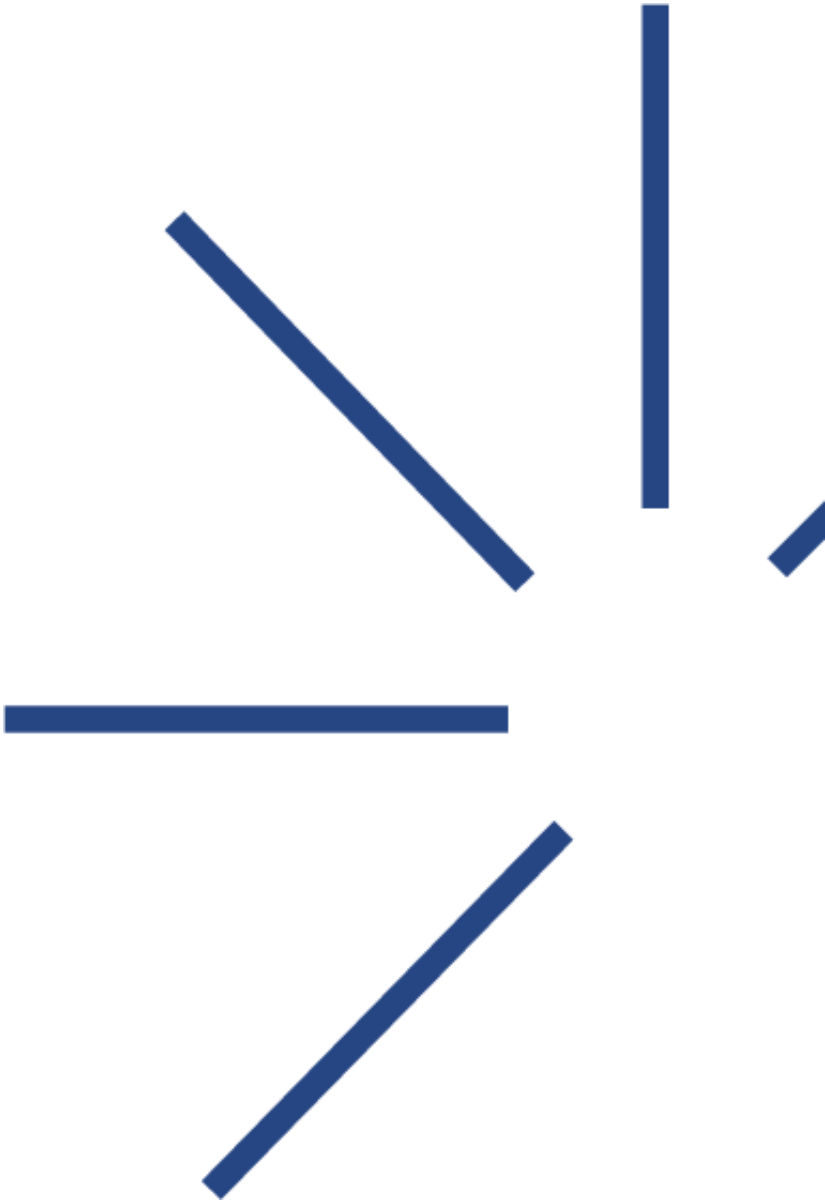
D4 – Sample data



Project funded by the European Union's Next Generation EU through the Slovak Republic's Recovery and Resilience Plan under project no. 09I05-03-V02-00049.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU
VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA
VÝSKUMNÁ AGENTÚRA

Introduction

The project **"Automated Legal Text Analysis based on Machine Learning**" (hereinafter "ALTAML") is a key initiative aimed at integrating innovative approaches in the field of data processing and subsequent analysis, with a focus on the field of law, which also includes information and communication technology law. The aim of this project is to develop and validate effective methods for the automated analysis of legal texts using machine learning techniques. This involves, in particular, the development of tools that facilitate the processing and analysis of large volumes of data in the form of various legal documents, as well as the extraction of relevant information (attributes) from such documents, including the identification of key terms, references to legislation and other attributes.

The aim of the ALTAML project is therefore to contribute to more efficient access to legal information and to the acceleration of legal processes, thereby ensuring a higher degree of legal certainty and improving the availability of legal texts, the results of their analysis and relevant legal information for both professionals and the general public.

As part of work package KPB1, deliverable D4 - **Sample data -** creating a set of sample data. Due to a large dataset of resources, only part of the processed documents will be published for demonstration

Contents

Introduction **1**

1 Sources of sample data **3**

2 Description of sample data **4**

1 Sources of sample data

The datasets contain decisions according to the source of acquisition in three formats: PDF, XML, and HTML.

