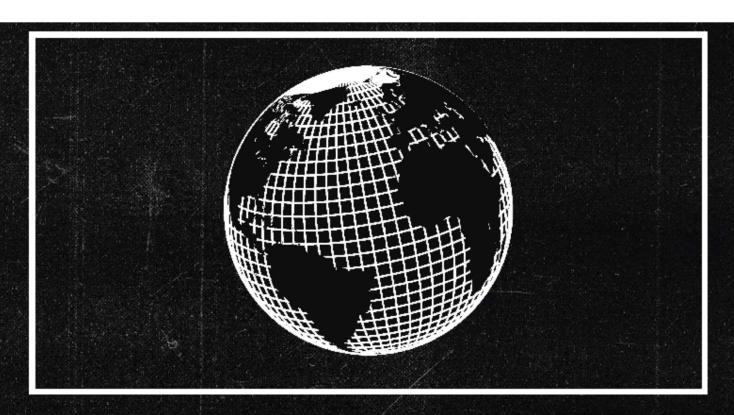
MY RIGHTS

HUDOC's FUNDAMENTAL FLAWS



Applying human centered design principles **1**n addition to alternative search methods to the European Court of Human Right's Judgments Database.



Accessibility Autonomy Unde rstanding · Ownership · Accessi bility · Autonomy · Understandi ng · Ownership · Accessibility · Autonomy · Understanding · Owne rship · Accessibility · Autonom





MY RIGHTS

MY RIGHTS

MY RIGHTS

MY RIGHTS















Tech Solutions for Human Rights

My-Rights is a Tech Solutions for Human Rights project (\mathbf{TSHR}) .

TSHR is a non-profit organisation that enables substantive access to human rights through the use of free and innovative technology while implementing Human-centered design principles.

TSHR has received funding from international NGOs and governments to produce reports, data gathering tools, web apps, or mobile apps for various human rights issues.

All TSHR products are free to users, collect no user data at all, and are designed with decolonial principles and accessibility concerns at their core.

This document provides details on the data pipeline that occurs from scraping the European Court of Human Rights' Database to the multiple inputs into www.My-Rights.Info (My-Rights). My-Rights is currently funded by the Prototype Fund and the Bundesministerium für Bildung und Forschung.

www.My-Rights.Info is the world's first and only Human-centered designed database for accessing human rights case law.

Viraaj Akuthota Founder of TSHR

www.tshr.io
www.my-rights.info
lawitapps@gmail.com



Executive Summary

The European Court of Human Rights is arguably one of the most important courts in the world as it delivers the greatest number of human rights judgments, its jurisdiction is wide and compulsory, and its decisions are highly respected outside of its jurisdiction. Unfortunately, its official database - $\underline{\text{HUDOC}}$ - is fundamentally flawed (HUDOC).

The current state of the HUDOC makes the job of already overburdened human rights practitioners even more challenging and, most disappointingly, impacts the ability to protect the rights of future potential human rights victims. Despite communicating these issues with the Head of HUDOC, little seems to have changed. It should be noted that the issues found within HUDOC were discovered as a result of creating the My-Rights.

Through the process of data extraction, TSHR identified many systemic issues the following provides a brief overview:

- when the same historical search was run on different days-covering the identical period from 1949 to 1991 the results vary, returning 4,125 cases one day and 4,117 another, with no basis to identify the discrepancy.
- of the hundreds of thousands of cases in the database, 24,343 result links lead nowhere, trapping users in endless redirects rather than delivering the promised legal decisions.
- 171 cases exist as empty shells appearing in search results with metadata but containing no actual judgment text.
- the database's search filters have become actively misleading. Language filters that ignore user selections and article filters that reference legal provisions that simply do not exist, such as the fictional "Protocol 4-6-2" and "Protocol 7-7-2."

The most disappointing aspect has been the institutional response to these documented problems. Since 2023, detailed reports of these issues have been submitted to HUDOC's management team, complete with specific case examples and technical evidence. While management acknowledged the problems and claimed they were resolved during a database upgrade, investigation revealed that only the sample cases provided were fixed—the underlying systematic issues remained untouched. This pattern suggests either a fundamental misunderstanding of the problems' scope or an inability to address them at the scale required.



