Presidente: Marta CARTABIA; Giudici : Giuseppe FRIGO, Paolo GROSSI, Giorgio LATTANZI, Aldo CAROSI, Mario Rosario MORELLI, Giancarlo CORAGGIO, Giuliano AMATO, Silvana SCIARRA, Daria de PRETIS, Nicolò ZANON,

Legal Data

Natural Language
Processing

Search Engine

Web Application