

International Legal Data in Action

Ideas and Applications for the ICJ and PCIJ

Seán Fobbe

Ludwig-Maximilians-Universität München

16 March 2023

Transforming the Role of International Courts and Tribunals in
a New Era of Adjudication

Fundação Getulio Vargas (FGV) Workshop

Structure of the Presentation

- 1 International Legal Data
- 2 Research Questions and Methods
- 3 Final Thoughts
- 4 Additional Slides

About this presentation

Scope

- Presentation
- Notes
- Further Reading

Available open access in one week

Direct Link www.doi.org/10.5281/zenodo.7738453

Web www.seanfobbe.com

Part I

International Legal Data

International Legal Data

Overview

- Data Science requires data
- Reasonable number of international relations data sets
- Availability of international legal data is very limited
- Very few high-quality legal corpora (= data sets composed of texts) available

International Legal Data

Selected Corpora

- Baturó et al. (2017): UN General Debate Corpus (UNGDC)
<https://doi.org/10.7910/DVN/0TJX8Y>
- Alschner et al. (2018): Text of Trade Agreements (TotA)
<https://github.com/mappingtreaties/tota>
- Schoenfeld et al. (2019): UN Security Council Debate Corpus (UNSCDC)
<https://doi.org/10.7910/DVN/KGVSYH>
- Alschner et al. (2020): Electronic Database of Investment Treaties (EDIT)
<https://github.com/mappingtreaties/tota>

Academic Paper

Fobbe, S. (2022). Introducing Twin Corpora of Decisions for the International Court of Justice (ICJ) and the Permanent Court of International Justice (PCIJ). *Journal of Empirical Legal Studies*, 19(2), 491-524. <https://doi.org/10.1111/jels.12313>

Features

- Complete collections of all ICJ and PCIJ majority and minority opinions from 1922 to 2022
- Public Domain (CC Zero 1.0)
- Open and platform independent file formats (PDF, TXT, CSV)
- Extensive documentation
- Open Source

New ICJ and PCIJ Corpora

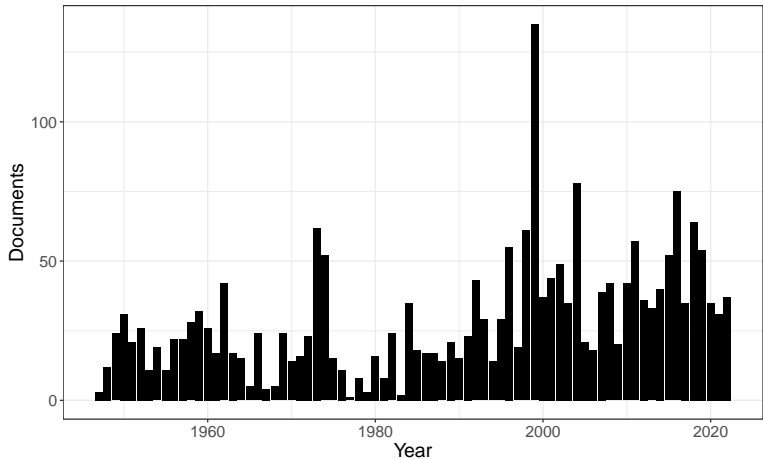
Comparison

Dataset	CD-PCIJ	CD-ICJ
Version	1.1.0	2022-09-07
Coverage	1922–1940	1947–2022
Variables	29	27
Documents (EN)	259	2,215
Documents (FR)	261	2,197
Tokens (EN)	1,296,536	15,421,131
Tokens (FR)	1,262,184	15,775,848
Legal Status	Public Domain	Public Domain

Corpus of Decisions: International Court of Justice

Version 2022-09-07

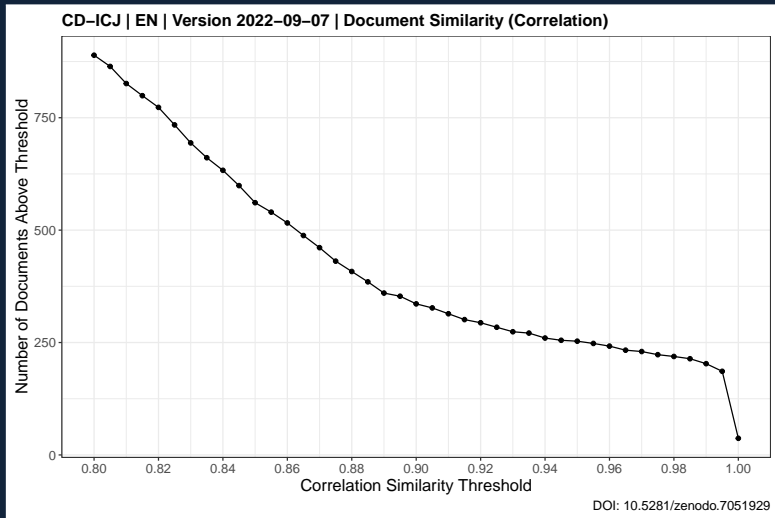
CD-ICJ | EN | Version 2022-09-07 | Documents per Year



