Daniel Braun, Elena Scepankova, Patrick Holl, Florian Matthes

The Potential of Customer-Centered LegalTech

Consumer Protection in the Digital Era

New technologies and tools, often summarised under the term "LegalTech", are changing the way in which legal professionals work. The digital transformation has changed many aspects of our daily life and democratised access to knowledge and services. In the legal domain, however, consumers rarely benefit from digitisation. On the contrary, they are often overpowered by big corporations and their well-equipped legal departments. In this paper, we outline how LegalTech can be used to empower consumers in the digital era, by building tools to support consumers and those who protect them. In order to show the potential of customer-centred LegalTech, we present two prototypes which semantically analyse, assess, and summarise Terms of Services from German web shops.

Daniel Braun

Wissenschaftlicher Mitarbeiter am Lehrstuhl für Software Engineering betrieblicher Informationssysteme der Technischen Universität München

E-Mail: daniel.braun@tum.de

Elena Scepankova

Wissenschaftliche Mitarbeiterin am Lehrstuhl für Software Engineering betrieblicher Informationssysteme der Technischen Universität München.

E-Mail: elena.scepankova@tum.de

Patrick Holl

Wissenschaftlicher Mitarbeiter am Lehrstuhl für Software Engineering betrieblicher Informationssysteme der Technischen Universität München

E-Mail: patrick.holl@tum.de

Prof. Dr. Florian Matthes

Inhaber des Lehrstuhls für Software Engineering betrieblicher Informationssysteme der Technischen Universität München

E-Mail: matthes@tum.de

1 Introduction

The digital revolution has democratised many aspects of our lives. Access to knowledge is no longer restricted to those who can afford 32 volumes of Encyclopædia Britannica or have access to university libraries, instead it is available to everyone with access to the internet. The access to the fine arts, but also to once expensive services like translation, has been opened up to new classes of citizens by digitisation.

For a long time, the legal domain was arguably one of the biggest resisters to digitisation efforts and in some aspects still struggles to catch up with other industries. A fact painfully displayed by the case of the "special electronic attorney mailbox" ("besonderes elektronisches Anwaltspostfach", beA) [1]. Nowadays, digitisation has entered the legal profession as so-called "LegalTech", a portmanteau word consisting of "legal services" and "technology", widely used as description for the support or automation of legal processes with software or online services.

However, unlike in other areas, mostly big companies and law firms benefit from these developments. Almost all of the existing LegalTech tools, like Lexis Advance, rfrnz, Juristische Textanalyse, and Lawlift, to name just a few, are made for companies and law firms, rather than consumers. Therefore, LegalTech tools are not only missing the opportunity to democratise the access to legal advice by making it more affordable and available, they are actively supporting the current imbalance of power, existing be-

tween companies and consumers[1], by providing companies with even more advantages over customers.

Currently, there are only a few examples of LegalTech tools, like Flightright or Chevalier, which are built for the benefits of consumers. And even these tools are still built to serve the commercial interests of their operators. In this paper, we want to advocate the idea of customer-centred LegalTech tools and present two prototypes which semantically analyse, assess, and summarise Terms of Services from German web shops.

2 Significance of Terms of Services

Standard form contracts trace back to the 19th century. In the age of industrialisation, entering into contracts has been accompanied by the unilateral use of pre-formulated rules tailored to one party's own interests and thus resulting in an imbalance of powers between the contracting parties. [2] Today, customers are confronted with standard form contracts every day in form of Terms of Services (ToS), for example when they buy something online or register for an online service. Studies with more than 45,000 participants have shown that only 0.1% to 0.2% of customers read ToS of online shops. [3]

While standard form contracts regularly reflect an imbalance of contracting power, this imbalance is even stronger in situations where one of the contracting parties is a consumer without a professional legal background and the other one is a company with a potentially huge legal department. The relevance of this is visible in the amount of jurisprudence in this area, comprising more than 28,000 judgements in Germany only. [4]

In acknowledgement of these facts, the European lawmaker has limited the creative leeway for companies, when it comes to standard form consumer contracts.[2] One might ask oneself why consumers should care about unlawful clauses in ToS, because in case of a dispute, these clauses are void anyway. In reality, however, at least for online shopping, the amount in dispute is often so low that consumers avoid legal steps, even if they are in the right.

