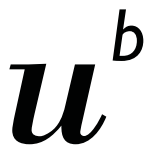
UNIVERSITY OF BERN



Bachelor Thesis

Judgment outcome extraction for Swiss court rulings

Author:
Marco Buchholz
15-118-987

Supervisors:
PD. Dr. Matthias Stürmer
Joël Niklaus

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science

at the

Research Center for Digital Sustainability Faculty of Science
University of Bern

November 2022

ABSTRACT

Apart from the Federal Supreme Court of Switzerland, there do not exist extracted judgment outcomes for Swiss court decisions. Although court decisions abide to a general structure, the underlying pattern and composition of words used to convey the final judgment may vary from court to court. This work will focus on analyzing said structure of judicial cases and explore to what extend a rule-based matching system can correctly identify the judgment outcomes across different courts in Switzerland. We present an overview of the entire pipeline which ranges from scraping the documents to the eventual extraction of the judgment outcome. We introduce an automated extraction system for commonly used patterns across different case documents which greatly reduces the required manual effort and analysis of such. With a coverage of 94.53% extracted judgment outcomes out of 613'629 rulings and a near perfect verification result, we have shown that a rule-based matching system is an excellent method of choice.

ACKNOWLEDGMENTS

I would like to thank my supervisors, PD Dr. Matthias Stürmer and Jöel Niklaus from the Research Center for Digital Sustainability, for their support and guidance during my bachelor thesis. Furthermore I would like to thank Loredana Coppe for always proofreading any English assignments in the past as well as this thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

1 INTRODUCTION

1.1 INFORMATION GATHERING TO BUILD A CASE

With digitalization at its peak, data a across all different fields is made available to the general public. This can especially be useful in the court of law by applying the method of analogical reasoning in order to build a case. On the basis of previous decisions of similar nature, one is able to draw similarities and identify patterns, thus forcing the court to be consistent in its verdicts [8]. It enables for a rational, statistical prediction about the outcome of a trial without any preconceived notions or personal experiences as previous cases should simply be a historical reflection of reality [9]. Entscheidsuche.ch[1] currently offers over hundreds of thousands of Swiss case documents with hundreds of cases being added daily. With data of such volumes and by using conventional methods of analysis, extracting relevant information and drawing conclusions related to a case would take an unreasonable amount of time. Through the help of machine learning and natural language processing information, extraction can be compressed to minutes.

1.2 LEGAL JUDGMENT PREDICTION

Information cost has been categorized as one of the primary transaction costs overall [10]. In the context of law, information about similar cases and their outcomes is a primary method of building a case. Systems which accelerate or even automate said process have been researched and implemented, such as retrieving relevant judgments using everyday vocabulary [12], or providing relevant law articles when given a case [6]. Legal judgment prediction is taking these methods one step further with the intend to predict the judgment outcome for a given case based on its provided facts. The use of its application greatly accelerates the process of analyzing and comparing decisions of similar nature. Such models assist legal practitioners and citizens to make more data- driven decisions, while also substantially reducing legal costs and improving access to justice [3]. For such a model to be as precise as possible, there needs to exist as much accurately labelled data as possible.

1 https://entscheidsuche.ch/

2

RELATED WORK

Regular expressions were used to extract judicial language entities by analyzing the structure of Chinese judicial texts [5]. The project briefly compares the methods of regular expressions, machine learning and GATE (General Architecture for Text Engineering) which offers a GUI and enables visualization of operations, but ultimately decide that regular expressions would be the best approach for their interests.

A paper in which regular expressions were used to extract metadata from scientific papers shows that using regular expressions is an effective and reliable way to not only distinguish between different types of references, such as journal articles, books, conference papers etc., but also for separating the metadata for which constitutes those references [11]. Categorization was very precise with the lowest values in precision being 87% for journal articles and 89% in recall for conference articles. Correctly identifying the constituting metadata was also found to be of high precision as well as recall for fields which usually encompass a strict pattern such as author name, publication date and DOI (Digital Object Identifier). Unstructured fields however, namely the title of the article and the name of the journal showed a recall of 53 and 64 percent. Nevertheless, their paper shows great promise for the success of my project as categorization between different regions and/or courts will most likely occur as well as establishing a pattern for judgement outcomes which encompass a stricter structure than titles and names of articles.

