Legal Knowledge and Information Systems
A. Wyner and G. Casini (Eds.)

© 2017 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/978-1-61499-838-9-145

Automated Detection of Unfair Clauses in Online Consumer Contracts

Marco LIPPI [a,1], Przemyslaw PALKA [b], Giuseppe CONTISSA [c], Francesca LAGIOIA [b], Hans-Wolfgang MICKLITZ [b], Yannis PANAGIS [d], Giovanni SARTOR [c], and Paolo TORRONI [e]

[a] *DISMI – Università di Modena e Reggio Emilia, Italy*
[b] *Law Department, European University Institute, Florence, Italy*
[c] *CIRSFID, Alma Mater – Università di Bologna, Italy*
[d] *iCourts, University of Copenhagen, Denmark*
[e] *DISI, Alma Mater – Università di Bologna, Italy*

Abstract. Consumer contracts too often present clauses that are potentially unfair to the subscriber. We present an experimental study where machine learning is employed to automatically detect such potentially unfair clauses in online contracts. Results show that the proposed system could provide a valuable tool for lawyers and consumers alike.

Keywords. Unfair terms detection, Consumer contract, Machine learning

1. Introduction

A PhD student from Poland plans to move to Italy. She will have to open a bank account, rent a flat, get a local phone number, etc. She will have to sign many lengthy contracts. Most of them will be only written in Italian. Can she simply focus on the costs and features of services described in the contracts? Or will she have to worry about possible 'legal traps' as well?

It is a fact that consumers rarely read the contracts they are required to accept [19], and even if they do, they have no means to influence their content. This created a need for limitations on contractual freedom [13], not only to protect consumer interests, but also to enhance the consumers' trust in transnational transactions and improve the common market [18]. The same considerations apply to online platforms, a necessary component of Junker Commission's Digital Single Market initiative.[2] Because consumers cannot realistically be expected to read and fully understand all the contracts they sign, European consumer law aims to prevent businesses from using so-called 'unfair contractual terms' in the contracts they unilaterally draft and require consumers to accept [20]. Law regarding such terms applies also to the Terms of Service (ToS) of online platforms [12]. Unfor-

[1]Corresponding Author: `marco.lippi@unimore.it`.
[2]Brussels, 6.5.2015COM(2015) 192 final. Communication: A Digital Single Market Strategy for Europe.

tunately, it turns out that owners of these platforms, such as Google, Facebook and Twitter, do use in their ToS unfair contractual clauses, in spite of the European law, and regardless of consumer protection organizations, which have the competence, but not the resources, to fight against such unlawful practices.

We propose to address this problem by partially automating the detection of (potentially) unfair clauses using machine learning. This paper follows and combines results of our earlier work. That includes an analysis of the legal issues involved in the automation of enforcement of consumer law regarding unfair contractual clauses, and have developed a software that detects unfair clauses, based on manually created rules encoding recurring textual structures, which gave promising results [15]. However, such an approach has a drawback, in that it is labor-intensive and struggles to cope with the diversity and rapid evolution of the language of ToS. In other recent work we trained a machine learning classifier on a corpus annotated by domain experts, and successfully used it to extract claims from legal documents [11]. Here we build on the work done so far by applying machine learning methods to the detection of unfair contract clauses.

We have structured this paper as follows. In Section 2 we introduce the legal problem. Section 3 describes the corpus and the document annotation procedure. Section 4 explains the machine learning methodology employed in the system, whereas Section 5 presents experimental results. Section 6 discusses related work and concludes with a look to future research.

2. Problem Description

In this section we briefly introduce the European consumer law on unfair contractual terms (clauses). We explain what an unfair contractual term is, present the legal mechanisms created to prevent business from employing them, and describe how our project will contribute to these mechanisms.

According to art. 3 of the Directive 93/13 on Unfair Terms in Consumer Contracts, a contractual term is unfair if: 1) it has not been individually negotiated; and 2) contrary to the requirement of good faith, it causes a significant imbalance in the parties rights and obligations, to the detriment of the consumer. This general definition is specified in the Annex to the Directive, containing an indicative and non-exhaustive list of the terms which may be regarded as unfair, as well in by more than 50 judgments of the Court of Justice of the EU [14]. Examples of unfair clauses encompass taking jurisdiction away from the consumer, limiting liability for damages on health and/or gross negligence, imposing obligatory arbitration in a country different from consumers residence etc.

