INWEND: Using CBR to automate legal assessment in the context of the EU General Data Protection Regulation

Clarissa Dietrich[1], Sebastian Schriml[2], Ralph Bergmann[1,3], Benjamin Raue[2]

[1] Business Information Systems II, University of Trier, 54286 Trier, Germany
http://www.wi2.uni-trier.de
[2] Institut für Recht und Digitalisierung Trier, University of Trier, 54296 Trier, Germany
https://irdt.uni-trier.de/
{dietrich,schriml,bergmann,raue}@uni-trier.de
[3] German Research Center for Artificial Intelligence (DFKI), Branch University of Trier, Behringstraße 21, 54296 Trier, Germany
ralph.bergmann@dfki.de

Abstract The European General Data Protection Regulation (GDPR), which governs the processing of personal data in all EU Member States, contains an exemption for "purely personal or household activities". Whether this so-called household exemption covers the setup of an online communication forum, particularly on a social media or chat platform, is a question of the individual case. We present a case-based reasoning approach to automatically assessing a scenario provided by the user and generating a tailored legal recommendation.

Keywords: Knowledge Engineering · Legal Tech · Case-Based Reasoning · General Data Protection Regulation

1 Introduction

The European General Data Protection Regulation (GDPR)[4] entered into force on 25 May 2018, establishing a new data protection law in all EU Member States. Despite a two-year transition period, the advent of this novel legal framework has caused considerable legal uncertainty: Companies, associations and other institutions have sought legal advice in order to ensure compliance with the regulation and to avoid its potentially harsh sanctions: Article 83(5) GDPR allows for administrative fines up to EUR 20 million, or in the case of an undertaking,

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119/1.

up to 4 per cent of the total worldwide annual turnover of the preceding financial year, whichever is higher.

The material scope of application of the GDPR, however, is not restricted to enterprises and organizations: Virtually every individual handles personal data of others incessantly and can therefore be subject to data protection law. Seemingly innocuous examples of data processing include private blogging, the use of social media accounts, taking and sharing pictures of others and a variety of similarly common activities.

Legal uncertainty entails negative consequences beyond mere lack of compliance, specifically so-called "chilling effects"[5]: Individuals may be deterred from socially desirable activities even though these could have been implemented in conformity with data protection laws. The need for high-quality, low-threshold legal expertise is therefore a worthy area of application for artificially intelligent systems.

In the interdisciplinary INWEND[6] project at the University of Trier, we are developing a software prototype that can elicit relevant factual information from non-lawyers and automatically generate a (preliminary) legal assessment using case-based reasoning (CBR) techniques. The focus of this project is the so-called household exemption in Article 2(2)(c) GDPR which excludes from the scope of the regulation the processing of personal data "by a natural person in the course of a purely personal or household activity".

This project has a pilot character for the research at the intersection of artificial intelligence and European law. In an interdisciplinary cooperation between legal experts and computer scientists the understanding of the respective other discipline is expected to be improved, which should also enable the expansion of research activities to other common interdisciplinary fields of research. The scientific relevance of the development of a structured legal case base which facilitates automated legal reasoning lies in the possibility to work with consistent and structured case data in the future in order to develop artificially intelligent systems.

The remainder of the paper is organized as follows. In section 2 we outline related work on the use of CBR systems in a legal context as well as recent but distinct research in legal information technology. In section 3 we provide an overview of the relevant legal domain and the scope of the prototype (in the interest of brevity called INWEND hereinafter). In section 4 we proceed to describe our approach to modelling this knowledge domain by using an initial case situation and approximately 200 case variations generated using a set of

[5] The term is defined by the Oxford Dictionary as "a discouraging or deterring effect on the behaviour of an individual or group, especially the inhibition of the exercise of a constitutional right, such as freedom of speech, through fear of legal action.", https://www.lexico.com/definition/chilling_effect.

[6] INWEND stands for "Intelligente Wissensbasierte Entscheidungsunterstützung für juristische Fragestellungen am Beispiel des Datenschutzrechtes", which translates to "Intelligent knowledge-based decision support for legal questions on the example of the data protection law".

parameters. We conclude with some remarks on the feasibility of our approach in civil law jurisdictions.