DOI: 10.5281/zenodo.7051929

Corpus of Decisions: International Court of Justice

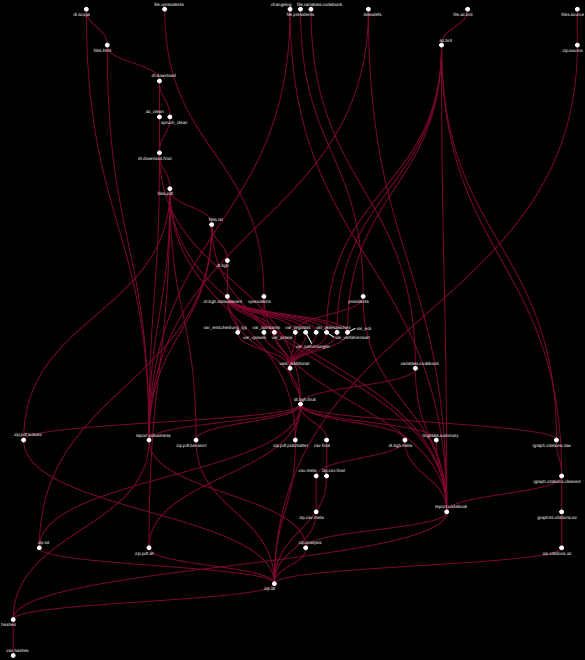
Version 2022-09-07



New ICJ and PCIJ Corpora

Feature Roadmap

- Refactor as strictly modular build in {targets} framework (example workflow on following slide)
- Part-of-Speech Tagging (POS)
- Named Entity Recognition (NER)
- Extract and include citation network as GraphML
- Text Segmentation



New ICJ and PCIJ Corpora

Download

Dataset	Download via Concept DOI
CD-ICJ	https://doi.org/10.5281/zenodo.3826444
CD-PCIJ	https://doi.org/10.5281/zenodo.3840479

Version and Concept DOI

- Concept DOI: Always download the *newest* version
- Version DOI: Always download a *specific* version

Part II

Research Questions and Methods

Doctrinal Analysis

Research Questions

- How do legal concepts evolve in the jurisprudence of the ICJ and PCIJ?
- What textual similarities can we discover in the jurisprudence of the Courts?
- Are certain legal concepts syntactically or semantically linked?
- Can we develop new classifications for legal concepts or factual problems?
- Do legal concepts appear in minority opinions before they enter a majority judgment/advisory opinion?

Recommended Reading

Alschner, Wolfgang, *The Computational Analysis of International Law* (July 29, 2019). Forthcoming in Rossana Deplano and Nicholas Tsagourias (eds), 'Research Methods in International Law: A Handbook', Ottawa Faculty of Law Working Paper No. 2019-33, Available at SSRN: <https://ssrn.com/abstract=3428762>

Citation Analysis

Research Questions

- What counts as a 'citation'?
- What do citations even measure?
- How does the ICJ use citations? How does the PCIJ use citations?
- Has the use of citations changed since 1922?
- Can citation analysis enhance doctrinal research in international law?

Recommended Reading

Alschner, W., & Charlotin, D. (2018). The growing complexity of the International Court of Justice's self-citation network. *European Journal of International Law*, 29(1), 83-112

Research Questions

- Who is the invisible college of international lawyers before the two World Courts? (Schachter 1977)
- Do ICJ/PCIJ judges favor their home State, region, culture, wealth level, political system? (Posner and De Figueiredo 2005)
- Which other persons are of explicit concern to the ICJ/PCIJ? Do they appear in more than one case?
- What role do NGOs play in ICJ opinions?
- What are the gender, origin, socio-economic status or other personal characteristics of actors in ICJ/PCIJ opinions? (Franck et al 2014)

Social Network Analysis II

Recommended Reading

- Schachter, O. (1977). The invisible college of international lawyers. *Northwestern University Law Review*, 72(2), 217.
- Posner, E. A., & De Figueiredo, M. F. (2005). Is the international court of justice biased? *The Journal of Legal Studies*, 34(2), 599-630.
- Franck, S. D. et al (2014). The diversity challenge: exploring the invisible college of international arbitration. *Columbia Journal of Transnational Law*, 53(3), 429.
- Pereira, L. L. S., & Ridi, N. (2021). Mapping the 'invisible college of international lawyers' through obituaries. *Leiden Journal of International Law*, 34(1), 67-91.