Loos and Luzak [12] identified five categories of *potentially unfair* clauses: 1) establishing jurisdiction for disputes in a country different than consumers residence; 2) choice of a foreign law governing the contract; 3) limitation of liability; 4) the provider's right to unilaterally terminate the contract/access to the service; and 5) the provider's right to unilaterally modify the contract/the service itself. Our research identified three additional categories: 6) requiring a consumer to undertake arbitration before the court proceedings can commence; 7) the provider retaining the right to unilaterally remove consumer content from the service; 8) having a consumer accept the agreement simply by using the service, not only without reading it, but even without having to click on “I agree/I accept.”

The 93/13 Directive creates two mechanisms to prevent the use of unfair contractual terms: *individual* and *abstract* control of fairness. The former takes place when a consumer goes to court: if a court finds that a clauses is unfair (which it can do on its own motion), it will consider that the clause is not binding on the consumer (art. 6). However, most consumers do not take their disputes to courts. That is why abstract fairness control has been created. In each EU Member State, consumer protection organizations have the competence to initiate legal proceedings aiming to obtain the declaration that clauses in consumer contracts are unfair, through judicial or in administrative proceedings. The national implementations of abstract control may differ—public authorities or civil society organizations may be involved; there may or may not be fines for using unfair contractual terms; etc. [21]—but what is common to all member states is that if a business uses unfair terms in their contracts, in principle there is always someone competent to make them stop.

Unfortunately, the legal mechanism for enforcing the prohibition of unfair contract terms have been unable to effectively counter this practice so far. As reported by some literature [12], and as our own research indicates [15], unfair contractual terms are, as of today, widely used in ToS of online platforms.

In our previous research [15] we developed a theoretical model of tasks that human lawyers currently need to carry out before legal proceedings concerning the abstract control of fairness of clauses can begin. Those include: 1) finding and choosing the documents; 2) mining the documents for potentially unfair clauses; 3) conducting the actual legal assessment of fairness; 4) drafting the case files and beginning the proceedings. Our project aims to automate the second step, enabling a senior lawyer to focus only on clauses that are found by a machine learning classifier to be potentially unfair, thus saving significant time and labor. Our classifiers will look not only for clearly unfair clauses but also for *potentially* unfair ones. The focus on potentially unfair clauses is due to two main reasons.

First, we may be uncertain on whether a certain type of clause falls under the abstract legislative definition of an “unfair contractual term”. One can only have legal certainty that a certain type of clause is unfair if a competent institution, such as the European Court of Justice, has decided so. That is the case for certain kinds of clauses, such as a jurisdiction clause indicating a country different from the consumer’s residence, or limitation of liability for gross negligence [15]. In other cases the unfairness of a clause, has to be argued for, showing that it creates an unacceptable imbalance in the parties’ rights and obligations. A consumer protection body might want to take the case to a court in order to authoritatively establish the unfairness of that clause, but a legal argument for that needs to be created, and the clause may eventually turn out to be judged fair.

Second, we may remain uncertain on the unfairness of a particular clause detected by the classifier, since its unfairness may depend not only on its textual content, but also on the context in which the clause is to be applied. For instance, a mutual right to unilaterally terminate the contract might be fair in some cases, and unfair in others, for example if unilateral termination would entail losing some digital content (purchased apps, email address, etc.) on the side of the consumer.

3. Corpus Annotation

In order to train machine learning classifiers we produced a corpus consisting of 20 relevant on-line consumer contracts, i.e. the ToS of the following on-line platforms: 9gag.com, Academia.edu, Amazon, eBay, Dropbox, Facebook, Google, Linden Lab, Microsoft, Netflix, Rovio, Snapchat, Spotify, Supercell, Twitter, Vimeo, World of Warcraft, Yahoo, YouTube and Zynga. When more than one version of the same contract was available, we selected the most recent version available on-line for the European customers. The corpus contains overall 5,103 sentences, 333 of which we marked as expressing (potentially) unfair clauses. If a clause span included multiple sentences, we decided to tag all such sentences. We used XML as a mark-up language.