Contents

Executive Summary	3
Introduction	5
Communication with HUDOC Management	6
Issues Found when Recreating the HUDOC database	7
Cannot Rely on HUDOC's Case Detail Section	8
HUDOC Search Filtering Issues	9
Language filters do not work	9
Impossible to Identify Number of Unique Cases	11
24,343 result links do not lead to a case	11
7 results do not exist entirely	11
171 cases exist but contains no judgment text with only the HUDOC logo being visible	12
Annexure A - 1	13
Annexure A = 2	14



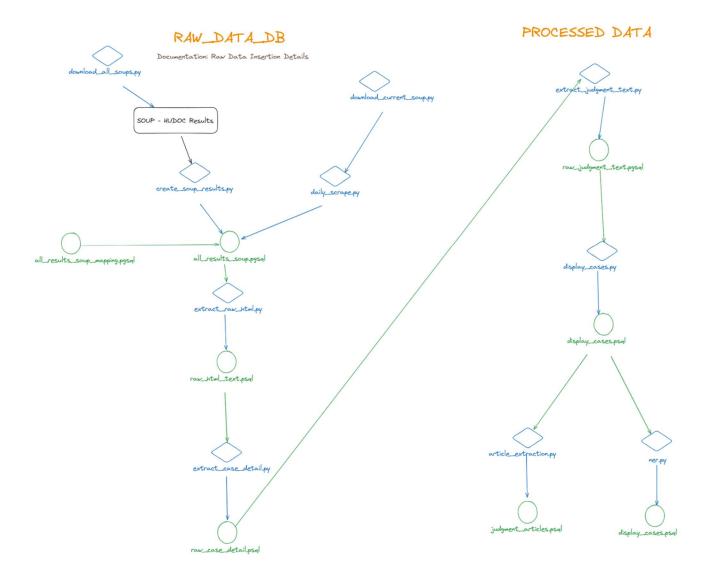
Introduction

My-Rights aims to be a free legal lawyer for rights holders and a research aid for practitioners. It applies decolonial and human-centered design principles to ensure the tool is transparent and meaningfully accessible. This tool is only applicable to the European Court of Human Rights. However, it is free and open-source software and can be easily adapted to any court.

My-Rights consists of four different entry points as it does not assume that there is only one method to obtain and understand information. These four entry points are:

- ❖ Semantic Text Search Tool
- ❖ Data Visualisation Navigation
- ❖ Binary Decision Branching Logic Questionnaire
- ❖ Human Rights Violation Prediction Tool.

This document details the data pipeline that feeds into all the above entry points. The following represents the current stage of the ETL process, which will be referenced throughout this document.





The data that feeds My-Rights is entirely from HUDOC. Unfortunately, HUDOC has significant data quality issues, which critically affect the ability to build a program that is based on top of this database. At each part of the data extraction process, unique issues are found, which makes it extremely difficult to effectively create safeguards for identifying specific issues or create code to extract specific fields. Instead, the entire database has to be recreated with an independent system. Therefore, the entirety of HUDOC must be scraped and the results (cases) recreated from scratch. Due to issues discussed later in this report, the recreation of the database must occur on a daily basis.

The remainder of the report provides detail to the range of issues found at each step of the ETL process.

Communication with HUDOC Management

Since 2023, I have been in contact with the Head of HUDOC regarding their data quality issues. I emailed HUDOC with a list of all their distinct issues, and then provided a sample Case ID to prove the existence of those issues. After sending several follow-up emails, HUDOC responded months later stating that they could not reply as they were in the process of upgrading their database. Approximately one year from the date of the initial emails, I observed that the HUDOC website fixed the issues that I highlighted only for the sample cases but not for all the cases that shared the same underlying issue. I raised this point to them via email. They responded but did not reply to the substantive content of the email. I reproduce the relevant emails in a redacted format to preserve confidentiality:

```
Dear X

I hope you are doing well.

I am hoping to follow up on the previous email chain.

I have noticed you made some corrections to the cases that I linked you. I was wondering if you made any further changes.

To note, I only provided you a sample of issues under each category but was wondering if your database upgrade solved the underlying nature of the data quality issues?

Kind regards,

...

Dear Viraaj,

My apologies for the belayed reply.

We are still working on some migrations but those are now internal so all glitches you saw in the past should be solved.
```

Shall you notice any issue in the future, don't hesitate to get back to me.