3 Related Work

A goal similar to the one outlined in this paper is pursued by the project «Terms of Service; Didn't Read» (ToS;DR). [5] Instead of automatically assessing and evaluating the ToS, ToS;DR is crowdsourced and provides manually generated summarisations of ToS from many major websites. However, the fact that ToS;DR is crowd-sourced affects their scalability and topicality.

An automated approach to analyse online standard form contracts was presented by Lippi et al. Their analysis focuses solely on so-called «unfair clauses» which are forbidden under the law of the European Union. [6] In their experiment, they analysed Terms of Services from 20 major websites regarding eight unfair clauses. In a leave-one-document-out evaluation, they achieved a precision of 0.62 using a Support Vector Machine. In contrast to our approach, Lippi et al. try to do a binary classification (unfair clause exists / does not exist), while we try to gather additional information and summarise them. In order to create these summaries, a system first has to obtain the relevant information from the text. Information Retrieval (IR) for legal texts has gained a lot of attention in recent years. Examples are McCallum [7], Grabmaier et al. [8], Francesconi et al. [9], and Shulayeva et al. [10], or, for German texts, Walter and Pinkal [11], and Watlt et al. [12]. The issue of simplifying legal texts was addressed by Bhatia et al. [13], Collantes et al. [14] and others.

4 Possible Approaches

We identified two possible approaches to tackle the imbalance of power between consumers and web shop operators:

- **Directly support consumers**: By automatically finding, assessing, and summarising ToS with regard to lawfulness and customer-friendliness, we can empower consumers to make educated decisions about where to buy or not. We first presented this idea in [15].
- **Support consumer protection agencies**: Instead of directly supporting consumers, we can also support those organisations that protect consumers by providing them with tools to automatically analyse large amounts of ToS. Unlike consumer themselves, such organisations are often willing to take on legal battles with companies about their ToS and can therefore support the enforcement of existing customer protection laws.

The consumer protection law aims to protect the consumer and allow for «optimal market decisions» [16]. The aim to enable consumers to take «optimal market decisions» is considered as fundamental and is intended to foster the faith of the consumers into market processes. [17, p. 19] German consumer law uses different regulatory instruments to achieve this, by preventively designing consumer protection law and by subsequently declaring contractual clauses not in accordance with those legal provisions as «void». In order to work effectively, legal regulations comprise both market-complementary as well as market-compensatory instruments [18, p. 47].

With the first approach, we likewise aim to enable the consumer to take «optimal market decisions». We consider both above-mentioned ways as important, but we think that there are obstacles which hinder legal regulations to achieve full value. The most important fact is that reading and understanding legal contracts is hindered by the fact that in comparison to regular language, legal language is characterised by a high degree of abstractness. In order to fulfil its function as a merit instance for any socially relevant behaviour, law itself needs to guard its capability of abstractly reacting even to unforeseen situations, which results in formulations characterised by a low level of comprehensibility. By summarising contractual terms and translating their legal and linguistic complexity into a simplified language (summary generation), we provide the customer with the possibility to understand his rights and duties and to take decisions based on knowledge, not on – justified or unjustified – trust towards the shop provider.

Secondly, standard form contracts like Terms of Services address fundamental conditions of performance of a certain business. In the context of online shopping, they set the provisions for e.g. payment, delivery, revocation or liability. Due to the abstract character of the legal language stipulating contractual rights and duties on the one hand, and normative requirements in legal regulations and judicial decisions on the other hand, the consum-

---

1 Cf. German Constitutional Court BVerfGE 89, 214 of 19 October 1993.
2 Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, OJ L 95/29 of 21 April 1993.

er often finds himself not capable of understanding and assessing the validity or invalidity of their contract. With our second functionality, we thus automatically (a) identify the differences between the clauses of (online shops as) companies and (b) assess their (un)lawfulness.