Although regular expressions is a widely used method for a quick and effective way to extract certain information from text, there are limitations that come with it such as extracting meaning from text or the amount of rules becoming unmanageably high when intending to induce grammar. Thus if one has access to large, annotated bodies of text (corpora), one can use machine learning techniques to overcome some of these limitations[7].

3 IMPLEMENTATION

3.1 STRUCTURAL FOUNDATION

In order to understand the process which leads to the extraction of the judgment outcome, the general structure of a Swiss court ruling is briefly introduced. A typical Swiss verdict is comprised of five sections: the header, the facts, the considerations, the ruling and the footer [1, 2]. Note that not all case rulings contain all listed sections below. Some sections may or may not be absent.

1. The header represents the introduction to the respective judgement. It names the court and its composition, as well as the place and date of the decision. It introduces the parties and their respresentatives and may briefly touch upon the matter of dispute, without providing a detailed explanation.
2. The facts serve to present the disputed and undisputed facts, the request as well as the factual assertion of the parties, the history of the case and any taking of evidence.
3. In the considerations, the court makes a legal assessment of the entire events. This constitutes the legal reasoning that later leads to the decision.
4. The ruling contains the decision on the merits and the costs.
5. The footer usually contains some information on the legal remedies. Often times an official footer section is omitted and the contents are added to the end of the ruling section.

3.2 WORKFLOW

The entire project's code and implementation can be found on the repository SwissCourtRulingCorpus[1] by Joël Niklaus. The required steps leading to the extraction of the judgment outcome for a given case document can be described in six steps. The order of tasks usually follow the numeration depicted in the figure 1 below, yet can vary as every step can be executed in isolation.

1 https://github.com/JoelNiklaus/SwissCourtRulingCorpus

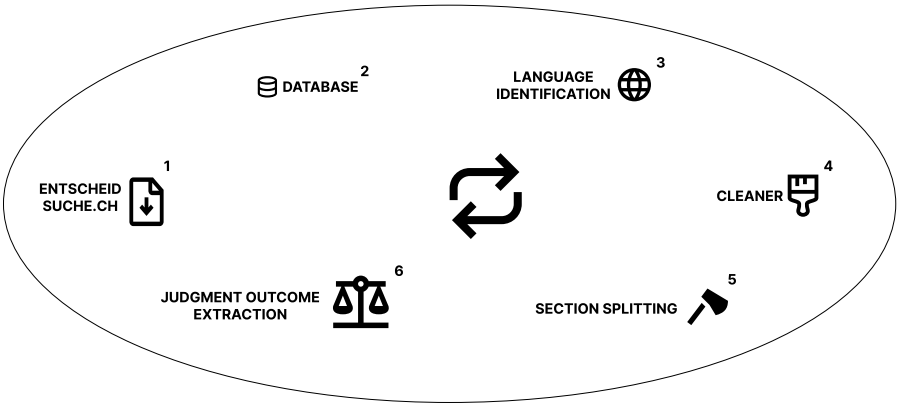


Figure 1: Workflow of SwissCourtRulingCorpus. The numbers depict the typical order of the executed steps.

3.2.1 *Scraping*

Entscheidsuche.ch[2] updates their website daily with all publicly available caseses from the courts in Switzerland. Per default, only case documents which do not already exist locally are considered in order to avoid rewriting the entire database. In a first step, all the files are scraped including the associated metadata of every court's folder from entscheidsuche.ch/docs[3]. A combination of BeautifulSoup's parsing library and Requests[4] is used to extract the complete list of URLs to each of the case documents. The list is then iterated to download all the files.

3.2.2 *Database*

For case documents associated with courts and chambers which do not already exist within our database, the required tables are created. We then extract the textual and meta information from every downloaded file and save them to our database. We again use BeautifulSoup's[5] parsing library to extract the textual contents of HTMLs. For PDFs we use the tika-python[6] library to extract the documents content.