Issues Found when Recreating the HUDOC database

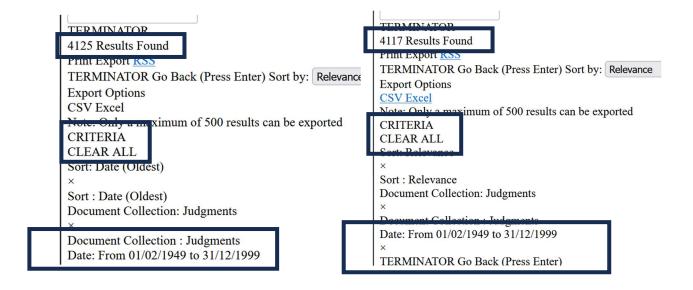
The primary issue identified when attempting to create a snapshot of HUDOC at any point in time is that:

Unfiltered searches of the same historical period produce different result counts when run at different times.

For example, on 18 October 2022, a search for all cases was run for the period from 1 February 1949 to 31 December 1991. The number of results found for this period was different to the number of results found for the same search run on 21 February 2024. It is difficult to identify why HUDOC or the European Court of Human Rights would release or remove a case historically. Therefore, it is most likely, a data quality issue.

To remedy this issue, the entire **HUDOC database has to be fetched on a daily basis** and the data cleaning process also has to occur on the same basis.

The following screenshots from the raw HTML file provides evidence to the figures cited above:



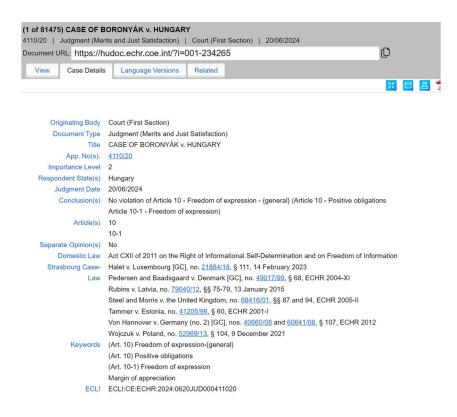
NEGATIVE IMPACT ON HUMAN RIGHTS RESEARCH

When lawyers conduct a keyword search, they typically identify how many results have been delivered, then read through a percentage of these cases. However, this approach does not work for HUDOC, as a keyword search might yield 500 cases on one day and only 6 cases on another for the same period. This makes it impossible to determine when research is complete or to gauge how many relevant cases have been reviewed as a percentage.



Cannot Rely on HUDOC's Case Detail Section

HUDOC's 'Case details' screen displays the metadata of a case. It is a separate screen from where the text of the case is located. The following screenshot displays an actual Case details screen:



For the entire HUDOC database, there exist numerous errors in the Case details screen. Some of these errors include:

- 1. Many different cases share the same Court Identifier number (ECLI), violating the system's requirement for unique identifiers (Annex A 1);
- 2. Case detail screens sometimes lack essential fields entirely, including missing ECLI fields (Annex A 2);
- 3. Case detail screens do not have consistent fields and some fields are completely empty (Compare Annex A 1 to 2).

For evidence relating to the above issues, please see Annexure A to this document.

NEGATIVE IMPACT ON HUMAN RIGHTS RESEARCH

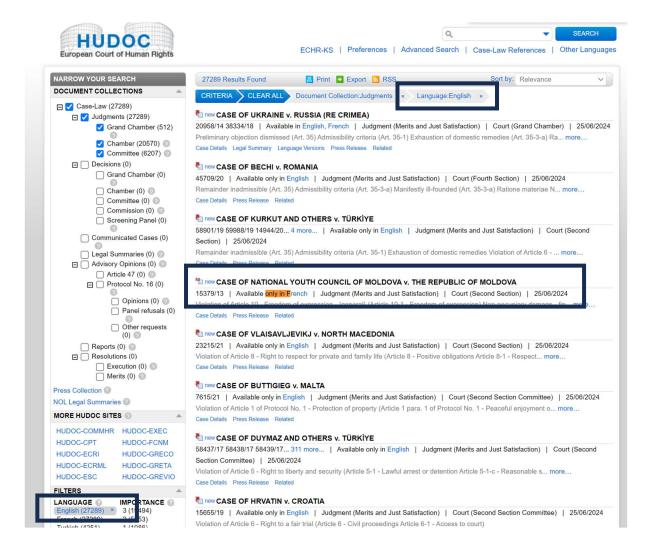
Due to the significant issues with the Case details screen, the advanced search filters for the HUDOC database become unreliable. When a lawyer utilizes the advanced search filters, the results are misleading, as they likely include cases that should not be included or omit cases that should be. This situation leads to lawyers being under the false impression that their research is complete, when it is in fact not.