While this approach directly supports consumers, it has one shortcoming: It does not help to make ToS fairer and only supports those who use the proposed tool. In order to get rid of illegal ToS clauses for good, legal actions are necessary in order to force companies to change their ToS. In Germany, the "Verbraucherzentralen" (customer protection agencies) are important actors when it comes to the legal enforcement of consumer interests. They regularly admonish web shop providers for illegal ToS clauses. Therefore, we teamed up with experts from the Verbraucherzentralen to develop them in their daily work. Unlike consumers, they are interested in checking large amounts of web shops in a short time and also in re-checking them after a while. Moreover, they often are interested in more complicated legal issues than consumers. A tool targeted at this group of professional users should take this into account.

5 Technology

As both approaches have different requirements regarding the depth of the legal analysis that needs to be performed, we decided to also implement them using different technologies. However, they still have very much in common: We use a pipes and filters architecture for both prototypes and the first steps of the pipeline are identical for both prototypes.

5.1 ToS Page Classification

First, we use a naive Bayes classifier to find the ToS page of a web shop by classifying each linked sub-page as "ToS" or "Other". The classifier was trained with a set of 400 manually annotated pages from web shops, 200 of them ToS pages and 200 of them other pages. While the classifier performs very well (cf. Section 7.1 for an evaluation), it is not very fast. In order to be able to present results to consumers as fast as possible, we decided to adopt a hybrid approach and developed a rule-based URL classifier, to pre-select sub-pages that potentially are ToS page and hence restrict the set of pages that have to be classified by the naive Bayes classifier.

The classification is realised using a rule-based approach that matches common patterns for ToS links. One common pattern we identified is that the URL often contains "*AGB*". The classifier separates URL strings into the following components: scheme specifier, network location part, path, query parameters. The path and query parameters are matched against a set of pre-defined, weighted rules. If a candidate reaches a certain threshold, we consider that a given URL points to a potential ToS page.

5.2 Content Extraction

The page which was classified as ToS is then further processed to extract the actual content from the website by e.g. removing headers, navigation etc. The current prototypes use the open source Java-library boilerpipe for this task.

5.3 Information Extraction

After this step, the pipelines for both prototypes differ. For the consumer-oriented prototype, we use simple POS-tagging, for the professional-oriented prototype we use the neural network from the StanfordCore NLP library to build a dependency tree for each sentence of the ToS. Afterwards, for both prototypes, the annotated ToS are analysed by a rule-based information extraction tool, which extracts the important information from the text and stores it in a JSON-format.

The information extraction is a two-step process. In the first step, we identify for each sentence whether it contains a topic of interest, like information about the return policy or limitations of the form of termination. For this step, we mainly use a keyword search which works on a stemmed version of the ToS. Due to the highly regulated nature of legal language this rather simple approach is still very effective. For example, in order to legally restrict the possibility to send back goods in Germany, the term "Widerruf" has to be used.

Once the relevant sentences are identified and labelled with their topics, the second step is to actually extract the information contained in the sentence. In order to do this, a set of rules is stored for each topic. Currently, these rules have to implemented in the source code, in the future we would like to provide users with a mean to create rules during run-time, e.g. with a graphical interface or a domain specific language.

For the consumer prototype, we use regular expressions for information extraction. Table 1 show examples of different extractions rules, translated from German.

Table 1 | Extraction rules from the consumer prototype		
	unlawful	**rules (translated from German)**
Right of Warranty	New goods: less than 2 years; used goods: less than 1 year	-warranty … ([0-9]* | [one | two | …]) [day(s) | month(s) | year(s)] AND used OR NOT used (goods | products) -warranty … used (goods | products) … excluded
Right of Withdrawal	Products have to be send back using the original packaging	- original (packaging | packed | …) … (return | send back)
Withdrawal period	Period of less than 14 days for shops trading in the EU	- withdraw … ([0-9]* | [one | two | …]) [day(s) | month(s) | year(s)]
Obligation to inspect	Warranty rights only if customer inspects and / or reports any product defects	- warranty … [inspect|report] AND NOT merchant
Risk of loss	In case of shipped sale the customer bears the risk of loss	[risk of loss | bearing the risk] … [shipped | carriage of goods] … consumer