3.2.3 *Language identification*

Since within a Swiss court case documents can vary in the language they are written in, every case must have assigned an appropriate language identifier which will be needed for the section splitting and

2 https://entscheidsuche.ch/
3 https://entscheidsuche.ch/docs
4 https://requests.readthedocs.io/en/latest/
5 https://pypi.org/project/beautifulsoup4/
6 https://pypi.org/project/tika/

judgment outcome extraction task. For every document within our database for which we were unable to extract the corresponding language from the metadata, we make use of the fastText language identification tool[7] to assign to each the appropriate language id.

3.2.4 *Cleaner*

Certain courts may include strange patterns or unnecessary text within their case documents which can make the entire process more prone to errors. To improve further processing, cleaning functions for certain courts have been implemented removing such patterns and texts. An example can be seen in the Listing 1 below.

Listing 1: Cleaning regexes for SH_OG

```
"SH_OG": [
  {
    "pattern": "\\n\\d+\\n",
    "replacement": "",
    "description": "remove page numbers"
  },
  {
    "pattern": "\\n{date.year}\\n",
    "replacement": "",
    "description": "remove year"
  },
  {
    "pattern": "^Microsoft Word.+\\.docx?\\n",
    "replacement": "",
    "description": "remove 'Microsoft Word' with file name at
        the beginning."
  }
]
```

3.2.5 *Section splitting*

The section splitting task serves the purpose of splitting the case ruling in its predefined sections introduced in 3.1. It is done so by defining a set of regex indicators, each of which implicate the start of one of the sections. The composition of words and special characters to indicate certain sections however, may differ for each court in Switzerland. Thus for every Swiss court, a different set of regex indicators has to be defined for each section which may appear within a ruling.

Given a case document, it is first broken down into its paragraphs. Using the set of regex indicators we iterate through each paragraph

7 https://pypi.org/project/fasttext-langdetect/

and search for a match. Once a match has been found, all subsequent paragraphs are appended to the specific section until we find a new match. We continue this process until we reach the end of the file.

3.2.6 *Section splitting helper module*

The main issue facing the section splitting task lies in the wide variety of words used to indicate the start of each section and the different variations and additives these words may encompass. Not only may courts amongst themselves use different variations but this can also occur within the same court. Because of this, a manual extraction of indicators for a single court may require hours of structural analysis of different case rulings within that court. In order to significantly reduce manual analysis while addressing maintainability for future courts and cases, a python module has been implemented which presents common indicators for a specific court for each section. It works by counting the occurrence of every paragraph appearing across all case rulings for a given court. If a court has certain patterns of indicating different sections in certain ways, those will be reflected by having a significantly higher occurrence count than a random paragraph appearing within a case ruling. Using this logic we can then filter out low occurrence paragraphs and make rough assignment choices by having a baseline of predetermined common indicators for each section. The result is an accurate overview of all appearing patterns of indicators for each of the sections.

Table 1: Top four entries for facts section of Verwaltungsgericht Bern

total count	indicator	coverage (%)
7157	Sachverhalt:	95.46
94	Sachverhalt und Erwägungen:	1.25
27	in Sachen B._	0.36
10	in Sachen	0.13

Table 2: Top four entries for consideration section of Verwaltungsgericht Bern

total count	indicator	coverage (%)
7160	Erwägungen:	95.51
125	Der Einzelrichter zieht in Erwägung:	1.67
94	Sachverhalt und Erwägungen:_	1.25
56	Der Einzelrichter zieht in Erwägung,	0.75

Table 3: Top four entries for ruling section of Verwaltungsgericht Bern

total count	indicator	coverage (%)
4865	Demnach entscheidet das Verwaltungsgericht:	64.89
2141	Demnach entscheidet der Einzelrichter:	28.56
441	Demnach entscheidet die Einzelrichterin:_	5.88
44	Demnach entscheidet das Schiedsgericht:	0.59

The tables represent the top four entries of the output of the pattern extractor module for the court Verwaltungsgericht Bern. Since for most courts there do not exist clear indicators for the footer section, it has been omitted for this court as well. By using all of the extracted indicators for each represented section, we are able to achieve a coverage nearing 100% without the need for manual analysis.