The sources of the decisions are described in the report and are as follows:

1) justice.gov.sk – website of the Ministry of Justice – decisions of general courts, the Supreme Court of the Slovak Republic, and the Supreme Administrative Court of the Slovak Republic,
2) nsud.sk – website of the Supreme Court of the Slovak Republic – decisions of the court,
3) nssud.sk – website of the Supreme Administrative Court of the Slovak Republic – decisions of the court,
4) ustavnysud.sk – website of the Constitutional Court of the Slovak Republic – decisions of the court,
5) nsud.sk - website of the Supreme Court of the Slovak Republic - Collection of decisions and opinions of the Supreme Court of the Slovak Republic,
6) nssud.sk - website of the Supreme Administrative Court of the Slovak Republic - Collection of opinions and decisions of the Supreme Administrative Court of the Slovak Republic,
7) ustavnysud.sk - website of the Constitutional Court of the Slovak Republic - Collection of findings and decisions of the Constitutional Court of the Slovak Republic.

2 Description of sample data

Legend:

source	*source of the decision*
url	*address of the specific decision*
id	*identification in our database*
sp_zn	*file number of the decision*
date	*date of decision*
typ	*type of decision*
titul	*title of the decision (either obtained from the source or created by WK)*
summary	*abridged content of the decision created by WK*
autor	*court that issued the decision*
pdf	*name of the original decision obtained in pdf format*
xml	*name of the xml file*
html	*name of the html file*
zip	*name of the zip file containing the decisions*

1.	**Ministry of Justice of the Slovak Republic**
source	justice.gov.sk
URL	https://www.justice.gov.sk/sudy-a-
ID	JUD4287149SK
sp_zn	37Ca/1/2016
date	9.12.2022
typ	Rozsudok
titul	Občianske právo - Zmluva o dielo; Občianske
summary	Ochrana autorského práva a zakázanie
autor	Okresný súd Košice I
pdf	Rozsudok_37Ca-1-2016.pdf
xml	ju_4287149_sk-JUD4287149SK.xml
html	ju_4287149_sk-JUD4287149SK.html
zip	01.zip

2.	**Supreme Court of the Slovak Republic**
source	https://www.nsud.sk/rozhodnutia/
URL	https://www.nsud.sk/rozhodnutia/4cdo2402020/
ID	JUD4347552SK
sp_zn	4Cdo/240/2020
date	28.3.2023
typ	Uznesenie

titul	určenie autorského práva s nárokom na náhradu nemajetk. ujmy
summary	Odmietnutie dovolania v spore o určenie autorského práva a náhradu nemajetkovej ujmy
autor	Najvyšší súd SR - senát
pdf	751519.9c7c1a.pdf
xml	ju_4347552_sk-JUD4347552SK.xml
html	ju_4347552_sk-JUD4347552SK.html
zip	02.zip

3.	**Supreme Administrative Court of the Slovak Republic**
source	https://www.nssud.sk/
URL	https://www.nssud.sk/wp-content/uploads/7Svk492022.pdf
ID	JUD4566407SK
sp_zn	7Svk/49/2022
date	27.3.2024
typ	Rozsudok
titul	Udelenie oprávnenia na výkon kolektívnej správy práv zvukárov
summary	Zrušenie rozhodnutia o kolektívnej správe práv zvukových
autor	Najvyšší správny súd SR - senát
pdf	Rozsudok_37Ca-1-2016.pdf
xml	ju_4287149_sk-JUD4287149SK.xml
html	ju_4287149_sk-JUD4287149SK.html
zip	03.zip

4.	**Constitutional Court of the Slovak Republic**
source	https://www.ustavnysud.sk/rozhodovacia_cinnost/rozhodnutia
URL	https://www.ustavnysud.sk/docDownload/4444fd9a-3648-4660-92dd-73b1150a1bbc
ID	JUD2616598SK
sp_zn	III. ÚS 651/2016 - 90
date	28.11.2017
typ	Nález
titul	Porušenie základných práv pri poskytovaní monitoringu médií bez súhlasu, čo zahŕňalo neoprávnené vyhotovovanie rozmnožení článkov
summary	Porušenie základných práv MAFRA Slovakia v súvislosti s autorskými právami
autor	Ústavný súd SR - senát
pdf	III. ÚS 651_2016.pdf

xml	ju_2616598_sk-JUD2616598SK.xml
html	ju_2616598_sk-JUD2616598SK.html
zip	04.zip

5.	**Collection of decisions and opinions of the Supreme Court of the Slovak Republic**
source	https://www.nsud.sk/zbierka-stanovisk-a-rozhodnuti/
URL	https://www.nsud.sk/data/att/676/637297.f1bf1b.pdf
ID	JUD4242507SK
sp_zn	2 CdoGp 1/2020
date	24.2.2022
typ	Uznesenie
titul	Absolútne majetkové autorské práva. Odúmrť. Voľné dielo.
summary	K otázke prechodu dedičstva v prípade absolútnych
autor	Najvyšší súd SR - senát
pdf	637297.f1bf1b.pdf
xml	ju_4242507_sk-JUD4242507SK.xml
html	ju_4242507_sk-JUD4242507SK.html
zip	05.zip

