A Test Collection for Case law Retrieval

Daniel Locke and Guido Zuccon
Queensland University of Technology
daniel.locke@hdr.qut.edu.au, g.zuccon@qut.edu.au

Aim

We provide a test collection for the evaluation of effectiveness of systems for case law retrieval.

Helping lawyers **find case law** to answer legal questions

Significance

Case law retrieval has no standard test collection for evaluation of effectiveness. Our collection contains **orders of magnitudes more documents**, as well as **nearly 10 times more relevance assessments** per topic than previous topic specific collections.

Collection statistics

Topics

12 topics

E.g.

"Whether the Federal Arbitration Act preempts a state-law contract rule that singles out arbitration by requiring a power of attorney to expressly refer to arbitration agreements before the attorney-in-fact can bind her principal to an arbitration agreement."

Documents

Our collection	Previous largest
3,597,230	63,916

compared to available search systems

Courtlistener	Austlii	Westlaw [3]
4,009,726	841,836	~13,000,000

Assessments per topic

Turtle [2]	Our collection	Locke et al. [1]
200	212	26

Relevance levels

Four levels of graded relevance

- **On point**
- **Explanatory**
- Background

Not relevant

Potential uses

1. **Evaluation of effectiveness**
2. **Citation network effects**
3. **Focused retrieval**

Each decision contains a list of citations. Evaluation of network measures on ranking remains unexplored in the literature.

Relevant documents contain tags of the relevant portion. These tags may be used for focused retrieval, for retrieval at a paragraph level rather than a document level.

Future work

We are **currently extending our collection to 38 topics**. We are also extending the collection by means of making it available to further users to provide relevance assessments.

The collection and interface are available at ***ielab.io/caselaw***.

Further Information

ielab.io/caselaw

[1] Daniel Locke, Guido Zuccon, and Harrisen Scells. 2017. Automatic Query Generation from Legal Texts for Case Law Retrieval. Springer International Publishing, 181–193.
[2] Howard Turtle. 1994. Natural Language vs. Boolean Query Evaluation: A Comparison of Retrieval Performance. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 212–220.
[3] Susan Mart. 2013. The case for curation: the relevance of digest and citator results in Westlaw and Lexis. Legal Reference Services Quarterly 32 (1-2), 13–53.