CEUR-WS.org/Vol-2738/LWDA2020_paper_14.pdf

Construction of a Corpus for the Evaluation of Textual Case-based Reasoning Architectures

Andreas Korger[1] and Joachim Baumeister[2,3]

[1] Angesagt GmbH, Dettelbachergasse 2, D-97070 Würzburg
[2] denkbares GmbH, Friedrich-Bergius-Ring 15, D-97076 Würzburg
[3] University of Würzburg, Am Hubland, D-97074 Würzburg

Abstract. Regulatory documents denote an interesting application domain for case-based knowledge management. These documents enumerate situations with conditions, that are often dangerous for human and environment and they give advice, rules, and instructions for prevention or handling. That type of documents is eminent in many domains and provides valuable experience knowledge which makes it a remarkable application and research domain for (textual) case-based reasoning. In this paper, an initial case-based representation of regulatory documents is introduced. We report on the construction of an open corpus of regulatory documents in the domain of nuclear safety regulations.

Keywords: Case-based Reasoning · Experience Management · Knowledge Management · Textual Case-based Reasoning · Corpus Annotation · Natural Language Processing.

1 Introduction

Case-based knowledge management approaches seem promising to handle regulatory documents. In general, regulatory documents are published by authorities covering the handling of situations of a specific domain. A document enumerates noteworthy situations with conditions, that are often dangerous for human and environment. Based on a detected situation the document gives advice, rules, and instructions for preventing or handling a particular situation.

Examples of regulatory documents are *compliance documents* of large companies, *safety documents* for conventions and festivals, and *legislative agreements* between parties.

This type of document is eminent in many areas and provides valuable experience knowledge. Furthermore, a single document often covers experience knowledge from different domains thus requiring the collaboration of many domain experts. For instance, the compliance document of a company will require prevention and handling rules defined by experts from the law and human resources departments. A safety document of a festival will cover regulatory situations defined by experts from the fire department and the police.

Copyright © 2020 by the paper's authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Today, the use of regulatory documents faces two essential challenges:

- Creation: Documents are commonly created manually and collaboratively by domain experts in a knowledge intensive process. The reuse of existing knowledge especially found in existing regulatory documents, is mostly not present.
- Retrieval: The documents are usually available in plain text and therefore the retrieval of suitable information for a given situation is complex and time intensive.

We see *regulatory documents* as an interesting application and research domain for (textual) case-based reasoning and we formulate the following hypotheses:

1. Case-based reasoning can provide a natural representation for covering experience knowledge.
2. Case-based reasoning can handle incomplete input for the formulation and retrieval of experience knowledge.
3. Case-based reasoning is very suitable to integrate background knowledge into the process provided, e.g., by ontologies.

To tackle these hypotheses, we started to build an open corpus of regulatory documents that can be used to evaluate interesting research questions:

- Case-oriented representation of regulatory documents
- Textual CBR in the domain of regulatory documents
- Retrieval of regulatory experience knowledge
- Reuse and generation of new experience knowledge
- Case-based review critique of regulatory documents
- (Textual) quality assessment of regulatory documents

To the knowledge of the authors there exists currently no open corpus of regulatory documents, that can be used to work on the research questions stated above.

In this paper, we introduce an initial case-based representation of regulatory documents and we report on the construction of an open corpus of regulatory documents in the domain of *nuclear safety regulations*. We invite volunteers to join this construction process. Preceding work in the field of knowledge management and case-based reasoning laying fundamentals was presented by Korger and Baumeister [22, 24].

The paper is organized as follows: In Section 2 we introduce the PIRI structure representing regulatory experience knowledge as a case-based interpretation and we sketch how case-based reasoning is used to work with regulatory documents. The construction of the corpus of nuclear safety regulations is described in Section 3, where we introduce the intention and the statistics of the corpus. We show the methods used for the construction of the corpus and we explain how to obtain the corpus for own research. In Section 4 we discuss a number of use cases applicable for the introduced corpus. The paper is concluded with related and future work in Section 5.

