FAU

Bayerisches Staatsministerium der Justiz

Friedrich-Alexander-Universität Erlangen-Nürnberg

Notare Bayern und Pfalz

BUNDESNOTARKAMMER
KÖRPERSCHAFT DES ÖFFENTLICHEN RECHTS

Follow us at direga.fau.de
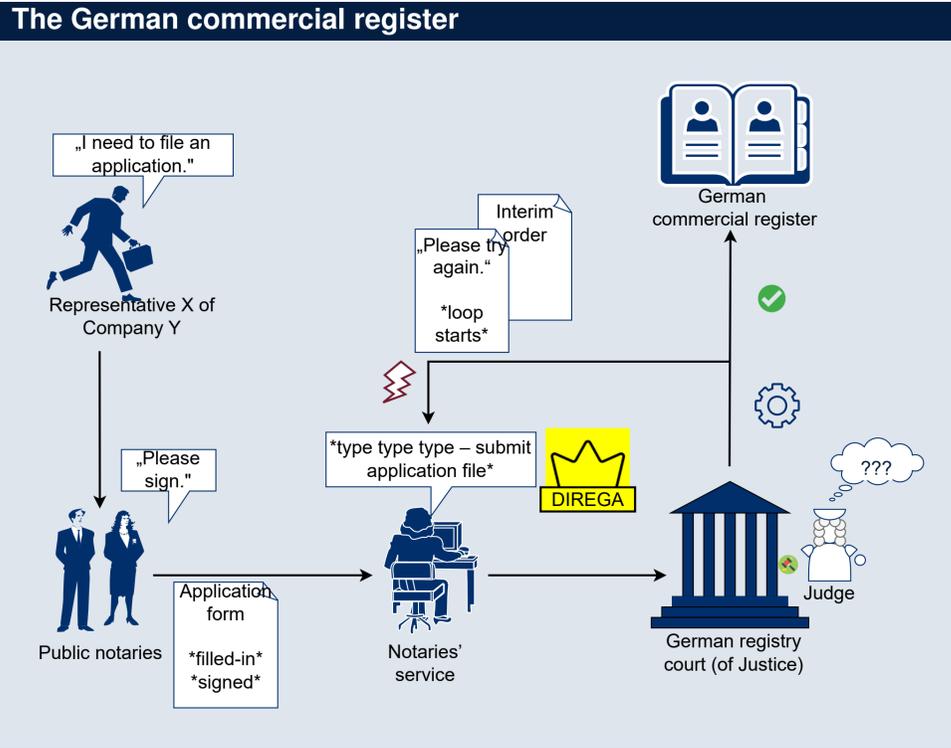
DIREGA – Building Decision Support for German Register Law
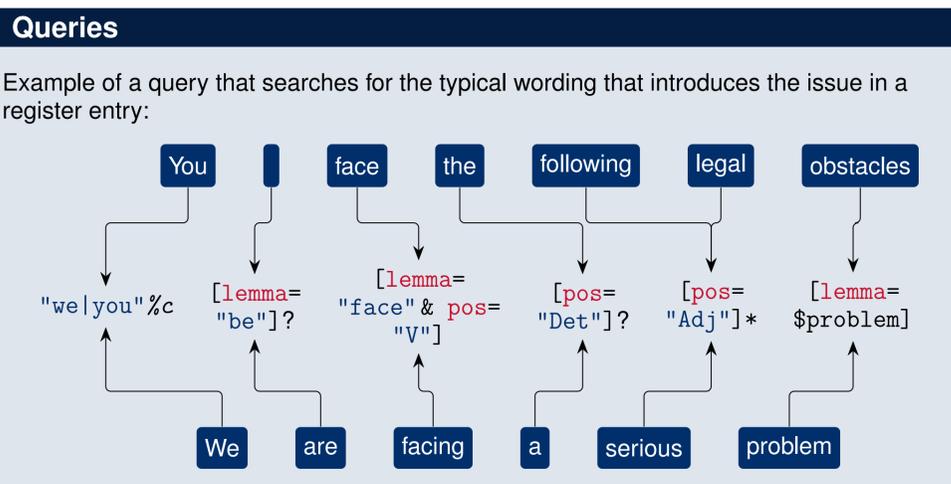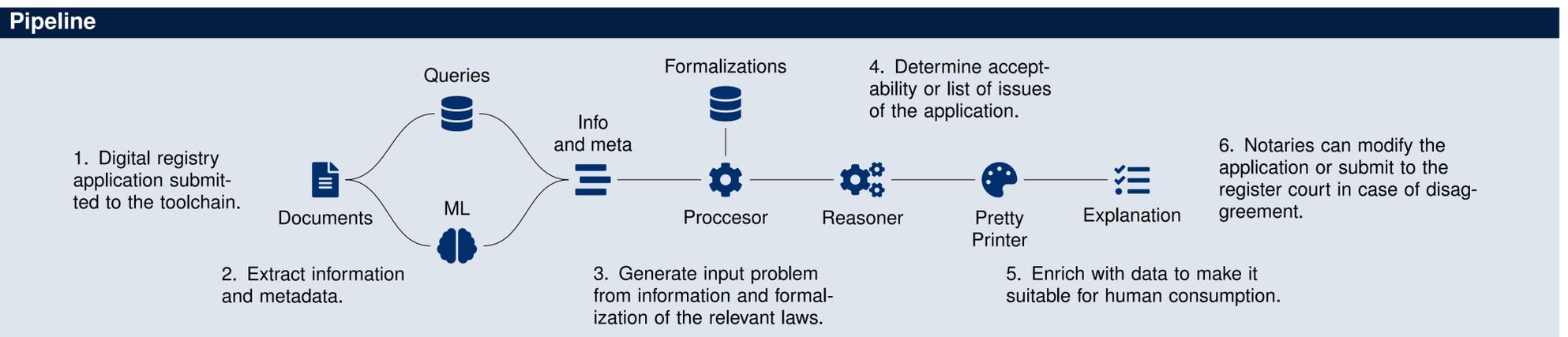
Axel Adrian   Osman Anil Basaran   Nathan Dykes   Stephanie Evert   Michael Gritz   Merlin Humml   Michael Kohlhase   Johannes Lindner   Andreas Maier   Stephan Prettner   Max Rapp   Lutz Schröder   Verena Stürmer