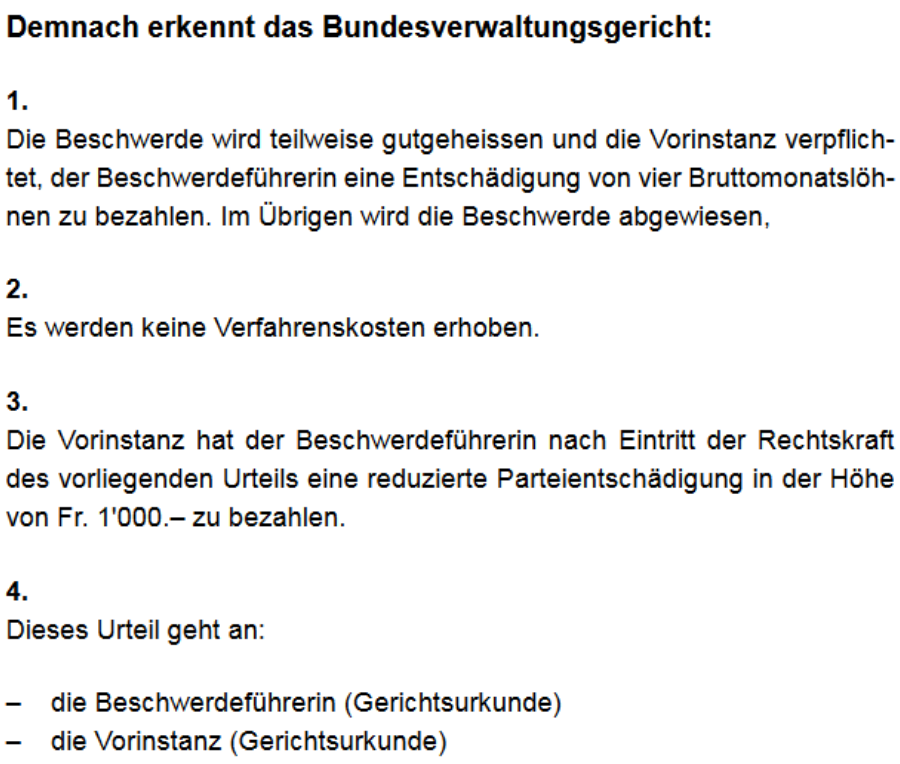
3.2.7 *Judgment outcome extraction*

The judgment outcome extraction step serves the purpose of extracting the judgment outcome based on a set of predefined indicators. We have defined seven possible outcomes so far:

1. Approval: The request is accepted in full
2. Partial approval: Part of the request is accepted
3. Dismissal: The request is dismissed in its entirety
4. Partial dismissal: Part of the request is rejected
5. Inadmissible: The court does not have jurisdiction over the request; Formal defects within the complaint
6. Write off: No reason for the proceedings (the decision is no longer needed, for example, because the parties have reached an out-of-court settlement or because two proceedings have been combined).
7. Unification: When two cases are about the same thing, they will get merged into one.

In order to map a ruling to one of the defined judgment outcomes, a set or combination of words is defined for each one of them. Since the meaning and implication of the words stay consistent across different courts, the same indicators can be used for all of them. As the indicators for the different judgment outcomes are not exclusive to this context, it is critical to consider only the ruling section of a case to avoid false positives as much as possible. A successful judgment

outcome extraction therefore goes hand in hand with a precise section splitting. For courts which abide to a numbered ruling structure, depicted in figure 2, we can restrict the context in which a judgment outcome is searched for to a minimum by only considering the final decisions while disregarding any additional text and information which may appear in the ruling section. However, for courts with an unnumbered ruling structure depicted in figure 3, we must consider the whole text which increases the likelihood of a false positive as there may exist additional text such as an intro or an outro which we cannot ignore.


Demnach erkennt das Bundesverwaltungsgericht:

1.
Die Beschwerde wird teilweise gutgeheissen und die Vorinstanz verpflichtet, der Beschwerdeführerin eine Entschädigung von vier Bruttomonatslöhnen zu bezahlen. Im Übrigen wird die Beschwerde abgewiesen,

2.
Es werden keine Verfahrenskosten erhoben.

3.
Die Vorinstanz hat der Beschwerdeführerin nach Eintritt der Rechtskraft des vorliegenden Urteils eine reduzierte Parteientschädigung in der Höhe von Fr. 1'000.– zu bezahlen.