For the professional prototype, we first generate a dependency tree for each relevant sentence. The nuances that come with more complex legal issues are often difficult to tackle with regular expressions. The limitation of the form of termination is a good example of this. Very often, one can find in the same sentence forms of termination being listed as permissible and others being ruled out. The sentence „Die Kündigung muss schriftlich oder telefonisch erfolgen, eine Kündigung per E-Mail ist nicht möglich." (The termination must be in written form or through phone, a termination via email is not possible.) specifies that terminations

have to be written or made through phone and at the same time rules out terminations via email. With regular expressions, it is difficult to write rules to cover all possible permutations of such a sentence, especially in German, and figure out which parts are included in the negation.

Dependency trees, such as the one shown in Figure 1, make the dependencies between words explicit and can therefore help to analyse which part of a sentence is negated and many other things. Instead of writing rules which just analyse the words, with dependency trees, we can also analyse the structure of the complete sentence and hence conduct more fine-grained analyses in order to extract structured information from the text. [19]

Figure 1 | Dependency Tree for the Sentence "Jede Kündigung muss schriftlich erfolgen." (Any termination must be in writing.)
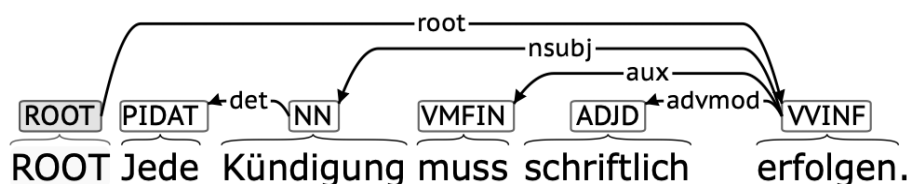
5.4 Assessment

The structured representation of the information contained in the ToS is used for the legal assessment. This assessment is based on a knowledge-base which contains information about legal regulations, like the right to return or warranty rights. The database contains for example the information that, in the European Union, customers have the right to return a product they bought online for at least 14 days.[3] For the consumer version, we use three labels: „legal", „illegal", and „customer friendly". For the given example of return policies, a time span of fewer than 14 days would be classified illegal, 14 days as legal, and everything above as customer friendly. For the professional prototype, we just use the labels „legal" and „illegal". Currently, the knowledge-base is stored in a simple JSON-format and has to be directly managed as a file. In the future, we would like to build a graphical user interface which allows domain experts to maintain and extend the knowledge-base more easily.

5.5 Summarisation

In order to summarise the extracted facts in a simplified language, we currently use a simple template-based approach. Examples of the generated summarisations can be found in Figures 3 and 5. In the future, we would like to make the text generation more flexible by replacing the templates with a surface realiser like SimpleNLG [20, 21].

---

[3] While there are different exceptions from this rule, e.g. for individualised goods, the current prototype does not take into account these special cases.

6 Prototypes

We developed two prototypes in order to show the potential of LegalTech applications to empower consumers and those who protect them. The first prototype, dubbed "consumer prototype" (cf. Section 6.1), was developed in order to serve people without any specialist legal knowledge while shopping online. The second prototype ("professional prototype", cf. Section 6.2) was built with experts from the customer protection agencies in mind as users who want to analyse large amounts of ToS at once. Both prototypes are implemented as web applications, so users do not need to locally install any software. Both backends were implemented using Java and a REST-API to communicate with the frontend.

6.1 Consumer Prototype

In the consumer prototype, the URL of a web shop can be entered through an input field. After clicking on the „Find ToS" button (cf. Figure 2), the

Figure 2 | Input mask of the consumer prototype

backend classifies all outgoing links of the input page with the classifier described in Section 5.1. A breadth-first search is conducted until a ToS page was identified. As an intermediate result, the URL of the ToS page is shown next to the button.