2 Case-Based Representation of Regulatory Documents

A regulatory document lists a collection of incidents of interest for a given life or work context. For each incident of interest, the document usually describes measures for prevention and measures for handling an occurring incident. To know, which incidents are actually of interest in a certain scenario and which measures are effectively applicable, is the result of experience collected in the past. A key to access this tacit knowledge is to identify parts of the document corresponding to incidents and measures, as well as to find a metric to make them comparable within a certain context. These considerations will be reflected in the case structure presented in the following.

2.1 The Domain of Regulatory Documents

A regulatory document consists of a sequence of text passages describing specific aspects of the domain. We call interesting passages within a document *information units*.

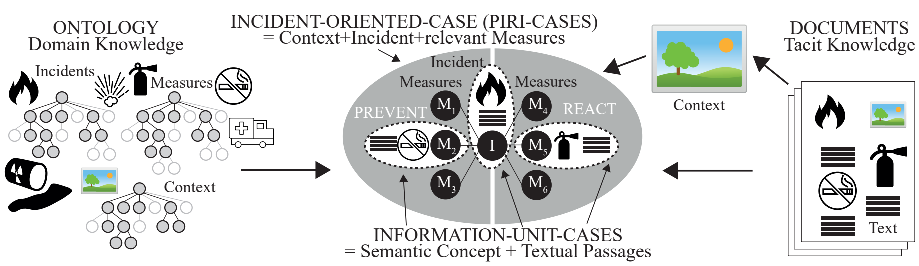

Fig. 1: Domain knowledge with semantic information represented in an ontology and tacit knowledge encoded in documents yielding two different case structures.

Definition 1 (Information Unit and Corpus). *An* information unit u *is defined as a text passage within a document, i.e.,* $u = (d, f, t)$*, where* d *is a document URI and* f, t *are offset information describing the extract of the document. For a corpus* $\mathcal{RD}$*, the universal set of all information units is defined as* $IU_{\mathcal{RD}}$*.*

For semantic interpretation of the text passage, such information units need to be annotated by metadata, explicitly describing the content.

Definition 2 (Ontology). *An ontology* $\mathcal{O} = (E, R)$ *contains metadata relevant for the considered domain. Here, relevant semantic concepts represented as entities* $e \in E$*, that are connected by relations* $r \in R$*.*

Typical examples for entities within a regulatory domain are *fire*, *smoking prohibition*, *explosion*, and *evacuation*. Common relations in this context are *partOf* and *requires*. A symbolic depiction of the ontology, the documents, and the case considerations can be seen in Figure 1.

When defining regulatory documents in a case-based representation, we distinguish two design alternatives:

- *Information unit cases:* A case is represented by a text passage (information unit) describing an incident or measure. The case also contains annotations explicitly describing the incident/measure using elements of the ontology. A case base then contains a collection of information units as cases and incident/measure annotations are attributes representing the particular cases.
- *Incident oriented cases:* One distinct case is represented by one incident together with all measures mentioned in a document for preventing and handling the incident. The case base then contains a collection of incidents with measures as attributes representing the particular cases. It is worth noticing, that incident oriented cases are constructed by aggregating information unit cases.

In the following, we describe both representation alternatives in more detail and we motivate their use by retrieval and reuse examples.

2.2 Information Unit Case Structure

To build a bridge between semantic concepts and free text an *information unit case* combines a distinct information unit with a corresponding metadata annotation. Having in mind that the overall goal is to reuse documents, it is convenient to consider textual passages as solutions to structured problem descriptions [14]. Subsequently, it is assumed that the textual passage is reusable for similar problems.

Definition 3 (Information Unit Case). *An information unit case c_u is defined as follows: $c_u = (a, u)$, where $u \in IU_{\mathcal{RD}}$ is an information unit from corpus $\mathcal{RD}$ and $a \subseteq \mathcal{O}$ is describing metadata from an ontology $\mathcal{O}$. We call the case base of information unit cases $\mathcal{CB}_U = \{c_{u_1}, ..., c_{u_n}\}$ the collection of all cases $c_{u_i}, \forall i \in \{1, .., n\}$ that are extracted and annotated from available regulatory documents.*