4.
Dieses Urteil geht an:

– die Beschwerdeführerin (Gerichtsurkunde)
– die Vorinstanz (Gerichtsurkunde)

Figure 2: Sample case document from court CH_BVGer of a numbered ruling structure

3.2.8 *Judgment outcome helper module*

A helper method similar to the section splitting can be used for the extraction of indicators in the judgment outcome task. Instead of looking at whole paragraphs however, the ruling section of each case will be split into different set of n-grams and compared to each other. Patterns which tend to remain consistent across many cases within a court will have a significantly higher occurrence count. These consistencies will primarily be the indicators responsible for the judgment outcome. Using this method, the manual effort of extracting patterns and indicators for judgment outcomes within the ruling sections of a

Swiss court is reduced significantly and thus maintainability is simplified.

Demgemäss erkennt das Einzelgericht:

://: Die Beschwerde wird abgewiesen.

Der Beschwerdeführer trägt die Kosten des Verfahrens mit einer Gebühr von CHF 500.–.

Mitteilung an:

– Beschwerdeführer

– Staatsanwaltschaft Basel-Stadt

Figure 3: Sample case document from court BS_Omni of an unnumbered ruling structure

Table 4: Top six entries for a combination of three neighboring words for the cantonal court of Vaud.

total count	indicator
22145	('Le', 'recours', 'est')
12292	('recours', 'est', 'rejeté.')
4238	('recours', 'est', 'admis.')
1971	('recours', 'est', 'irrecevable.')
1493	('recours', 'est', 'partiellement')
1422	('est', 'partiellement', 'admis.')

Table 5: Top five entries for a combination of four neighboring words for the cantonal court of Vaud.

total count	indicator
12279	('Le', 'recours', 'est', 'rejeté.')
4237	('Le', 'recours', 'est', 'admis.')
1968	('Le', 'recours', 'est', 'irrecevable.')
1487	('Le', 'recours', 'est', 'partiellement')
1369	('recours', 'est', 'partiellement', 'admis.')
1072	('dans', 'la', 'mesure', 'où')

4 EXPERIMENTAL RESULTS

4.1 RESULT ANALYSIS

The evaluation of the results is done by analysing the coverage of both the section splitting and judgment outcome extraction task and the verification of correctness. The complete results with the exact numbers can be found in the appendix A. A summary and visualization in form of a bar chart can be found in the figure 4 below.

4.1.1 *Ruling section*

We count the number of total case documents for each court and compare them to the amount of ruling sections we were able to extract. Courts with very low coverages usually include many case documents which either have the ruling section included within the considerations or do not include a ruling section at all. Since the judgment outcome extraction task is heavily reliant on the accuracy of the ruling section, case documents and entire courts with merged considerations and ruling sections have been ignored in order to reduce false positives. This issue, however, only consistently appeared for smaller courts, namely SH_OG, TG_OG, UR_Gerichte, AI_Aktuell, AI_Bericht and OW_Gerichte ,which combined amount to 4'056 case documents in total.
Altogether we were able to extract the ruling section of 613'629 out of 681'360 scraped case documents. An overall coverage of 88.88%.

4.1.2 *Judgment outcome extraction*

Analogous to the evaluation of the ruling section, the number of total judgments we were able to extract is compared to the number of extracted ruling sections for each court. Due to the nature of a pattern matching system, there may exist some cases in which the chosen indicators for certain courts wrongly indicate the start of a ruling section. In such cases there may exists a substantial amount of wrongly identified text. However, since the structure of a ruling section does not need to follow a strict guideline there may also exist cases in which the ruling section consists of unnecessary text which is not directly relevant to the judgment outcome. The more unnecessary text exists, the higher the likeliness of a false positive. We can address both issues of wrongly identified text as well as ruling sections which consist of too much text by setting a character limit of 3000 characters.

If the ruling section exceeds said limit we dismiss the case document as we assume it to be faulty or too ambiguous in order for us to be able to apply the pattern matching extraction method.
In total we were able to extract a judgment outcome for 580'042 out of 613'629 ruling sections. An overall coverage of 94.53%.