Once the URL is identified, the content extraction, information extraction, assessment and summarization are conducted. The results of all these steps are presented as shown in Figure 3. The results are split into two topics: revocation & right to return ("Widerruf & Rückgaberecht") and warranty („Gewährleistung"). On the left side of the table, the original texts are shown which are excerpted from the ToS. Colour-highlighting is used to explain based on which parts of the text the assessment algorithm made its decision. Green highlights show identified time limits and blue highlights show identified topics. In the middle of the table, the assessment can be found. A „thumbs up" means a clause was identified to be customer-friendly, i.e. exceeds the legal minimum in a way which is beneficial for consumers. A „thumbs down" means a clause was identified as illegal and a „neutral position" means that a clause fulfils the legal minimum. On the right side, the automatically generated summarisation is shown.

While the presented prototype is currently just a proof of concept, running such a service „in production", i.e. giving consumers direct access to it, does have legal implications. First, liability is a question, but second, there is the German Act of Out-of-Court Legal Services which stipulates that in individual cases legal advice against payment shall only be provided by legal person-

Figure 3 | Consumer prototype (mock ToS have been used as input data)

Analysis

Widerruf & Rückgaberecht

Kunden können von uns erhaltene Ware ohne Angabe von Gründen innerhalb von 30 Tagen durch Rücksendung der Ware zurückgeben .

👍 Dieser Shop gewährt ein Widerrufsrecht von 30 Tagen.

Gewährleistung

Gewährleistungsfristen Die Gewährleistung für neueWare beträgt 24 Monate .

✊ Dieser Shop gewährt Verbrauchern eine Gewährleistungsfrist von 24 Monaten für neue Waren.

Für gebrauchte Waren beträgt die Gewährleistungsfrist 10 Monate .

👎 Der Betreiber gewährt Verbrauchern eine Gewährleistungsfrist von 10 Monaten für gebrauchte Waren.

tiple URLs, separated by commas. In addition, it is possible to upload a PDF file or directly paste plain text into the input box.

As a first step, all inputs are converted to plain text. In the case of a URL this means that the same steps are performed as in the consumer prototype: identification of the ToS page and content extraction. In the case of a PDF file as input, the text is extracted using Apache PDFBox. Afterwards, information extraction, assessment, and summarization are performed and the results are displayed as shown in Figure 5.

Other than the consumer prototype, which is optimised to show the analysis of one ToS page, the professional prototype is optimised to show

nel[4], e.g. lawyers, legal counsels, and tax consultants. One could argue that our functionality serves as a mere clarification tool operated on standard form contracts in a general way. As we neither intend nor are capable of providing specific legal advice in individual cases, our tool focuses on enhancing the understanding of legal language. However, these are open questions which go beyond technical feasibility, which have to be solved in order to allow consumers to benefit from LegalTech.

Moreover, it needs to be clearly communicated to potential users which aspects of the ToS are covered by the analysis and that the tool just rates these specific aspects and does not make any statements about other aspects of the ToS, nor does it provide an assessment of the ToS as a whole.

6.1 Professional Prototype

The prototype which we developed to support domain experts from the customer protection agencies has a different interface which is tailored towards their own needs. As shown in Figure 4, the interface has three possible input types. On the top, there is an input field where the user can enter a URL. In contrast to the consumer prototype, it is also possible to enter mul-