2 Foundations and Related Work

The automation of legal tasks in the broadest sense is currently a topic of intense economic and scientific interest, with a variety of projects and products being discussed under the broad umbrella term of "Legal Tech" [6,8]. While the upsurge in public attention afforded to this field is a more recent phenomenon, the roots of legal informatics are much older: Many influential works date back to the second half of the twentieth century and predate the development and accessibility of powerful computers, large data storage capacities and the Internet.

The general suitability of CBR techniques for handling legal tasks was discovered early, partly because the CBR approach exhibits structural similarities to legal reasoning that are particularly apparent in the common law with its doctrine of precedent and the principle of *stare decisis*[7]. We will first outline some pioneering research work conducted in the 1980s and 1990s and then contrast the INWEND project to this and other contemporary research with a different outlook.

2.1 CBR in the Legal Domain

The possibility to associate the CBR approach with legal reasoning and argumentation was discovered in as early as the 1980s [17]. The pioneering legal reasoning system HYPO [16] used CBR techniques to generate arguments by comparing and contrasting cases with the use of a case base [3]. HYPO then inspired a multitude of further approaches and systems: The successor system CABARET [18] (which stands for "Case-Based Reasoning Tool") integrated CBR with a rule-based approach. The system CATO [2] ("Case Argument Tutorial") was designed to teach argumentation skills to students, whereas IBP [7] ("Issue-Based Prediction") was aimed at the prediction of case outcomes. While these approaches were all rooted in the common law legal system, the INWEND project studies, inter alia, the applicability of these approaches and findings to German and European law.

2.2 Other Related Work

More recently, the ARGUMENTUM [11] project at the German Research Center for Artificial Intelligence has worked on argument extraction from German Federal Constitutional Court decisions. The relevant text passages are extracted from a corpus of decisions to justify or refute assertions in order to construct a convincing argument. While in the German legal system statutory law is regarded as a main source of reasonable legal arguments, decisions makers are to a large extent guided by the reasoning brought forward in court decisions.

[7] Latin for: "Let the decision stand."

The INWEND project takes a different approach, setting aside argument mining and starting from highly structured legal case information provided by domain experts.

3 Legal Background and Project Scope

We begin our account with a brief introduction to the legal domain of our projected CBR system, in order to both provide some context for our development strategy and highlight the practical need for the solution we propose.

3.1 Significance of the Household Exemption in Art. 2(2)(c) GDPR

According to Article 2(1) of the GDPR, the regulation applies, inter alia, to the "processing of personal data wholly or partly by automated means", comprising practically any use of data processing systems [14][8]. The broad material scope of application of the GDPR is counterbalanced by exemptions listed in Article 2(2) GDPR, in particular the so-called household exemption in Article 2(2)(c) GDPR.

This clause excludes from the scope of application the processing of personal data "by a natural person in the course of a purely personal or household activity". This means that any such activity may be carried out without having to observe any of the obligations under the GDPR. These comprise a variety of legal duties, such as the principles relating to the processing of personal data set forth in Article 5, the need to justify the processing according to Article 6 and to obey the rights of the data subject set forth in Articles 12 et seqq. Being an exception to the entire data protection regime, there is general agreement that the clause must be interpreted restrictively [12][9].

3.2 Scope of the INWEND System

A general problem of legal interpretation, including the construction of statutory law, is the open-textured nature of legal concepts [1] leading to ambiguity. For instance, under which circumstances does the household exemption cover (private, not business-related) communication in a chat or online forum? As the exchange of information via technical communication devices generally entails the processing of personal data, the applicability of the GDPR hinges upon the question whether such communication is to be deemed a "purely personal or household" activity.

The INWEND system is designed to address the need for an automated legal assessment of the question: Does a natural person opening a communication forum on an online platform comply with the household exemption? The practical implication of this system is that a user may enter the circumstances of their

[8] Paal/Pauly-*Ernst*, Art. 2 GDPR margin number 5.
[9] Kühling/Buchner-*Kühling/Raab*, Art. 2 GDPR margin number 21.

(intended) use of such a forum and be immediately provided with (preliminary) legal feedback. The availability of such a tool may, as a secondary effect, incentivise platform providers to make available privacy-friendly options to attract more users. In the following section we will describe our strategy for modelling this legal problem and the first results of our research.

4 Approach to Modelling the Household Exemption with CBR

In case-based reasoning the task of solving problems is based on previous experience, which is stored in the form of cases in a case base [4]. This method of experience storage must be reconciled with the experience knowledge of the household exemption. A legal question should be represented in a way a CBR system can interpret as a case. We select a structural approach, thus a case is represented in attribute-value form using a structured vocabulary [4]. This chapter describes how we determine suitable attributes and create a functional case base in order to appropriately represent legal thinking for CBR.