4.1.3 *Verification of results*

Although a broad coverage is beneficial for both the section splitting and judgment outcome extraction task, it is more important that the extracted rulings and outcomes are correct. For the two biggest courts we chose a verification sample of 100 and for the remaining courts a total of 50 random case documents. The issue discussed in chapter 3.2.7 on sub optimal precondition for the judgment outcome extraction task due to a unnumbered ruling structure is reflected within the verification results of BS_Omni, ZH_Baurekurs and GE_Gerichte. All three of these courts abide to a either mixed numbered and unnumbered structure or a purely unnumbered structure for their case documents. With a ratio of 45/50 for GE_Gerichte, 46/50 for BS_Omni and 47/50 for ZH_Baurekurs the verification result of these courts are some of the few outliers which did not obtain a near perfect validation of 49/50 and 50/50 correct case documents.
In order to determine the validity of the verification, the Clopper Pearson Exact method [4] is used approximate a confidence interval. For a verification of 100/100 correct case documents applying this method results in a confidence interval of:

$$CI_{95\%}(0.96378, 1.0000)$$

Meaning for the largest court CH_BGer totaling 164'402 extracted judgment outcomes, our error margin would lie between:

$$[0, 5'955]$$

Since the Clopper Pearson exact method is merely an estimation, we can assume that for courts with very strict numbered ruling structures under which CH_BGer falls as well, the real amount of falsely extracted ruling sections and judgment outcomes lies within the lower section of the error margin.