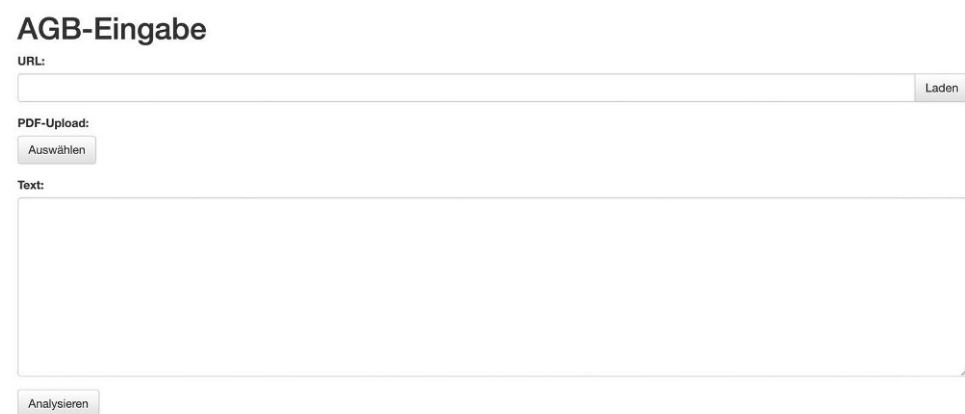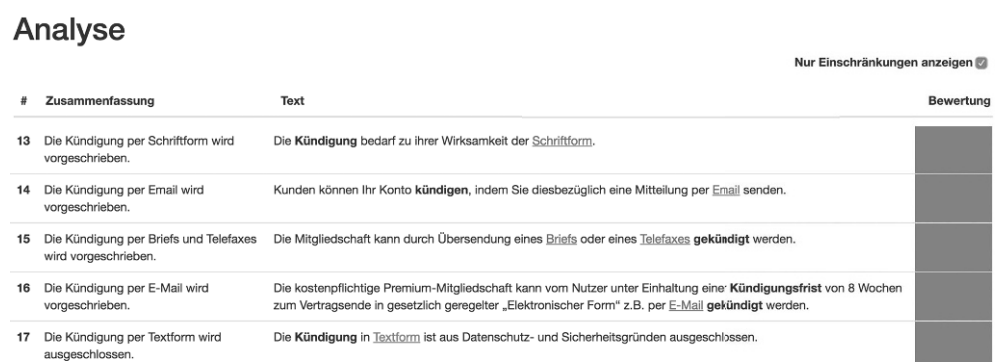**Figure 4 | Input mask of the professional prototype**

AGB-Eingabe

URL:

[   ] [Laden]

PDF-Upload:

[Auswählen]

Text:

[   ]

[Analysieren]

Figure 5 | Professional prototype (mock ToS have been used as input data)

Analyse

Nur Einschränkungen anzeigen ☑

#	Zusammenfassung	Text	Bewertung
13	Die Kündigung per Schriftform wird vorgeschrieben.	Die **Kündigung** bedarf zu ihrer Wirksamkeit der Schriftform.	
14	Die Kündigung per Email wird vorgeschrieben.	Kunden können Ihr Konto **kündigen**, indem Sie diesbezüglich eine Mitteilung per Email senden.	
15	Die Kündigung per Briefs und Telefaxes wird vorgeschrieben.	Die Mitgliedschaft kann durch Übersendung eines Briefs oder eines Telefaxes **gekündigt** werden.	
16	Die Kündigung per E-Mail wird vorgeschrieben.	Die kostenpflichtige Premium-Mitgliedschaft kann vom Nutzer unter Einhaltung eine **Kündigungsfrist** von 8 Wochen zum Vertragsende in gesetzlich geregelter „Elektronischer Form" z.B. per E-Mail **gekündigt** werden.	
17	Die Kündigung per Textform wird ausgeschlossen.	Die **Kündigung** in Textform ist aus Datenschutz- und Sicherheitsgründen ausgeschlossen.	

large amounts of data. It therefore includes a filtering function which allows to only show clauses which were identified as illegal („Nur Einschränkungen anzeigen"). Other than that it contains the same three elements: the original text with highlighting to make the reasoning transparent (this time in the middle), the assessment on the right, which is only binary in this prototype (red = illegal, green = legal), and a short summary on the left.

---

4 Cf. § 8 Abs. 1 Act of Out-of-Court Legal Services.

7 Evaluation

Since the professional prototype uses more advanced technology, the evaluation will focus on this prototype. Moreover, since all further analyses are based on the correct classification of ToS pages, we conducted a separate evaluation for the ToS page classification.

7.1 ToS Page Classification

As mentioned in Section 5.1, the prototypes use a hybrid approach of rule-based and machine learning classification for ToS pages. We collected a dataset of 3424 pages from web shops. 2592 ToS pages, manually labelled by a price comparison website, and 832 other web shop pages. We split the dataset into training (200 ToS and 200 Other) and test (2392 ToS and 632 Other) data.

The results of the evaluation are shown in Table 2. It is visible that the ML approach performed significantly better. Given the relatively small dataset which was used to train the classifier, these results are very promising. The fifth column in Table 2 shows the average time in seconds needed to classify a URL. If successful, the rule-based approach is, as expected, much faster.