4.1 Case Base Construction from an Initial Case Situation

The CBR approach relies on the existence of a case base, from which relevant reference experiences can be retrieved. With the GDPR being a novel legal framework, the availability of pertinent court decisions – particularly by the authoritative European Court of Justice (ECJ) – is low. Thus, there is little documented experience available that could be turned into a case base.

We therefore took a different approach in acquiring knowledge from the knowledge domain: Based on an interpretation of the household exemption, we developed an exemplary initial case situation. This is a short factual description of the scenario in which the question concerning the applicability of the household exemption arises. In particular, we consider the situation in which a user wants to open a forum on a communication platform and invite others to join the discussion.

This allowed us to identify a number of parameters which pertain to the applicability of the household exemption: Who is the intended circle of participants (such as family members, close friends, acquaintances), and how many people will be accessing the forum? How much personal information of other participants in the discussion will the user gain? Are there technical asymmetries by which the user will have more insight into other participants' personal data than vice versa?

Since legal ontologies are commonly used in legal informatics as a formal knowledge model [13], we considered using an existing legal ontology as a means of knowledge representation. Existing ontologies include LKIF Legal Ontology [9], Criminal Law Ontology [19] and PrOnto [15]. While the LKIF Legal Ontology and the Criminal Law Ontology are not focussed on the EU General Data Protection Regulation, PrOnto models the GDPR main concepts but does not

provide a sufficiently sophisticated representation of the household exemption. As none of these ontologies seemed apt to model our situation adequately, we decided to define a custom set of parameters in order to represent cases with a basic ontology.

4.2 Impact of Legal Reasoning on Case Base Structure

With the help of domain experts we created a documentation of the selected parameters and their features, providing strategies for the assessment of problematic cases as well as sets of positive and negative examples. In the next step, we used these parameters to synthetically generate a case base of approximately 200 cases. Since not all combinations of the parameter features were meaningful from a legal standpoint, the case base was slightly smaller than the product of all features: With five parameters, each of which ranging between two and five features, a total of 240 combinations would have resulted; since 48 of these combinations involved irreconcilable case details, there were only 192 legally meaningful cases.

We then assessed the applicability of the household exemption in the cases of the case base. It is worth noting that, rather than judging cases on an individual basis, legal domain experts employ two strategies to determine the outcome of cases by groups: Firstly, they try to identify so-called "edge cases", these being parts of the case base where the legal assessment reaches a "tipping point" between two contrary results. Secondly, they extensively employ reasoning *a fortiori*[10], arguing for instance that a case cannot qualify for the household exemption where a comparable case with a more restricted circle of participants has already been rejected: Since a wider circle of participants is an argument against the applicability of the household exemption, such a decision would be inconsistent.

The consistency of a legal case base is a notion which was discussed earlier in the context of the common law, specifically with regard to the doctrine of precedent [10]: In a legal system where court decisions can have a binding effect, it is of evident importance that courts are able to determine whether their envisaged decision is commensurate with the already existing legal framework. With our approach being rooted in a different jurisdiction, it is an important question of our research how this notion of consistency is to be construed in civil law and attained in our system. This is true in both a conceptual sense, which refers to the structure of the case base, and in a pragmatic sense, which is concerned with actual decision making. It is therefore yet to be explored how a consistent case base can be generated from expert knowledge in an expedient and robust way.

4.3 Implementation of the Model in ProCAKE

Drawing on the aforementioned domain knowledge, we then implemented the case representation using the CBR system ProCAKE [5]. This is a framework

[10] Latin for: "from the stronger".

for structural and process-oriented CBR applications developed by the Department of Business Information Systems II at the University of Trier. The case representation comprises the parameters as well as the range of acceptable values corresponding to the contents of the case base. Additionally, the user is provided with an option to make no specification in order to allow for factual uncertainty.

Considering the argumentation pattern mentioned in 4.2, whereby the applicability of the household exemption can be determined by *a fortiori* reasoning, we chose to model the parameter *circle of participants* as a polyvalent attribute. Cases that differ merely with regard to this parameter can be summarized as one case by representing the values of the respective cases for this parameter in a set. This step reduces the case base by 30 cases. At the time of writing, further modifications of the model are still in progress.