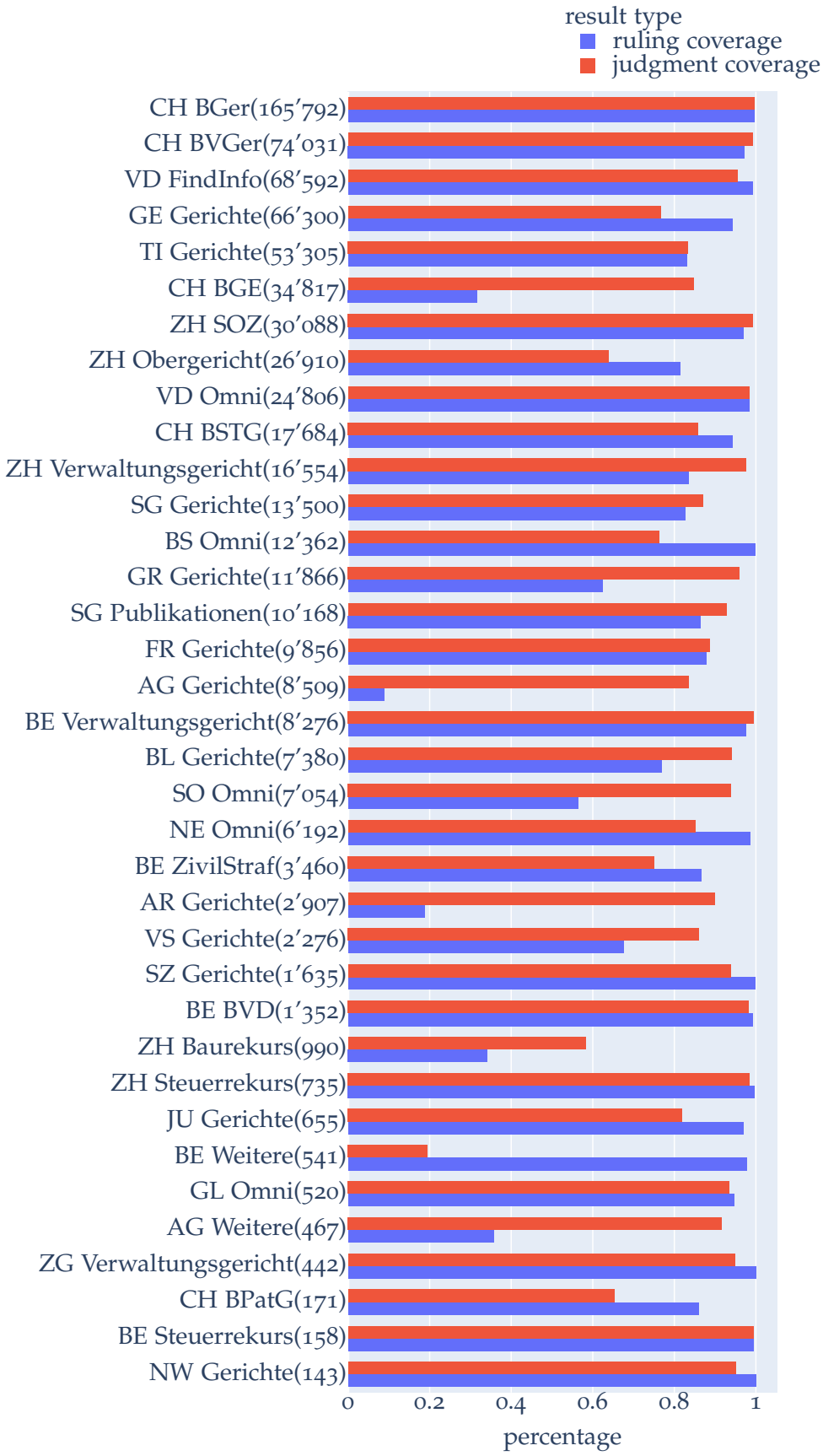


Figure 4: Coverage results ordered by total case documents per court

5 CONCLUSION

In this thesis we extracted the ruling sections and judgment outcomes from as many Swiss case documents as possible through a rule based pattern matching system. By defining different sets of regex indicators we were able to recognize and assign each of the predefined sections to the appropriate segments of a case documents as well as identify the correct judgment outcome of a majority of the total amount of scraped case documents.

We presented a maintainable system for extracting the patterns which indicate different outcomes and beginning of sections. We introduced the basic pipeline of the SwissCourtRulingCorpus which leads to the eventual extraction of the judgment outcome. In total we were able to extract 613'629 rulings sections out of 681'360 case documents and 580'042 judgment outcomes out of the 613'629 ruling sections.

6 OUTLOOK

The work of this thesis builds the basis for the judgment prediction task. Naturally, the future direction involves using the results of this work to train NLP models to make prediction about the outcome of a case based on its facts. Using the most accurate and largest courts such as CH_BGer and CH_BVGer and then continuing to increase the dataset by adding additional courts which achieved a good verification result as the correctness of the results is an important requirement so that the judgment prediction task may perform as accurately as possible. The methods and patterns for courts with relatively high error margins can be reevaluated and improved upon.

A

APPENDIX TEST

In this chapter the exact results of the section splitting and judgment outcome extraction task can be found.

Table 6: Ruling section extraction

court	extracted rulings	total cases
AG_Gerichte	755	8'509
AG_Weitere	167	467
AR_Gerichte	545	2907
BE_BVD	1341	1'352
BE_Steuerrekurs	157	158
BE_Verwaltungsgericht	8076	8'276
BE_Weitere	529	541
BE_ZivilStraf	2'996	3'460
BL_Gerichte	5'672	7'380
BS_Omni	12'344	12'362
CH_BGE	11'030	34'817
CH_BGer	165'091	165'792
CH_BPatG	147	171
CH_BSTG	16'656	17'684
CH_BVGer	71'994	74'031
FR_Gerichte	8'657	9'856
GE_Gerichte	62'474	66'300
GL_Omni	492	520
GR_Gerichte	7'398	11'866
JU_Gerichte	635	655
NE_Omni	6'103	6'192
NW_Gerichte	143	143
SG_Gerichte	11'150	13'500
SG_Publikationen	8'793	10'168
SO_Omni	3'975	7'054
SZ_Gerichte	1'631	1'635
TI_Gerichte	44'282	53'305

Continued on next page

Table 6 – *Continued from previous page*

court	extracted rulings	total cases
VD_FindInfo	68'056	68'592
VD_Omni	24'391	24'806
VS_Gerichte	1'537	2'276
ZG_Verwaltungsgericht	442	442
ZH_Baurekurs	338	990
ZH_Obergericht	21'917	26'910
ZH_Sozialversicherungsgericht	29'160	30'088
ZH_Steuerrekurs	733	735
ZH_Verwaltungsgericht	13'822	16'554