Table 2 | Evaluation ToS Classification

approach	precision	recall	F-score	Øt in s
ML	0.9115	0.8219	0.8644	1.435
rule-based	0.7953	0.5393	0.6428	0.001

7.2 Classification of Limitations of the Termination Form

In order to evaluate the classification of limitations of the termination form, which is used by the professional prototype, we used a hand-picked sample of 25 ToS pages from online subscription services, like fitness platforms. In each ToS page, sentences which contain limitations of the termination form were manually labelled with either legal or illegal. In total, the 25 ToS contained 23 sentences with limitations of the termination form of which six contained illegal limitations.

Our algorithm detected all 23 sentences correctly that contained limitations (true positives) but falsely identified five more sentences (false positives). On this test data set, our algorithm achieved therefore a recall of 1.0, a precision of 0.81 and an F-score of 0.9 when it comes to the detection of sentences which contain limitations of the termination form. When it comes to classifying whether these limitations are illegal or not, the algorithm correctly labelled all six illegal limitations (true positives). One legal limitation was falsely classified as illegal (false positive). This means a recall of 1.0, a precision of 0.85, and an F-score of 0.92.

While the data set might be too small to draw general conclusions, the results are very promising and confirm the assumption that dependency trees might be a viable technology for the problem at hand. The algorithm was designed in a way which is optimised towards a higher recall which is also reflected by the results of the evaluation. Since the system is designed as a mean of support for human experts, false positives can be quickly identified by the humans in the loop. False negatives, on the other side, may never be noticed because the amount of data is too large to manually check every sentence for limitations of the termination form.

In the future, it would be desirable to investigate how experts and consumers react to the respective prototypes, in order to find out whether they could really have a real-world impact.

8 Vulnerabilities

The evaluation above measures the performance of the proposed approach in a controlled environment. However, assuming one of the described systems would be successfully applied in the real world, there are other aspects which would have to be taken into account. Most importantly, companies could try to cheat the system by optimising their ToS to disguise their consumer unfriendly clauses from the algorithm. A similar situation can be seen when it comes to optimising websites for search engine algorithms. [22] This is a possible vulnerability of the consumer prototype. For the professional prototype, we expect organisations to not share their set of rules by which they evaluate ToS. This makes it difficult for companies to find out based on which formulation an assessment was made.

In general, there is no easy solution to fix this vulnerability. In the area of search engine optimisation, we have seen an arms race between search engines and dubious website providers for years. Consequently, it would also be necessary in the case of ToS assessment to regularly update the rules to detect the latest fraud attempts of unfaithful shop providers. This also emphasises the importance of customer protection agencies, which could help to hold such providers legally accountable for their deception.

9 Conclusion

In this article, we present two approaches to consumer protection in the digital era by empowering consumers and those who protect them with LegalTech. We chose ToS as the subject of our research because consumers are confronted with them multiple times every day. Beyond presenting two prototypical implementations, which semantically analyse and assess ToS from German web shops, our goal is to raise awareness for the fact that, as of today, mostly companies benefit from the digitisation in the legal industry. This cements the already existing imbalance of power between consumer and companies. Moreover, we want to spark a discussion about the legal issues which have to be resolved, before consumers can truly benefit from LegalTech applications (namely the questions of liability and out-of-court legal services).

Literatur

[1] Möllers, Frederik; Hessel, Stefan: Das Sicherheitsgutachten zum besonderen elektronischen Anwaltspostfach (beA). Computer und Recht, 34(7):413–417, 2018.

[2] Zerres, Thomas: Principles of the German Law on Standard Terms of Contract. Jurawelt, 2014.

[3] Bakos, Yannis; Marotta-Wurgler, Florencia; Trossen, David R: Does anyone read the fine print? Consumer attention to standard-form contracts. The Journal of Legal Studies 43(1):1–35, 2014.

[4] JURIS Rechtsinformationsdatenbank. https://www.juris.de/r3/search, 2017. Last accessed: 2017-10-29.