Based on the case model and the case base, a first similarity model was created. For calculating the global similarity the individual parameters are weighted by their impact on the assessment of the case. For example, the aforementioned parameter *circle of participants* has a particularly strong bearing on the outcome of the assessment.

The similarity for the parameter *circle of participants* is computed with a set mapping. If the value of the query matches a value in the set of a case of the case base, the similarity is high, otherwise low. The other local similarities are implemented with simple similarity measures, with equal values being assigned a high similarity score and different values being assigned a low similarity score. Again, at the time of writing, further modifications of the model are in progress.

4.4 Realization of a Graphical User Interface

We developed a prototypical Graphical User Interface (an excerpt is shown in Figure 1), to visualize the intended interaction with the program. The user has to respond to questions by choosing the answer most fitting to the situation at hand. As mentioned under 4.3, the questions also include the option to abstain from answering any individual question.

Each question corresponds to an attribute in our case model, with the answers determining their possible values. As can be seen in Figure 1, we guide the user through the interaction by providing additional information on questions and answers as needed. For additional reference, the factual situation described by the selected parameters is summarized for the user, which can be seen in Figure 2. Specifically, the text seen underneath the headline “Sachverhalt” (facts of the case) is a summary of the factual situation automatically generated from the answers given by the user; the text underneath the headline “Einschätzung” (legal opinion) is an easily comprehensible assessment of this factual situation, pertaining to the plain result – whether the household exemption is applicable or not – as well as its legal certainty.

On the basis of the answers selected by the user a query case is created. When the user clicks the submit button, this query is transmitted to the back end as a JSON object and then relayed as a query to a ProCake server, where it

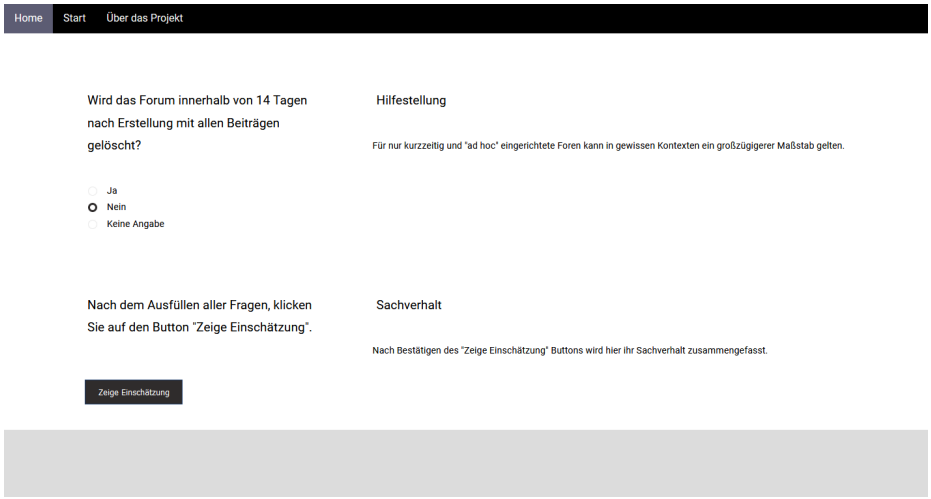

Fig. 1. Excerpt of the GUI

is used for a retrieval on the case base. The retrieval result containing the most similar case is then returned to the web interface. The outcome of this retrieved case is displayed on the GUI (see Figure 2) and suggested to the user as the closest match to their query.

Fig. 2. Exemplary assessment of a submitted case.

In summary, this first version of a GUI generates a legal assessment in the context of the household exemption for a user case, based on the presented parameters, case base and similarity model.

5 Conclusion and Future Work

This paper presents a first CBR approach to modelling the legal assessment of a factual scenario in the context of the GDPR household exemption. The prototype system developed in the INWEND project is designed to elicit relevant facts from the user and to automatically generate a tailored legal recommendation. In the current stage of development, we have generated a substantial case base to reflect the structure of the GDPR household exemption. The system is controlled via a web interface and is able to retrieve reference cases from its case base.

In a next step, we plan to refine our ontology which represents our concept of the case representation. The ontology serves as a translation help between users and domain experts.

We judge the scalability of this approach to a wider field of law as rather difficult, because the expert knowledge required is expensive and the acquisition effort is high. Thus, a further interesting task to work on is to generate new cases semi-automatically: An algorithm suggests a scenario based on the case parameters and an expert decides the applicability of the household exemption. This algorithm should determine the potential edge cases and disregard cases that do not improve the ability to decide a case.