A case $c_u = (p_u, s_u)$ representing one information unit is defined by a set of named entities and relations between them as problem description p_u and textual passage contained in a document indicating the fulfillment as solution s_u to the problem. These textual passages are referred as fulfilling textual features. A textual passage may be just one word up to some sentences. Depending on the use case scenario the description and the solution of the case might switch. For instance given a textual passage as problem description the metadata is the solution. An exemplary information unit case is:

$c_x =$ (piri:manualFireFighting, "find a fire extinguisher and put out the fire")

Similarity of Information Units To define similarity functions for information unit cases, we exploit the taxonomic interrelation of semantic concepts. Both incidents and measures can be well classified into a taxonomy, building the base for the similarity assessment and the adaptation. The similarities are calculated via the taxonomic order of its elements. Each element of the hierarchy is assigned with a likelihood symbolizing the similarity of its sub-elements. The similarity of the leaf elements is set to *1* and to *0* for the root element. The similarity increases with depth d of the element according to for instance $sim_d = 1 - 1/2^d$, [11]. With this atomic case structure, basic retrieval and reuse is possible. Picking up the previous example one could state the question, what to do, if there is no fire extinguisher in reach. This can be solved e.g. by the retrieval of measures more special than *manual fire fighting*, like *manual fire fighting with clothes*. A retrieved case would be for instance:

c_x = (piri:manualFireFightingWithClothes, "take off your jacket and use it to put out the fire by throwing it onto the flame")

2.3 Incident-Oriented Case Structure (PIRI)

A reoccurring pattern sharpens the focus on the essence of the regulatory documents. The pattern selected for this purpose bases on the assumption that a regulatory document delivers a core message. In terms of safety, the most important content are the incidents it mentions as well as the measures to prevent them or to react to their consequences. For a given context and relevant incident induced by the context the according measures are ordered by importance and classified into preventive and reactive measures. Other patterns might be considered for other tasks in an analogous way.

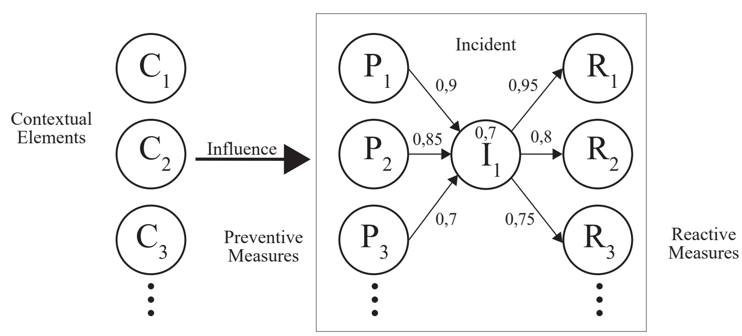

Fig. 2: PIRI-diagram under a context $C = (C_1, ..., C_j)$ showing the ranked preventive (P) and reactive measures (R) with the according importance weights for the incident and measures.

We call the instances of these information patterns PIRI-snippets (Preventive-Incident-Reactive-Interrelation) [23]. The presented model reduces the complexity of the real world for facilitation of assessment. Typically there is a cascade of measures that are executed in a specific order. For instance in the case of fire, first evacuate all people, then close the doors and windows. The PIRI-pattern in its graphical representation can be seen in Figure 2. The corresponding case structure is given in the following definition:

Definition 4 (The Incident-Oriented PIRI-Case). *A PIRI-case is defined as follows: $c_{PIRI} = (P_I, I, R_I, C_I)$, where $I \in \mathcal{CB}_U$ is a case describing a specific incident, $P_I \subseteq \mathcal{CB}_U$ denotes a set of cases describing preventive measures for I, $R_I \subseteq \mathcal{CB}_U$ denotes a set of cases describing reactive measures for I, $C_I \subseteq \mathcal{CB}_U$ denotes a set of cases describing the context of the incident I. We call the case base of PIRI-cases $\mathcal{CB}_{PIRI} \subseteq \mathcal{P}(\mathcal{CB}_U)$, where $\mathcal{P}(\mathcal{CB}_U)$ is the set of all subsets from $\mathcal{CB}_U$.*