HUDOC Database Search Filtering Issues

Language filters do not work

Filters, such as language filters, do not function appropriately for the HUDOC website. For example, if the language filter is set to English, cases that are delivered **only in French**, still appear:





Fictional Article Choices in HUDOC Search Filters

A number of Articles referenced in the Article search filter do not exist in the European Convention on Human Rights. Such Articles include but are not limited to the following:

- Protocol 4-6-2
- Protocol 7-7-2

The following screenshot demonstrates the filters that exist for the above fictional articles:

P4-4 (34)
P4-6 (1)
P4-6-2 (1)
P6-1 (9)
P6-2 (1)
P7-1 (26)
P7-1-1 (23)
P7-1-2 (5)
P7-2 (176)
P7-2-1 (3)
P7-2-2 (2)
P7-3 (7)
P7-4 (126)
P7-4-1 (2)
P7-5 (12)
P7-7 (1)
P7-7-2 (1)

NEGATIVE IMPACT ON HUMAN RIGHTS RESEARCH

The malfunctioning language filters on the HUDOC website impede efficient legal research by displaying irrelevant cases, leading to wasted time and potential oversight of relevant information.

The presence of fictional articles in HUDOC's search filters further undermines the reliability of the database. Lawyers relying on these filters may spend considerable time examining non-existent articles, leading to confusion and potential misinterpretation of legal frameworks.



Impossible to Identify Number of Unique Cases

Due to a combination of data quality issues, missing cases, and links that do not directly lead to a case, it is impossible to readily identify how many unique cases the European Court of Human Rights has delivered.

To combat this issue, TSHR was required to create its own unique identification system since HUDOC's system cannot be relied upon. The remainder of the section highlights the distinct issues that make identifying the number of unique cases impossible.

24,343 result links do not lead to a case

When the scraping first occurred, there were roughly 24,434 cases where, if a user clicked on a result, they would not be redirected to the judgment itself. Instead, they would be redirected to a page that specifies the language in which the case is provided. Usually, this occurs if a result item displayed in the search is in English, but the underlying case is available only in French. Unfortunately, this often meant being redirected to an already existing case. It should be noted that, the European Court of Human Rights officially delivers cases only in English, French, or both.

7 results do not exist entirely

After accommodating for the incorrect result links, seven cases were identified where the HUDOC website stated that no official language versions exist. However, a link to the case was still present in the search results, and the respective case details also exist.

These issues can be grouped into cases that were either struck out or are translations. For the cases that are struck out:

- a. AFFAIRE KILIC KALKAN c. TURQUIE: Has no other related judgment, so no judgment text for this case can be found at all
- b. AFFAIRE OZCETIN c. TURQUIE: Itself does not contain judgement text. A search does lead to another case with judgement text in English, that it likely relates to. Further, this case does not have an ECLI number in its case details.



Two examples of the above issue are:

Id	Title	Col5_id/URL
5671	"AFFAIRE KILIC KALKAN c. TURQUIE"	"/eng#{""tabview"":[""document""],""itemid"":[""001-63635""]}"
5675	"AFFAIRE OZCETIN c. TURQUIE"	"/eng#{""tabview"":[""document""],""itemid"":[""001-63634""]}"

171 cases exist but contains no judgment text with only the HUDOC logo being visible

To identify the number of unique cases that exist, a de-duplication test was run. This test identified 171 cases where a judgement existed, but for that judgment there was no judgment text content.

The following is the SQL code to identify the duplicates and the relevant item IDs:

a. The md5 value is = 9d50214e3ab049b8d3d1266a91e080a4

SELECT
 MD5(judgment_html) AS judgment_hash,
 COUNT(*) AS duplicates_count

FROM
 raw_html_text
GROUP BY
 judgment_hash
HAVING
 COUNT(*) > 1
ORDER BY
 duplicates_count DESC;

For evidence, the following table provides 5 example cases, displaying the ECLI, where the issue can be found.

```
ECLI:CE:ECHR:200 | ECLI:CE:ECHR:200 | ECLI:CE:ECHR:200 | ECLI:CE:ECHR:200 | CLI:CE:ECHR:200 | ECLI:CE:ECHR:200 | ECLI:CE:ECHR:2
```

NEGATIVE IMPACT ON HUMAN RIGHTS RESEARCH AND HUMAN RIGHTS VICTIMS

Unlike the other issues referred to above, this issue is uniquely significant as human rights victims are left unable to identify the case that applies to them if they are caught by the above data quality issues.