Further work needs to be undertaken to identify the most relevant cases in the current case base and to adapt the domain and similarity model. We would also like to evaluate our approach with test users in order to verify its reliability and practical utility.

We conclude that CBR systems are generally capable of providing initial legal assessments for non-lawyers in civil and EU law jurisdictions. We could imagine combining our approach with text mining strategies to extract case features from a plain text entered by the user instead of generating the query from a user's questions and answers.

Acknowledgements. This work is funded by the German Federal Ministry of Education and Research (BMBF) under grant number 01UG1920.

References

1. Aikenhead, M.: The uses and abuses of neural networks in law. Santa Clara High Technology Law Journal **12**, 31–70 (1996)
2. Aleven, V.: Teaching Case-Based Argumentation through a Model and Examples. Ph.D. thesis, University of Pittsburgh (1997)
3. Ashley, K.D.: Reasoning with cases and hypotheticals in hypo. International Journal of Man-Machine Studies **34**(6), 753 – 796 (1991)
4. Bergmann, R.: Experience Management: Foundations, Development Methodology, and Internet-Based Applications, LNCS, vol. 2432. Springer (2002)
5. Bergmann, R., Grumbach, L., Malburg, L., Zeyen, C.: Procake: A process-oriented case-based reasoning framework. In: Proceedings of International Conference on Case-Based Reasoning (ICCBR) (2019)

6. Breidenbach, S., Glatz, F.: Rechtshandbuch Legal Tech. Beck C. H., München (2020)
7. Brüninghaus, S., Ashley, K.D.: Predicting outcomes of case based legal arguments. In: Proceedings of the 9th International Conference on Artificial Intelligence and Law. p. 233–242. ICAIL '03 (2003)
8. Hartung, M., Bues, M.M., Halbleib, G.: Legal Tech - A Practitioner's Guide. Bloomsbury Academic, London (2018)
9. Hoekstra, R., Breuker, J., Di Bello, M., Boer, A.: The lkif core ontology of basic legal concepts. In: Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques. CEUR Workshop Proceedings, vol. 321, pp. 43–63 (2007)
10. Horty, J., Bench-Capon, T.: A factor-based definition of precedential constraint. Artificial Intelligence and Law **20**, 181–214 (2012)
11. Houy, C., Niesen, T., Calvillo, J., Fettke, P., Loos, P., Krämer, A., Schmidt, K., Herberger, M., Speiser, I., Gass, A., Schneider, L., Philippi, T.: Konzeption und Implementierung eines Werkzeuges zur automatisierten Identifikation und Analyse von Argumentationsstrukturen anhand der Entscheidungen des Bundesverfassungsgerichts im Digital-Humanities-Projekt ARGUMENTUM. Datenbank-Spektrum **15**(1), 15–23 (2015)
12. Kühling, J., Buchner, B.: Datenschutz-Grundverordnung. No. 2, C.H.Beck (2018)
13. Liebwald, D.: Auf dem Weg zum Begriff: Vom Rechtswort zur Rechtsontologie – Automatisierte Verfahren zur semantischen Erschließung von Texten. Wort - Bild - Zeichen: Beiträge zur Semiotik im Recht **13**, 203–223 (2012)
14. Paal, B., Pauly, D.: Datenschutz-Grundverordnung. No. 2, C.H.Beck (2018)
15. Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L.: Legal ontology for modelling gdpr concepts and norms. In: Proceedings of the 31st International Conference on Legal Knowledge and Information Systems (JURIX). pp. 91–100 (2018)
16. Rissland, E., Ashley, K.: Hypo: A precedent-based legal reasoner. Advanced Topics in Law and Information Technology pp. 213–234 (1989)
17. Rissland, E., Ashley, K., Branting, L.: Case-based reasoning and law. Knowledge Engineering Review **20**, 293–298 (2005)
18. Skalak, D., Rissland, E.: Arguments and cases: An inevitable intertwining. Artificial Intelligence and Law **1**, 3–44 (1992)
19. Soh, C., Lim, S., Hong, K., Rhim, Y.Y.: Ontology modeling for criminal law. In: Pagallo, U., Palmirani, M., Casanovas, P., Sartor, G., Villata, S. (eds.) AI Approaches to the Complexity of Legal Systems. pp. 365–379. Springer (2018)