A case describing one PIRI-snippet is the subset of the case base $\mathcal{CB}_U$ containing only cases of information units that are related to an incident I. In this case definition the importance weights are neglected. We see, that PIRI-cases are an aggregated form of information unit cases, and more sophisticated retrieval and reuse scenarios are possible. An exemplary PIRI-case for the incident *fire* looks as follows.

c_x = ([piri:fireAlarmSystem, piri:fireDrill], piri:fireIncident,
[piri:callFireDepartment, piri:manualFireFighting, piri:evacuate],
piri:doc_id123_fire_fighting_in_power_plants)

Similarity of PIRI Cases To retrieve similar cases for instance similar PIRI-snippets, the case base is searched for similar problem descriptions p_i to the query q_1. A query is made up by a set of named entities, relations between them and textual passages. With an aggregation function a global similarity measure is composed by weighting the previously described local similarity functions of information unit cases with the parameters $(\omega_P, \omega_I, \omega_R)$ and summed up as follows:

$$Sim_{\mathrm{PIRI}}(c_k, c_l) = \frac{\omega_P Sim_P(P_k, P_l) + \omega_I Sim_I(i_k, i_l) + \omega_R Sim_R(R_k, R_l)}{\omega_P + \omega_I + \omega_R} \quad (1)$$

If we want to compare two PIRI-snippets, then it is desirable to consider the context. For this reason we define the following extended similarity measure under the context C:

$$Sim_{\mathrm{PIRI+C}}(c_k, c_l) = \frac{\omega_1 Sim_{\mathrm{PIRI}}(c_k, c_l) + \omega_2 Sim_C(Cont_k, Cont_l)}{\omega_1 + \omega_2} \quad (2)$$

A similarity measure Sim_C of two documents for context comparison can be obtained from the background information provided in the documents as free text. The available documents most often contain brief background information and summaries of document objective, scope, and structure. This information can be exploited for context assessment. These passages can be compared to approximate the documents similarity using textual similarity measures [18, 25]. The specific case-based modeling of the context is in scope of future work.

2.4 Corpus of Nuclear Safety Regulations

For the initial construction of the corpus we annotated 143 documents summing up to about 17.500 pages of nuclear safety regulations as described before. The International Atomic Energy Agency (IAEA) granted their permission to be the source for all these documents [5]. Yet there are many more sources of documents in the same domain as partially summarized in Table 1.

Table 1. Overview of available documents in the domain.

Source	Language	Documents
International Atomic Energy Agency [5]	English	150
Bundesamt für Strahlenschutz [2]	German	80
Consejo de Seguridad Nuclear [4]	Spanish	100
Canadian Nuclear Safety Commission [3]	English	80
U.K. Office for Nuclear Regulation [7]	English	250
U.S. Nuclear Regulatory Commission [8]	English	2000
Autorité de Sûreté Nucléaire [1]	French	40
Ispettorato Nazionale per la Sicurezza Nucleare [6]	Italian	20

The corpus described in this work can be rebuilt following some basic steps. In the future (CRC of this paper), we will provide a ready-to-use download. All resources are available via GitHub [21]. Tools to aid in building and working on the corpus by oneself are explained in the source file or readme files. The corpus consists of three major parts. Scripts written in Java provide functionality, ontologies contain the data, a case base provides case-based similarity assessment, retrieval, and adaptation capacity. The ontology was implemented using the semantic wiki KnowWE [12]. For the case-based implementation we made use of the framework myCBR [11]. For the textual structuring of the regulatory documents an ontology was implemented [22]. Core components make use of the SKOS ontology (Simple Knowledge Organization System) [28] and the PROV ontology [26] as upper ontologies.

To access documents without annotation the present annotation information can be used as seeds for a case-based bootstrapping strategy to mine new cases [16]. This semi-supervised approach for entity and relation extraction supports the user in an active learning scenario. The algorithm compares the unknown corpus $\mathcal{B}$ with all n-grams retrieved from the already annotated corpus $\mathcal{A}$. It uses different similarity measures for n-grams of different sizes. To compare large text passages a text based similarity measure like tf-idf or sentence embeddings is used, smaller information units are compared using case-based similarities. The result is the set of automatically extracted annotations most similar (to a certain threshold) to manually verified information units existent in the corpus $\mathcal{A}$. The user reviews the generated annotations, adjusts them if necessary, and adds them to the case base as *manually verified annotations*.