Geospatial Analysis

Research Questions

- Which identifiable sub-State geographic areas are the subject of ICJ and PCIJ proceedings? What is their distribution?
- Are there surprising links between specific doctrines and spatial features?
- Which States care about which issues in which areas?
- Do the majority and minority opinions have different geographical interests?
- Are there commonalities or differences in the judicial treatment of land, sea, air, space and cyberspace?

Example

Extract geographic terms from the text of opinions, map them to points/polygons on a projection and visualize on a world map

Part III

Final Thoughts

Final Thoughts

- 1 Replication, replication! Publish your source code and research data!
- 2 Found an error? Need a feature? Submit a GitHub issue or send me an e-mail (fobbe-data@posteo.de)
- 3 Publish anything based on the CD-ICJ or CD-PCIJ? Send me a link, I'll include it in the Codebook as an example!

Questions?

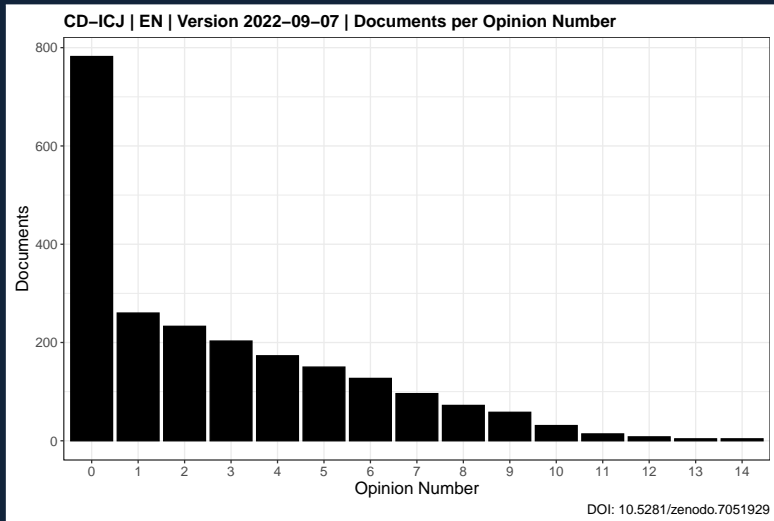
www.seanfobbe.com

Part IV

Additional Slides

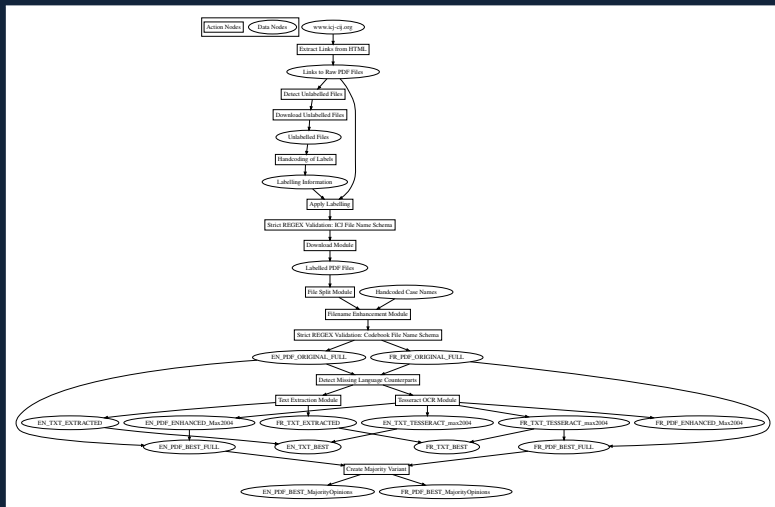
Corpus of Decisions: International Court of Justice

Version 2022-09-07



Corpus of Decisions: International Court of Justice

Workflow | Version 2022-09-07



Corpus of Decisions: International Court of Justice

Workflow | Version 2022-09-07