An initial analysis of our corpus enabled us to identify 8 different types of clause, for which we defined 8 corresponding XML tags: jurisdiction (`<j>`), choice of law (`<law>`), limitation of liability (`<ltd>`), unilateral change (`<ch>`), unilateral termination (`<ter>`), arbitration (`<a>`), contract by using (`<use>`), and content removal (`<cr>`). We assumed that for each type of clause we could distinguish three classes: (a) clearly fair, (b) potentially unfair, and (c) clearly unfair. In order to mark the different degrees of (un)fairness we appended a numeric value to each XML tag, with 1, 2, and 3, meaning clearly fair, potentially unfair, and clearly unfair respectively. For instance, the tag `<j3>` indicates that the tagged clause is classified as a clearly unfair jurisdiction clause.

A **jurisdiction** clause specifies what courts will adjudicate the disputes arising from the contract. If a jurisdiction clause gave consumers the right to bring disputes in their place of residence, the clause was marked as clearly fair, whereas it was marked as clearly unfair if it stated that any judicial proceeding takes a residence away (i.e. in a different city, different country). As an example consider the following clauses taken from the Dropbox ToS:

```
<j3>You and Dropbox agree that any judicial proceeding to resolve claims
relating to these Terms or the Services will be brought in the federal or state
courts of San Francisco County, California [...]</j3>

<j1>If you reside in a country (for example, European Union member states) with
laws that give consumers the right to bring disputes in their local courts,
this paragraph doesn't affect those requirements.</j1>
```

The second clause introduces an exception to the general rule stated in the first clause, so the first one was marked as clearly unfair and the second as clearly fair.

A **choice of law** clause specifies which law will govern the relations arising from the agreement, and according to which law a potential dispute will be adjudicated. If the applicable law was determined based on the consumer's country of residence, the clause was marked as clearly fair. In any other case the choice of law clause was considered to be potentially unfair. The following example is taken from the Facebook ToS:

```
<law2>The laws of the State of California will govern this Statement, as well
as any claim that might arise between you and us, without regard to conflict of
law provisions</law2>
```

A **limitation of liability** clause specifies the amount and types of damages that the service provider will be obligated to provide to consumers under terms and conditions stipulated in the service agreement. Clauses that did not exclude or limit the liability were marked as clearly fair. Potential unfairness was attributed to clauses that reduced, limited, or excluded the liability of the service provider for damages (such as any harm to the computer system because of malware or loss of data), and for the suspension, modification, discontinuance or lack of the availability of the service. This classification was also applied to clauses as well as those containing blanket phrases like "to the fullest extent permissible by law". Clauses that reduced, limited, or excluded the liability of the service provider for physical injuries, intentional damages as well as in case of gross negligence, were marked as clearly unfair.

A **unilateral change** clause in favour of the provider specifies the conditions under which the service provider can amend and modify the ToS. Such clauses were consistently marked as potentially unfair.

A **unilateral termination** in favour of the provider details the circumstances under which the provider can suspend and/or terminate the service and/or the contract. We marked such clauses as follows: potentially unfair if the suspension or termination was allowed only under specific reasons and conditions; clearly unfair if they empowered the service provider to suspend or terminate the service at any time for any or no reasons and/or without notice. That was the case in the Academia terms of use:

```
<ter3>Academia.edu reserves the right, at its sole discretion, to discontinue
or terminate the Site and Services and to terminate these Terms, at any time
and without prior notice.</ter3>
```

A **contract by using** clause states that the consumer is bounded by the terms of use of a specific service, simply by using the service. We consistently marked such clauses as potentially unfair.

A **content removal** clause specifies the conditions under which the service provider may remove the user's content. We marked the clause as follows: potentially unfair if the clause specified reasons and conditions for such a removal; clearly unfair if it stated that the provider may remove content in his full discretion, and/or at any time for any or no reasons and/or without notice nor possibility to retrieve the content.

Finally, an **arbitration** clause requires the parties to resolve their disputes through an arbitration process, before the case can go to court. Tt is thus considered as a kind of forum selection clause. Such a clause may or may not specify that arbitration occur within a specific jurisdiction. We marked such a clause as follows: clearly fair if it defined the arbitration as fully optional; clearly unfair if it stated that the arbitration (1) takes place in a state other then the state of consumer's residence and/or (2) it is not based on law but on arbiter's direction; potentially unfair in all other cases.