Algorithm 1: Algorithm for semi-supervised case-based bootstrapping.

Data: Annotated corpus $\mathcal{A}$, not annotated corpus $\mathcal{B}$
Result: Set of new annotations for corpus $\mathcal{B}$ with high similarity
gram_max=length of largest annotated n-gram in $\mathcal{A}$;
v=similarity switch (e.g. 10 words);
Query ontological entity labels to the corpus $\mathcal{B}$;
Construct new ontology with retrieved machine annotations;
Get all annotations from the corpus $\mathcal{A}$ in gram-size-order;
while $i < gram\_max$ **do**
  if $i < v$ **then**
    $sim_{gram} = sim_{case-based}$
  else
    $sim_{gram} = sim_{text-based}$
  Query i-grams to machine annotations of $\mathcal{B}$ with sim_{gram};
  Request user review for most similar information units;

3 Use Cases

We exemplify the previous approach by use case scenarios contained in the corpus to verify the hypothesizes stated in the introducing section. In the first part of this section the semantic retrieval is illustrated. Afterwards, we demonstrate how textual passages can be adapted using the hierarchical relation of entities. In the following, we describe the usefulness of the PIRI-approach for the generation of new document sketches.

3.1 Semantic Search and Retrieval of Documents

A semantically annotated corpus allows for semantic search of documents fitting to a given problem description. Therefore, the user query is translated into a structured query to the case base. For instance, the question *"Which measures prevent a fire?"* can be analyzed in a first step retrieving the contained entities *Measure*, *Fire*, and the relation *isPreventiveMeasureFor*. The most special relational triple according to their hierarchical interrelation, fulfilling the elements extracted out of the user query, is <*Measure isPreventiveMeasureFor Fire*>. All ontological elements containing entities of the hierarchy, that are equal or more special, should be respected and thus retrieved. For instance:

- *piri:fireWatch piri:isPreventiveMeasureFor piri:fireIncident*
- *piri:fireBarriers piri:isPreventiveMeasureFor piri:fireSpreadingIncident*
- *piri:smokeDetector piri:isPreventiveMeasureFor piri:smolderingIncident*

Following this retrieval, the user can research the particular text passages, that were annotated with these concepts.

3.2 Adaptation of Textual Passages

Text snippets can be adapted by specification or generalization of entities according to the ontological classification hierarchy. Additionally the corpus aims to support the mining of transformation knowledge. For instance, to find a process how measures suitable for certain scenarios are automatically adaptable to new and unknown incidents. An example for adaptational capacity using the hierarchical structuring of incidents and measures bases on the following textual snippet.

Example 1. "In general, the fire containment approach is preferred, since it emphasizes passive protection and thus the protection of safety systems does not depend on the operation of a fixed fire extinguishing system." [9].

It contains the entities *passive protection*, *safety systems*, and *fixed fire extinguishing system*. The entity *passive protection* has a related entity *active protection*. The entity *fixed fire extinguishing system* has the related entity *mobile fire extinguishing system*. Those two are related in an appropriate manner. We can check the correct adaptation by simply querying for cases, where a *mobile fire fighting system* is used for *active protection*.

3.3 Generation of Document Plots

Usually, documents are created by first sketching a document plot. Here, graph-based and case-based data management complement to each other. We assume that a document about *maintenance of fire safety systems* has to be written, which also means to create a new case for this scenario [14]. In our corpus there exist two documents, one for fire safety in general and one for maintenance of power plants. In a first step, the ontological models associated to the documents are extracted and united using a convenient unification strategy [22]. The new model then contains a relevant set of entities and relations with corresponding information units. In a next step, the PIRI-snippets for the new document are generated automatically. Measures that target the same incident are accumulated into one PIRI-snippet. Afterwards, the document plot is presented to the user. The user now adapts and revises the generated plot. If needed, further user support is available with a case-based query for similar information units. For each component of the maintenance requirements special documents may give deeper advice than available in the two merged documents. We selected an interesting textual passage for fire safety containing advice for maintenance in the following example.