The German commercial register

„I need to file an application."

Representative X of Company Y

„Please sign."

Public notaries

Application form
filled-in
signed

Notaries' service

type type type – submit application file

DIREGA

Interim order

„Please try again."

loop starts

German commercial register

German registry court (of Justice)

Judge

???

Data in the register

1	**Name of the company** … which includes the entity of the company	COMPANY
2	**Business address / administrative address** … which is relevant for judicial jurisdiction of registry courts … which is relevant for filing a suit against the company	
3	**Owners** … who can be liable Associates / non-liable Shareholders … who need to be described with the amount of ownership	
4	**Representatives / Rules of representation** … if not stated differently from the law, default articles apply	
5	**Capital, i.e. liability mass** … which can be a very complex from a legal standpoint	
6	**Statute** … which is relevant is relevant for the legal effectiveness of resolutions of the shareholders	

- 11,917,322 files of various types
- 1,219,814 cases for commercial register entries
- 445,023 commercial register entries
- 30 legal forms
- 96 document types
- 124 case types, e.g. lists of shareholders, annual financial statements

Challenges

Explainability — enables → Scrutability — ensures → Correctness

Pipeline

1. Digital registry application submitted to the toolchain.

Documents

Queries

ML

Info and meta

Formalizations

Proccesor

2. Extract information and metadata.

3. Generate input problem from information and formalization of the relevant laws.

4. Determine acceptability or list of issues of the application.

Reasoner

Pretty Printer

Explanation

5. Enrich with data to make it suitable for human consumption.

6. Notaries can modify the application or submit to the register court in case of disagreement.

Queries

Example of a query that searches for the typical wording that introduces the issue in a register entry:

You | face the following legal obstacles

"we|you"%c   [lemma="be"]?   [lemma="face" & pos="V"]   [pos="Det"]?   [pos="Adj"]*   [lemma=$problem]

We   are   facing   a   serious   problem

This classic, symbolic approach is:
- Predictable
- Lightweight, inexpensive
- Fast
- Good for legal compliance
- High precision generates reliable training data

Logic

```
protocol_correctly_signed(Application, Company) :-
  not uses_default_regulation(Company),
  different_protocol_satisfied(Company).
protocol_correctly_signed(Application, Company) :-
  uses_default_regulation(Company),
  not missing_signatures_executives(Application, Company).
```

Metadata enrichment using machine learning

1. **Document metadata**
   - Structures within the commercial register database
   - Metadata additionally submitted by notaries
   - Interim orders (commercial register responses)
   - Existing documents referring these submitted documents
2. **Extracting information**
   - Structured Information: Pre-existing metadata or structured fields can be accessed directly and reliably
   - Unstructured Information: Data from documents can be extracted after OCR using different approaches:
     - ► Rule-Based Systems: Applying rules to extract specific patterns
     - ► Prompts and LLMs: Using fine-tuned Large Language Models to interpret and extract relevant features
3. **Validation dataset**
   - We manually annotated a new dataset consisting of interim orders related to shareholder lists
   - These include submission failure reasons classified by three legal experts
   - Will be used in the next phase of our research to validate our different approaches