4. Machine Learning Methodology

From a machine learning point of view, the problem of detecting unfair clauses within a contract can be seen as a sentence classification task. Given a sentence belonging to a contract, the goal is to classify it as *positive* (if the sentence expresses a clearly or potentially unfair clause) or *negative* (otherwise). In order to train a machine learning system able to distinguish positive from negative sentences, a supervised learning algorithm is typically employed. This framework assumes the availability of a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ made up of N pairs (x_i, y_i) where x_i is the representation of a sentence, and y_i is its corresponding label (also named class or target), which is the category to be predicted. In this case, we consider only two sentence categories, namely positive (clearly or potentially unfair) and negative (clearly fair), thus we deal with a binary classification task.

There are many approaches for the classification of sentences. In this paper we consider and compare two of them. The first one, known as *bag-of-words* (BoW), consists in representing a sentence as a vector of features that is as large as the dimension of the vocabulary of words within the dataset. Each feature is either zero (if the corresponding word does not appear in the sentence) or different from zero (if it does). The non-zero value in the feature vector associated to each word is the so-called TF-IDF score, that is the number of times the word appears in the sentence (Term Frequency, TF) multiplied by a term that amplifies the contribution of rare words (Inverse Document Frequency, IDF) [22]. A sentence representation, such as its BoW, can be fed to different types of machine learning classifiers. In this work we employ support vector machines (SVMs), as they are widely used in text classification [6]. Extensions of the BoW approach consider so-called n-grams, i.e. features extracted from the sentence by taking into account the frequencies of consecutive word combinations, and grammatical information such as part-of-speech tags, i.e., word categories such as nouns, verbs, etc. [7]. The BoW approach is thus built to leverage the *lexical* information within a sentence, and in particular the presence of keywords and phrases that are highly discriminative for the detection of unfair clauses.

The second approach we consider in our study is that of *tree kernels* [17] (TK). This approach takes into account the similarity between the *structure* of sentences and has been shown to offer state-of-the-art performance in related classification tasks, such as those typical of argumentation mining [10], for example claim detection [8]. The structure of a sentence is naturally encoded by its *constituency parse tree*, which describes the syntactic and grammatical characteristics of a sentence. A TK consists of a *similarity measure* between two trees, by taking into account the number of common substructures or *fragments*. Different definitions of fragments induce different TK functions. In our study we use the SubSet Tree Kernel (SSTK) [4] which counts as fragments those subtrees of the constituency parse tree terminating either at leaves or the level of non-terminal symbols. SSTK have been shown to outperform other TK functions in several argumentation mining sub-tasks [9].

5. Experimental Results

We performed experiments on the dataset described in Section 3, following a standard *leave-one-document-out* (LOO) procedure, whereby each document in the corpus is used, in turn, as test set for our classifier, while the remaining documents constitute the training set. In this way, we obtain predictions for each document in the dataset, and we measure the performance on each contract separately, thus evaluating the generalization capabilities of the system. In particular, we compute precision, recall and F_1 for each contract, and we finally compute the average for each of these three metrics (this is called macro-average [22]). Precision (P) is defined as the fraction of examples predicted as positives, which are actually labeled as positive. Recall (R) is the fraction of positive examples that are correctly detected. F_1 is finally the harmonic mean between precision and recall ($F_1 = \frac{2PR}{P+R}$).

As customary in studies of this kind, the above performance measures are compared with baselines that give an indication of the difficulty of the problem at hand. We aim to compare three systems:

1. a single SVM exploiting BoW (unigrams and bigrams), considering as the positive class the union of all tagged sentences;
2. a combination of eight SVMs exploiting the same features as above, but each considering as the positive class only one specific tag; a sentence is then predicted as unfair if at least one of the SVMs predicts it as such;
3. a kernel machine exploiting TK, considering as the positive class the union of all tagged sentences.

We adopt two standard baselines: a *random* baseline, which predicts unfair clauses at random,[3] and an *always positive* baseline, which predicts every sentence as unfair. If any of these baselines provided a result with acceptable accuracy, that would mean that the classification task has a trivial solution.