Table 7: Judgment outcome extraction

court	extracted judgments	verification
AG_Gerichte	630	50/50
AG_Weitere	153	50/50
AR_Gerichte	490	49/50
BE_BVD	1'318	50/50
BE_Steuerrekurs	156	50/50
BE_Verwaltungsgericht	8'036	50/50
BE_Weitere	103	50/50
BE_ZivilStraf	2'250	49/50
BL_Gerichte	5'331	50/50
BS_Omni	9'413	46/50
CH_BGE	9'344	50/50
CH_BGer	164'402	100/100
CH_BPatG	96	47/50
CH_BSTG	14'296	50/50
CH_BVGer	71'390	100/100
FR_Gerichte	7'672	49/50
GE_Gerichte	47'876	45/50
GL_Omni	460	50/50
GR_Gerichte	7'099	48/50
JU_Gerichte	520	49/50
NE_Omni	5'194	50/50

Continued on next page

Table 7 – *Continued from previous page*

court	extracted judgments	verification
NW_Gerichte	136	50/50
SG_Gerichte	9'697	50/50
SG_Publikationen	8'163	50/50
SO_Omni	3'732	50/50
SZ_Gerichte	1'530	49/50
TI_Gerichte	36'903	46/50
VD_FindInfo	65'005	50/50
VD_Omni	23'984	50/50
VS_Gerichte	1'323	50/50
ZG_Verwaltungsgericht	419	50/50
ZH_Baurekurs	197	47/50
ZH_Obergericht	14'000	50/50
ZH_Sozialversicherungsgericht	28'958	50/50
ZH_Steuerrekurs	722	50/50
ZH_Verwaltungsgericht	13'498	50/50

BIBLIOGRAPHY

[1] Luzerner Moot Court 2015. *VADEMECUM ZUM VERFASSEN EINES U RTEILS*. 2015. URL: https://www.unilu.ch/fileadmin/fakultaeten/rf/furrer/Luzerner_Moot_Court/Luzerner_Moot_Court_2015_Allg/03_Vademecum_Urteil_2015.pdf (visited on 10/18/2022).

[2] *Aufbau eines Gerichtsentscheides*. URL: https://www.zivilprozess.ch/gerichtsentscheide/aufbau-eines-gerichtsentscheides (visited on 10/18/2022).

[3] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. *Neural Legal Judgment Prediction in English*. 2019. DOI: 10.48550/ARXIV.1906.02059. URL: https://arxiv.org/abs/1906.02059.

[4] Stephanie Glen. *Clopper-Pearson Exact Method*. URL: https://www.statisticshowto.com/clopper-pearson-exact-method/ (visited on 11/11/2022).

[5] JIAO Kainan. "Extraction Method of Judicial Language Entities Based On Regular Expression." In: Apr. 2021. URL: http://scholarship.law.ufl.edu/flr/vol67/iss4/1.

[6] Chao-Lin Liu and Ting-Ming Liao. "Web Technologies Research and Development." In: 2005, pp. 64–75. URL: https://link.springer.com/book/10.1007/b106936.

[7] Prakash Nadkarni, Lucila Ohno-Machado, and Wendy Chapman. *Natural language processing: an introduction*. Sept. 2011. URL: https://academic.oup.com/jamia/article/18/5/544/829676.

[8] Emily Sherwin. "A Defense of Analogical Reasoning in Law." In: 66 (1999). URL: https://chicagounbound.uchicago.edu/uclrev/vol66/iss4/3.

[9] Dru Stevenson and Nicholas J. Wagoner. "Bargaining in the Shadow of Big Data." In: 2015, pp. 1345 –1346. URL: https://scholarship.law.ufl.edu/flr/vol67/iss4/1/.

[10] Dru Stevenson and Nicholas J. Wagoner. "Bargaining in the Shadow of Big Data." In: 2015, p. 1375. URL: https://scholarship.law.ufl.edu/flr/vol67/iss4/1/.

[11] Xiaoyu Tang, Qingtian Zeng, Tingting Cui, and Zeze Wu. "Regular expression-based reference metadata extraction from the web." In: Oct. 2010. URL: https://ieeexplore.ieee.org/abstract/document/5607427.

[12] Yi-Hung Liu Yen-Liang Chen and Wu-Liang Ho. "text mining approach to assist the general public in the retrieval of legal documents." In: (Jan. 2013). URL: https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.22767.

DECLARATION

Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.
Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.

Schwyz, November 2022

Marco Buchholz