Example 2. "The inspection, maintenance and testing programme should cover the following fire protection measures:
—fire barrier closures such as fire doors and fire dampers;
—fire detection and alarm systems, including flammable gas detectors;
—emergency lighting systems;
—water based fire extinguishing systems;

—a water supply system including a water source and distribution pipe;
—gaseous and dry powder fire extinguishing systems;" [9]

For instance, the maintenance of the *water supply system* is in the focus of the fire safety document. The corpus contains a document subjected to the *radioactive contamination of water* which provides helpful suggestions for the safe maintenance of the water supply system. In this manner, noteworthy measures and incidents can be added to complete the generated document plot.

3.4 Results

The use cases gave a first outline how case-based reasoning can provide a natural representation for covering experience knowledge. They showed that especially the concept of similarity-based retrieval is suitable to handle incomplete input for the formulation and retrieval of experience knowledge. What exceeded the capacity of this work was to show how background knowledge could be integrated into the process in a case-based manner. Nevertheless, it was suggested how strategies basing on the similarity of free text can be intuitively integrated into a structured case-based architecture.

4 Conclusions

This paper introduced an approach for constructing a corpus exploiting graph-based, case-based, and textual information in the domain of nuclear safety regulations. It showed aspects of research work necessary in this field. A proposal for the case-based structuring of a collection of similar documents and text snippets, respectively, was made. Finally selected use cases showed how the corpus can be used practically and will be beneficial for further research work in the domain.

4.1 Related Work

An approach for information retrieval using nuclear safety ontologies was presented by Ogure et. al. [27]. Bouchet and Eichenbaum-Voline [15] presented some early work on case-based search in experience feedback reports from nuclear power plants. Ahmad et al. [10] pick out the issue that frequent changes in daily life induce an increasing amount and heterogeneity of safety-related data. Grabmair et al. [19] presented first results of a feasibility experiment to annotate documents on sub-sentence level with the goal of ranked document retrieval in a certain medical law domain.

The idea to use patterns for relation extraction dates back to Hearst [20] and was evolved by many following authors. A work by Krug et al. [25] using augmented rule-based relation extraction combining active learning with supervised strategies has inspired the presented way of corpus construction. Fundamentals for the use of case-based bootstrapping in document indexing was presented early by Brüninghaus and Ashley [16, 17]. These approaches are well known but came into focus again meeting state-of-the-art computational and natural language processing capacity.

In this work a simplified data structure is used to assess the content of an entire document. A similar problem statement was handled by Caro-Martinez et al. [18] to use case-based strategies in an environment of incomplete information to find explanatory examples in recommender systems. An approach for the semi-automated proof of correctness of case solving strategies in the domain of German law was presented by Beck et al. [13]. This inspired aspects of this work, such as it showed capacities of a rule-based approach in a similar use case.

4.2 Future Work

For future work the annotation will be extended to more documents of the domain provided by other authorities and countries. We expect benefits from this extension, such as the chance to build a multilingual textual case-based corpus [14]. Basic applications will be evaluated more thoroughly on a broader data basis. The goal is to exploit the tacit corpus knowledge to develop more sophisticated applications. Incorporating more NLP technologies is expected to further facilitate text generation in this and similar knowledge intensive domains. Other researches are welcome to use the presented corpus for their work and to contribute to extensions of it.

4.3 Acknowledgments

We wish to thank the International Atomic Energy Agency (IAEA) for their support and the consent to use their publications as a base for this work [5].