Table 1 shows the results achieved by each of these variants. We notice that the precision of baseline classifiers is below 8%, and that the precision of either BoW and TK is above 57%. Moreover, we notice how the single-model SVM performs best, outperforming both Tree Kernels, which exploit the same setting for the definition of positive class, and the combined-model SVM, which separately trains a different model for each category (tag) of unfair clauses.

These figures tell us something about the nature of the task. First, the better performance of the single model with respect to the combined model implies that knowing unfair clauses of different categories is useful to correctly predict the unfair clauses of a specific category. This is particularly important for corpora where few tagged examples exist for a certain category, but it is also interesting from a computational linguistic and legal point of view, since it seems to suggest the existence of *a common lexicon for unfair clauses*, which spans across several tag categories. Second, the worse performance associated with TK suggests that the syntactic structure of the sentence is probably not very indicative of the presence of an unfair clause—or, at least, that it is less informative than the lexical information captured by n-grams. This makes the task of detecting unfair

[3]Sampling takes into account the class distribution in the training set.

Table 1. Results on leave-one-document-out procedure.

Method	P	R	F_1
SVM – Single Model	0.620	0.715	0.648
SVM – Combined Model	0.576	0.621	0.582
Tree Kernels	0.571	0.665	0.603
Random Baseline	0.071	0.071	0.071
Always Positive Baseline	0.075	1.000	0.138

Table 2. Recall of abusive clauses for each tag category for the single and combined SVM models, micro-averaged on the whole dataset.

Tag	Single	Combined
Arbitration	0.531	0.344
Unilateral change	0.809	0.723
Content removal	0.677	0.645
Jurisdiction	0.826	0.826
Choice of law	1.000	0.778
Limitation of liability	0.614	0.602
Unilateral termination	0.780	0.744
Contract by using	0.579	0.342

clauses different from other text retrieval problems in the legal domain, such as, for example, the detection of claims and arguments [10].

In Table 2 we also report the recall of the single- and combined-model SVM for each separate tag category, micro-averaged on the whole dataset. The results show that all the categories benefit from the knowledge of unfairness given by the other categories: this is particularly significant for the "Arbitration" and "Contract by using" categories, which still remain the hardest to detect.

Interestingly, preliminary experimental results provided some feedback to the tagging: a number of apparent false positives where due to mistakes in tagging; they concerned unfair clauses that had escaped the analysts, due to the length and complexity of the ToS.

6. Conclusions

The use of machine learning and natural language processing techniques in the analysis and classification of legal documents is gaining a growing interest. Moens et al. [16] proposed a pipeline of steps for the extraction of arguments from legal documents, exploiting supervised classifiers and context-free grammars, whereas Biagioli et al. [3] proposed to employ multi-class SVM for the identification of significant text portions in normative texts. Recent approaches have focussed on the detection of claims [11] and of cited facts and principles in legal judgments [23], as well as on the prediction of judicial decisions [1]. A case study regarding the construction of legal arguments in the legal determinations of vaccine/injury compensation compliance using natural language tools was given in [2]. It is worth remarking that, in most of these works, classic lexical features such as BoW still represent a crucial ingredient of automated systems. Finally, privacy policies represent another strictly related application where machine learning approaches have proved effective (e.g., see [5] and references therein).

This paper presented a first experimental study that used machine learning to address the automated detection of potentially unfair clauses in online contracts. Our results seem encouraging: using a small training set we could automatically detect most unfair clauses, and with acceptable precision. Given that most unfair clauses are currently hidden within long and hardly readable ToS, the recall and precision offered by our approach may be already sufficient for practical applications.

Interesting and to some extent unexpected outcomes included the comparatively better performance of the BoW approach, and the fact that the automated detection method we developed was able to highlight a number of unfair clauses that human analysts had failed to identify in the first place.

This study was motivated by a long-term goal such as the pursuit of effective consumer protection by way of tools that support consumers and their organizations in detecting unfair contractual clauses. That is also the objective of a research and development project (CLAUDETTE) that has recently kicked off at the European University Institute. Looking to the future, we plan to carry out further analyses that enable us to determine what machine learning methods should be implemented in such future tools. Accordingly, we plan to conduct a qualitative analysis of the errors performed by our system, in order to identify weaknesses and improve performance. We are also working on the construction of a larger corpus, with the intention of improving training as well as providing a suitable dataset for testing other machine learning algorithms, such as deep networks, which have proven effective in several other natural language processing tasks. Finally, we are studying ways to exploit contextual information, since it was pointed out that a clause might be fair in a context but not in others.