[5] Binns, Reuben; Matthews, David: Community structure for efficient information flow in'ToS; DR', a social machine for parsing legalese. In: Proceedings of the 23rd International Conference on World Wide Web. ACM, pp. 881–884, 2014.

[6] Lippi, Marco; Palka, Przemyslaw; Contissa, Giuseppe; Lagioia, Francesca; Micklitz, Hans-Wolfgang; Panagis, Yannis; Sartor, Giovanni; Torroni, Paolo: Automated Detection of Unfair Clauses in Online Consumer Contracts. In: JURIX. pp. 145–154, 2017.

[7] McCallum, Andrew: Information extraction: Distilling structured data from unstructured text. Queue, 3(9):48–57, 2005.

[8] Grabmair, Matthias; Ashley, Kevin D; Chen, Ran; Sureshkumar, Preethi; Wang, Chen; Nyberg, Eric; Walker, Vern R: Introducing LUIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools. In: Proceedings of the 15th International Conference on Artificial Intelligence and Law. ACM, pp. 69–78,.2015.

[9] Francesconi, Enrico; Montemagni, Simonetta; Peters, Wim; Tiscornia, Daniela: Semantic processing of legal texts: Where the language of law meets the law of language, volume-6036. Springer, 2010.

[10] Shulayeva, Olga; Siddharthan, Advaith; Wyner, Adam: Recognizing cited facts and principles in legal judgements. Artificial Intelligence and Law, 25(1):107–126, 2017.

[11] Walter, Stephan; Pinkal, Manfred: Automatic extraction of definitions from German court decisions. In: Proceedings of the workshop on information extraction beyond the document. Association for Computational Linguistics, pp. 20–28, 2006.

[12] Waltl, B.; Landthaler, J.; Scepankova, E.; Matthes, F.; Geiger, T.; Stocker, C.; Schneider,C.: Automated extraction of semantic information from german legal documents. In: IRIS: Internationales Rechtsinformatik Symposium. 2017.

[13] Bhatia, Vijay K: Simplification v. easification-the case of legal texts. Applied linguistics,4:42, 1983.

[14] Collantes, Miguel; Hipe, Maureen; Sorilla, Juan Lorenzo; Tolentino, Laurenz; Samson,Briane: Simpatico: A Text Simplification System for Senate and House Bills. In: Proceedings of the 11th National Natural Language Processing Research Symposium, pp. 26–32, 2015.

[15] Braun, Daniel; Scepankova, Elena; Holl, Patrick; Matthes, Florian: SaToS: Assessing and Summarising Terms of Services from German Webshops. In: Proceedings of the 10th International Conference on Natural Language Generation. Association for Computational Linguistics, Santiago de Compostela, Spain, pp. 223–227, 2017.

[16] Oehler, Andreas: Zur ganzheitlichen Konzeption des Verbraucherschutzes – eine ökonomische Perspektive. Verbraucher und Recht, 21(8):294–300, 2006.

[17] Tamm, Marina: Verbraucherschutzrecht: Europäisierung und Materialisierung des deutschen Zivilrechts und die Herausbildung eines Verbraucherschutzprinzips. Mohr Siebeck, 2011.

[18] Reichardt, Vanessa: Der Verbraucher und seine variable Rolle im Wirtschaftsverkehr. Duncker & Humblot, 2008.

[19] Braun, Daniel; Scepankova, Elena; Holl, Patrick; Matthes, Florian: Customer-centered LegalTech: Automated Analysis of Standard Form Contracts. In: Tagungsband Internationales Rechtsinformatik Symposium (IRIS) 2018. Editions Weblaw, pp. 627–634, 2018.

[20] Gatt, Albert; Reiter, Ehud: SimpleNLG: A realisation engine for practical applications.In: Proceedings of the 12th European Workshop on Natural Language Generation (ENLG2009). pp. 90–93, 2009.

[21] Bollmann, Marcel: Adapting simplenlg to german. In: Proceedings of the 13th European Workshop on Natural Language Generation. pp. 133–138, 2011.

[22] Malaga, Ross A: Worst practices in search engine optimization. Communications of the ACM, 51(12):147–150, 2008.