6.	**Collection of opinions and decisions of the Supreme Administrative Court of the Slovak Republic**
source	https://www.nssud.sk/zbierka-stanovisk-a-rozhodnuti-najvyssieho-spravneho-sudu-slovenskej-republiky/
URL	https://www.nssud.sk/wp-content/uploads/Zbierka_2_2023.pdf
ID	JUD4390154SK
sp_zn	33 D 17/2021
date	17.11.2022
typ	Rozhodnutie
titul	Disciplinárne previnenie; viazanosť právnym názorom Ústavného súdu Slovenskej republiky
summary	porušenie povinnosti sudcu pri nerešpektovaní rozhodnutia ústavného súdu
autor	Najvyšší správny súd - senát (ostatné)
pdf	Zbierka_3_2023.pdf
xml	ju_4390154_sk-JUD4390154SK.xml
html	ju_4390154_sk-JUD4390154SK.html
zip	06.zip

7.	**Collection of findings and decisions of the Constitutional Court of the Slovak Republic**
source	https://www.ustavnysud.sk/rozhodovacia_cinnost/zbierka-n%C3%A1lezov-a-uznesen%C3%AD
ID	JUD3178033SK
sp_zn	PL. ÚS 9/2016
date	7.2.2018
typ	Nález
titul	Hudobné kvóty
summary	Spor o ústavnosť imitácie slovenských kultúrnych prvkov v rozhlasovom vysielaní, definícia slovenského hudobného diela prostredníctvom národných prvkov
author	Ústavný súd SR - plénum
pdf	Dokument-28.pdf
xml	ju_3178033_sk-JUD3178033SK.xml
html	ju_3178033_sk-JUD3178033SK.html
zip	07.zip

This document also includes 2 annexes:

1. Data sets – description I (Part I: the "Copyright" and "Trademarks / Unfair Competition" domains)

- The document describes 17 data packages created for the copyright and trademarks/unfair competition domains, containing documents in html format.
- The documents were selected into the individual packages on the basis of full-text queries or metadata queries in the database, with the queries themselves formulated according to the specifications provided by the project partners.
- The material is divided into four parts: the partners' specifications with references to the package numbers (e.g. provisions of Copyright Act No. 185/2015 Coll., Sections 442a and 458a of the Civil Code, Directives 2001/29/EC and 2006/115/EC, CJEU case law in cases C-162/10 and C-306/05, Act No. 506/2009 Coll. on Trademarks and Sections 44 to 48 of the Commercial Code), a detailed overview of the individual packages including the wording of the meta-queries and full-text queries, a description of the result sets, and a list of annexes. Each package has a corresponding overview in .xlsx format (datovy-balik-1.xlsx to datovy-balik-17.xlsx) with an "altaml" sheet listing all documents and, in most cases, a "po_redukcii" sheet containing a narrowed selection based on an analysis of the case file number (registers "Ca" and "Co" for copyright, "CbPV" and "CoPV" for trademarks). The document also explains the meaning of the individual columns in the result sets (query, case file number, database ID, date of decision, source, legal area, register, ECLI and others).

2. Data sets – description II (Part II: the "Digital Evidence in Criminal Proceedings"

domain)

- The document describes 14 data packages prepared for the digital evidence in criminal proceedings domain, again containing documents in html format selected by means of full-text and metadata queries based on the partners' specifications.
- The specifications draw primarily on provisions of the Code of Criminal Procedure (Sections 3, 89a, 90, 91, 92, 96d, and 100–118), Act No. 236/2017 Coll. on the European Investigation Order, the Convention on Cybercrime, Directive 2013/40/EU on attacks against information systems, and Regulation (EU) 2023/1543 on European Production Orders and European Preservation Orders for electronic evidence.
- Alongside standalone search criteria, cumulative criteria were also applied, combining individual provisions with keywords from the field of digital data (content of communication, traffic and location data, data carrier, mobile phone, server, cloud, e-mail, GPS data, encrypted communication, cryptocurrency and others). As in Part I, each package here is accompanied by result sets in overviews (datovy-balik-18.xlsx to datovy-balik-31.xlsx) in both a full and a reduced version; the reduction was carried out through an analysis of the case file number with a focus on criminal agenda (registers "T", "To", "Tdo", "Tk", "Nt" and "ÚS"). Struck-through entries in the overview indicate queries that returned no results or results that were not relevant.