References

[1] N. Aletras, D. Tsarapatsanis, D. Preoiuc-Pietro, and V. Lampos. Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. *PeerJ Computer Science*, 2:e93, Oct. 2016.

[2] K. D. Ashley and V. R. Walker. Toward constructing evidence-based legal arguments using legal decision documents and machine learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, pages 176–180. ACM, 2013.

[3] C. Biagioli, E. Francesconi, A. Passerini, S. Montemagni, and C. Soria. Automatic semantics extraction in law documents. In *Proceedings of ICAIL*, pages 133–140. ACM, 2005.

[4] M. Collins and N. Duffy. New ranking algorithms for parsing and tagging: Kernels over discrete structures, and the voted perceptron. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 263–270. ACL, 2002.

[5] B. Fabian, T. Ermakova, and T. Lentz. Large-scale readability analysis of privacy policies. In *Proceedings of the International Conference on Web Intelligence*, pages 18–25. ACM, 2017.

[6] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *ECML 98*, pages 137–142, 1998.

[7] E. Leopold and J. Kindermann. Text categorization with support vector machines. how to represent texts in input space? *Machine Learning*, 46(1-3): 423–444, 2002.

[8] M. Lippi and P. Torroni. Context-independent claim detection for argument mining. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 185–191, 2015.

[9] M. Lippi and P. Torroni. Margot. *Expert Syst. Appl.*, 65(C):292–303, Dec. 2016. ISSN 0957-4174. doi: 10.1016/j.eswa.2016.08.050.

[10] M. Lippi and P. Torroni. Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Technol.*, 16(2):10:1–10:25, Mar. 2016.

[11] M. Lippi, F. Lagioia, G. Contissa, G. Sartor, and P. Torroni. Claim detection in judgments of the EU Court of Justice. In *VI Int. Workshop on Artificial Intelligence and the Complexity of Legal Systems (AICOL)*, 2015.

[12] M. Loos and J. Luzak. Wanted: a bigger stick. on unfair terms in consumer contracts with online service providers. *Journal of Consumer Policy*, 39(1): 63–90, 2016.

[13] H.-W. Micklitz. On the intellectual history of freedom of contract and regulation. *Penn State Journal of Law & International A airs*, 4(1), Dec. 2015.

[14] H.-W. Micklitz and N. Reich. The court and sleeping beauty: The revival of the unfair contract terms directive (uctd). *Common Market Law Review*, 51 (3):771–808, 2014.

[15] H.-W. Micklitz, P. Pałka, and Y. Panagis. The empire strikes back: Digital control of unfair terms of online services. *Journal of Consumer Policy*, pages 1–22, 2017.

[16] M.-F. Moens, E. Boiy, R. M. Palau, and C. Reed. Automatic detection of arguments in legal texts. In *Proceedings of the 11th international conference on Artificial intelligence and law*, pages 225–230. ACM, 2007.

[17] A. Moschitti. Efficient convolution kernels for dependency and constituent syntactic trees. In J. Frnkranz, T. Scheffer, and M. Spiliopoulou, editors, *ECML 2006*. Springer Berlin Heidelberg, 2006.

[18] P. Nebbia. *Unfair contract terms in European law: a study in comparative and EC law*. Bloomsbury Publishing, 2007.

[19] J. A. Obar and A. Oeldorf-Hirsch. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. 2016.

[20] N. Reich, H.-W. Micklitz, P. Rott, and K. Tonner. *European consumer law*. Intersentia, 2014.

[21] H. Schulte-Nölke, C. Twigg-Flesner, and M. Ebers. *EC consumer law compendium: the consumer acquis and its transposition in the member states*. Walter de Gruyter, 2008.

[22] F. Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34(1):1–47, 2002.

[23] O. Shulayeva, A. Siddharthan, and A. Wyner. Recognizing cited facts and principles in legal judgements. *Artificial Intelligence and Law*, 25(1):107–126, Mar 2017.