Despite the European Court of Human Rights not being bound by strict precedent, past decisions are still highly influential. Unfortunately, all the beneficial information in these cases for future human rights victims is inaccessible for human rights victims.



Annexure A - 1

Title PUDAS v. SWEDEN

App. No(s). <u>12119/86</u> Decision Date <u>06/09/1989</u>

Conclusion(s) Partly admissible

Partly inadmissible

Article(s) 6

6-1

8

0

8-2 11

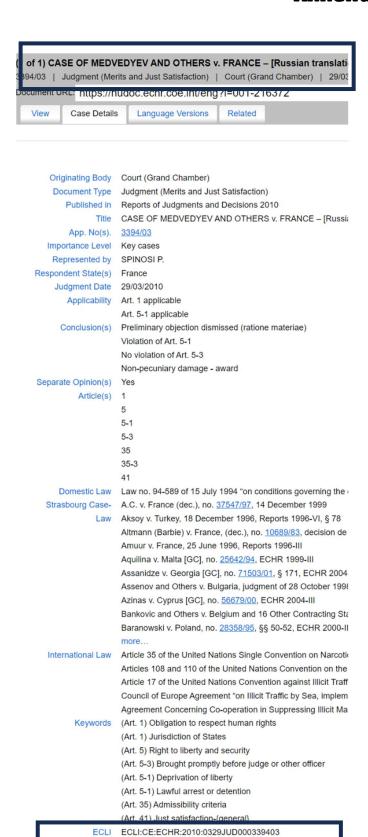
11-1

P1-1

P4-2



Annexure A - 2



(1 of 1) CASE OF LLOYD AND OTHERS v. THE UNITED KINGDOM 29798/96 30395/96 34327/96... | Judgment (Merits and Just Satisfaction) | Court (Fourth Section) 01/03/2005 Case Details Language Versions Related Originating Body Court (Grand Chamber) Document Type Judgment (Merits and Just Satisfaction) Published in Reports of Judgments and Decisions 2010 Title CASE OF MEDVEDYEV AND OTHERS v. FRANCE – [Russian translation] by Developme App. No(s), 3394/03 Importance Level Key cases Represented by SPINOSI P Respondent State(s) France Judgment Date 29/03/2010 Applicability Art. 1 applicable Art. 5-1 applicable Conclusion(s) Preliminary objection dismissed (ratione materiae) Violation of Art. 5-1 No violation of Art. 5-3 Non-pecuniary damage - award Separate Opinion(s) Yes Article(s) 5-1 5-3 35 35-3 Domestic Law Law no. 94-589 of 15 July 1994 "on conditions governing the exercise by the State of its po Strasbourg Case- A.C. v. France (dec.), no. 37547/97, 14 December 1999 Law Aksoy v. Turkey, 18 December 1996, Reports 1996-VI, § 78 Altmann (Barbie) v. France, (dec.), no. 10689/83, decision de la Commission du 4 July 198 Amuur v. France, 25 June 1996, Reports 1996-III Aquilina v. Malta [GC], no. 25642/94, ECHR 1999-III Assanidze v. Georgia [GC], no. 71503/01, § 171, ECHR 2004-II Assenov and Others v. Bulgaria, judgment of 28 October 1998, Reports 1998-VIII, §§ 146 Azinas v. Cyprus [GC], no. 56679/00, ECHR 2004-III Bankovic and Others v. Belgium and 16 Other Contracting States (dec.) [GC], no. 52207/9 Baranowski v. Poland, no. 28358/95, §§ 50-52, ECHR 2000-III International Law Article 35 of the United Nations Single Convention on Narcotic Drugs of 30 March 1961 Articles 108 and 110 of the United Nations Convention on the Law of the Sea, signed at Me Article 17 of the United Nations Convention against Illicit Traffic in Narcotic Drugs and Psyc Council of Europe Agreement "on Illicit Traffic by Sea, implementing article 17 of the United Agreement Concerning Co-operation in Suppressing Illicit Maritime and Air Trafficking in N Keywords (Art. 1) Obligation to respect human rights (Art. 1) Jurisdiction of States (Art. 5) Right to liberty and security (Art. 5-3) Brought promptly before judge or other officer (Art. 5-1) Deprivation of liberty (Art. 5-1) Lawful arrest or detention (Art. 35) Admissibility criteria ECLI:CE:ECHR:2010:0329JUD000339403