ESRA: An End-to-End System for Re-identification and Anonymization for Swiss Court Rulings

Joel Niklaus[†], **Matthias Stürmer**[†]

[†]*Research Center for Digital Sustainability (RCDS)*, University of Berne, Switzerland, {joel.niklaus}/{matthias.stuermer}@inf.unibe.ch

Key Contributions

We designed a high-level system architecture for re-identification and anonymization of Swiss court rulings

This system a) enables the verification of existing anonymizations and b) reduces anonymization costs and errors

Why is it Hard?

Court rulings as textual data are inherently unstructured

Research on anonymization on textual data is very thin and even more so on court decisions

Difficult balance between open justice and privacy

Why Should One Care?

Privacy is a fundamental human right

Court rulings should be public to facilitate jurisprudence and make justice system accountable

Manual anonymization is expensive, error-prone and not scalable

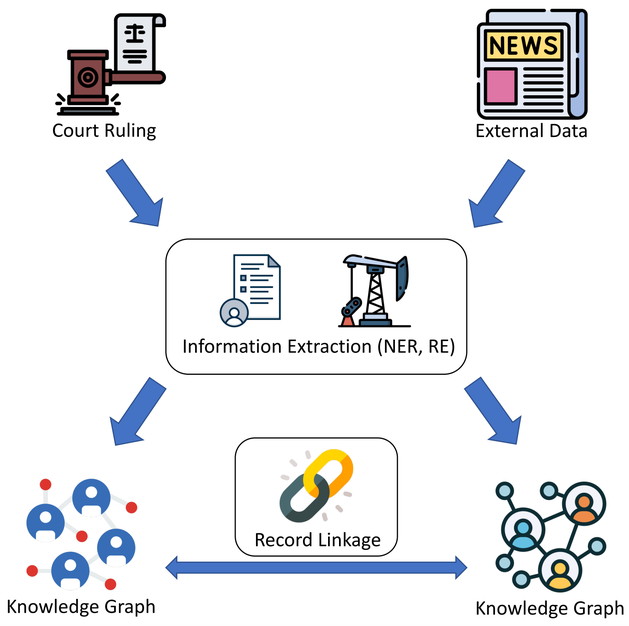
Re-Identification Module

Information Extraction from *court ruling* and *external data* with

- Named Entity Recognition (NER)
- Relation Extraction (RE)

=> Saved in two knowledge graphs

Use record linkage to link anonymised entities from the court ruling to real entities from external data



Data

> 300K Swiss Federal Court Rulings

newspapers, trade registers, government publications, death notices, etc.

Anonymization Module

Named Entity Recognition to find strings to be anonymized

Four methods for anonymization:

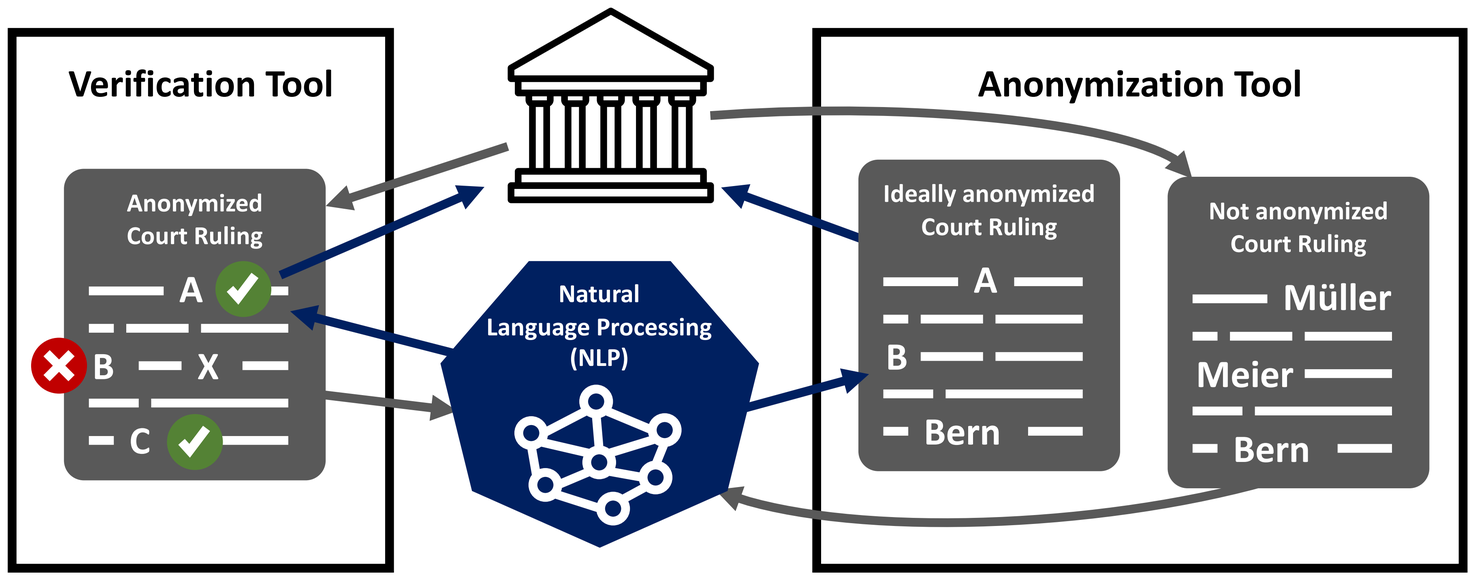
1. Suppression (Uni Bern => XXX)
2. Tagging (Uni Bern => Uni_1) (current practice)
3. Random substitution (Uni Bern => ETH)
4. Generalization (Uni Bern => the Uni)

Methods

BERT Pretraining and Finetuning

Find a Balance

We see the anonymization as an iterative process of using the anonymization module to anonymize an entity and then applying the re-identification module to determine how safe the ruling is from re-identification. These two steps are applied until the safety score is sufficiently high.



Conclusions

We designed a system for re-identification and anonymization of Swiss court rulings

Re-identification module verifies safety of existing anonymizations => promotes privacy

Anonymization module reduces anonymization costs and errors => more comprehensive publication practice