References

1. Autorité de Sûreté Nucléaire: https://www.asn.fr
2. Bundesamt für Strahlenschutz: https://www.bfs.de
3. Canadian Nuclear Safety Commission: https://www.cnsc-ccsn.gc.ca
4. Consejo de Seguridad Nuclear: https://www.csn.es
5. International Atomic Energy Agency: https://www.iaea.org
6. Ispettorato Nazionale per la Sicurezza Nucleare: https://www.isinucleare.it
7. U.K. Office for Nuclear Regulation: https://http://www.onr.org.uk
8. U.S. Nuclear Regulatory Commission: https://www.nrc.gov
9. Fire safety in the operation of nuclear power plants: safety guide. International Atomic Energy Agency, Vienna (2000)
10. Ahmad, J., Kostov, B., Lalis, A., Kremen, P.: Ontological foundation of hazards and risks in stamp. Semantic Web Journal (to appear) (2018)
11. Bach, K., Althoff, K.D.: Developing case-based reasoning applications using myCBR3. In: Agudo, B.D., Watson, I. (eds.) Case-Based Reasoning Research and Development. pp. 17–31. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
12. Baumeister, J., Reutelshoefer, J., Puppe, F.: KnowWE: A semantic wiki for knowledge engineering. Applied Intelligence **35**(3), 323–344 (2011)

13. Beck, P.D., Ifland, M., Kronbach, J., Puppe, F., Schenke, R.: Semi-automatische Korrektur von juristischen Lösungsskizzen. Zeitschrift für Didaktik der Rechtswissenschaft **3**, 242–251 (01 2016)
14. Bergmann, R.: Experience Management. Springer, Berlin, Heidelberg (2002)
15. Bouchet, J.L., Eichenbaum-Voline, C.: Case-based reasoning techniques applied to operation experience feedback in nuclear power plants. In: Smith, I., Faltings, B. (eds.) Advances in Case-Based Reasoning. pp. 497–511. Springer Berlin Heidelberg, Berlin, Heidelberg (1996)
16. Brüninghaus, S., Ashley, K.D.: Bootstrapping case base development with annotated case summaries. In: Althoff, K.D., Bergmann, R., Branting, L. (eds.) Case-Based Reasoning Research and Development. pp. 59–73. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)
17. Brüninghaus, S., Ashley, K.D.: Reasoning with textual cases. In: Muñoz-Ávila, H., Ricci, F. (eds.) Case-Based Reasoning Research and Development. pp. 137–151. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
18. Caro-Martinez, M., Recio-Garcia, J.A., Jimenez-Diaz, G.: An algorithm independent case-based explanation approach for recommender systems using interaction graphs. In: Bach, K., Marling, C. (eds.) Case-Based Reasoning Research and Development. pp. 17–32. Springer International Publishing, Cham (2019)
19. Grabmair, M., Ashley, K.D., Chen, R., Sureshkumar, P., Wang, C., Nyberg, E., Walker, V.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. p. 69–78. ICAIL '15, Association for Computing Machinery, New York, NY, USA (2015)
20. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: COLING 1992 Volume 2: The 15th International Conference on Computational Linguistics. pp. 539–545 (1992)
21. Korger, A., Baumeister, J.: GitHub: https://github.com/regdoc/piri
22. Korger, A., Baumeister, J.: The SECCO ontology for the retrieval and generation of security concepts. In: Cox, M.T., Funk, P., Begum, S. (eds.) ICCBR. Lecture Notes in Computer Science, vol. 11156, pp. 186–201. Springer (2018)
23. Korger, A., Baumeister, J.: Case-based retrieval and adaptation of regulatory documents and their context. In: LWDA Berlin. pp. 292–303 (2019)
24. Korger, A., Baumeister, J.: Case-based generation of regulatory documents and their semantic relatedness. In: Arei, K., Kapoor, S., Bhatia, R. (eds.) Future of Information and Communication Conference San Francisco. Advances in Information and Communication, vol. 1130, pp. 91–110. Springer (2020)
25. Krug, M., Reger, I., Jannidis, F., Weimer, L., Madarász, N., Puppe, F.: Overcoming data sparsity for relation detection in german novels. In: DHd 2017 Digital Humanities: multimedial & multimodal, Montreal. pp. 490–493 (2017)
26. Moreau, L., Groth, P.: Provenance: An Introduction to PROV. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan and Claypool (2013)
27. Ogure, T., Furuta, K.: Information retrieval using ontology for sharing knowledge on safety. In: Spitzer, C., Schmocker, U., Dang, V.N. (eds.) Probabilistic Safety Assessment and Management. pp. 531–536. Springer London, London (2004)
28. W3C: SKOS Simple Knowledge Organization System Reference: http://www.w3.org/TR/skos-reference (August 2009)