Rule-based Semantic Relation Extraction in Regulatory Documents

Andreas Korger[1] and Joachim Baumeister[1]

University of Würzburg, Am Hubland, D-97074 Würzburg
a.korger@informatik.uni-wuerzburg.de
joba@uni-wuerzburg.de

Abstract. Regulatory documents are present in many domains of daily living, technical, business, and political context. The knowledge underlying such documents is most often structured using semantic concepts from narrow to broad scope. Those concepts are immanent to the text making up a document. Narrow semantic concepts are described by some words or sentences. Semantic concepts of a broader sense are more complex in their textual representation. This work gives examples of textual characteristics of semantic concepts in the domain of nuclear safety, and that of public events. It shows a rule-based approach for the handling of these concepts and extracting the relations between them.

Keywords: Knowledge Management · Semantic Concept Extraction · Relation Extraction · Ontology Population · Natural Language Processing · Machine Learning.

1 Introduction

In Japanese language many figurative terms exist that stand for broader semantic concepts. For instance the term *shuden o nogasu* means "Being stranded somewhere having missed the last train". The aggregated consequences of this situation are pretty clear. The stranded person has to find alternative transportation home or shelter for the night, get something to eat, inform relatives about the situation, probably postpone dates the next morning, etc. Yet the semantic relations of necessary reactions may differ dependent of the context and the affected person. The same phenomenon is present in many regulatory domains. For instance, the semantic concepts of a *fire alarm* and a *tire change*. In general the concepts are clear but the specific characteristics can be very different. A fire alarm in a nuclear power plant is different from such in a school, and a tire change on a car is different from such on a harvesting machine. When concepts and relations of this kind are used in textual documents special attention has to be payed to their semantics. The lack of clear definitions can lead to misconception. Vice versa, when semantic concepts have to be extracted or assessed in a given text, information may be missing. Often broad concepts cannot be described in full textual detail. Some things are left to understanding by prior knowledge. How deep a semantic concept is or should be elaborated depends for instance on the context, the kind of audience, the available resources, and the way of human computer interaction. Named entities consist of only some words. Semantic concepts may be composed by whole textual passages. They can be distributed across the document or more than one document. Their semantic bases not only in the language but also the document structure as well as in human prior knowledge. Even so a clear perception of such concepts is present to the affiliated audience. These facts make it difficult to define and compare broad concepts solely on a textual basis and to extract relations between them out of a given text.

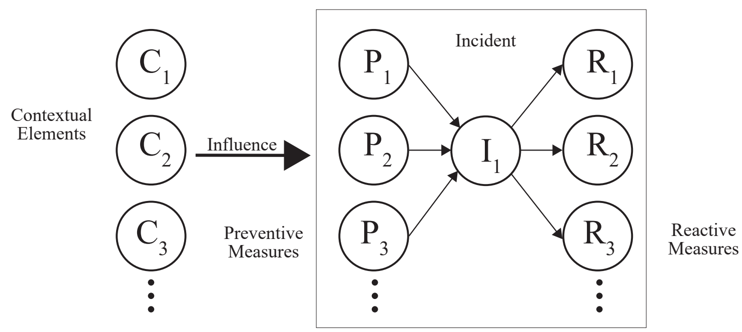

Fig. 1: Measure-Incident-Context scheme to represent semantic concepts together with their relations (PIRI-Scheme) [15].

The presented approach to extract and structure information from available *regulatory documents* enables semantic search, comparison, adaptation, helps to reveal tacit knowledge from textual data, and subsequently facilitate the knowledge management in the domain [10]. We assume that the information about featured entities in a text together with their relations and textual evidence holds as a description. Therefore the regulatory scenario is broken down into the semantic concepts of *relevant incidents* and *connected measures* as depicted in Figure 1. Both can be structured hierarchically from broader to narrower scope. Additionally, incidents and measures are related to one another depending on different contexts. These characteristics can be exploited to transform available information into a graph-based representation. Previous works for this were presented by Korger and Baumeister [14, 15]. The paper is organized as follows: In Section 2, we show different aspects of a rule-based approach for information extraction of semantic concepts and their relations. The further structuring of the extracted information into an ontology is described in Section 3. In Section 4, we discuss the results of the processing architecture applied to documents of the domain of nuclear safety. The paper is concluded with related work, future work and acknowledgments in Section 5.

2 Rule-based Approach for Relation Extraction

Whilst manually analyzing and annotating regulatory documents textual patterns apparently stood out. A big part of semantic information could be made available at acceptable efforts by the manual extraction of textual rules exploiting these textual patterns. Subsequently the decision to build an information extraction system on the base of phrase matching rules was evident. The presented architecture is used for the tasks of named entity recognition (NER), semantic concept extraction (SCE), and relation extraction (RE) [13]. Classical rule-based information extraction systems show a good performance [13]. Therefore this approach is intended to hold as an evaluation benchmark for further work of extracting information from regulatory documents. From the current point of our research a single technique will not be capable of achieving the best results but we expect a combination of different methods to be superior. In the following, we describe a selection of suitable methods.

2.1 Architecture and Preprocessing

In the scenario a basic set of semantic concepts manually defined by domain experts is already available. These concepts are enriched by labels, sometimes descriptions and annotated textual examples. The elements in focus of extraction are named entities, semantic concepts, and relations between them. The SKOS standard (Simple Knowledge Organization System) [23] is used to represent these elements with their hierarchical relations. On the processing side we use the open source library spaCy [12] for natural language operations and Apache UIMA [6] for visualization and annotation standardization.

An overview of the processing steps that are applied can be seen in Figure 2. The following sections describe the particular steps in more detail.

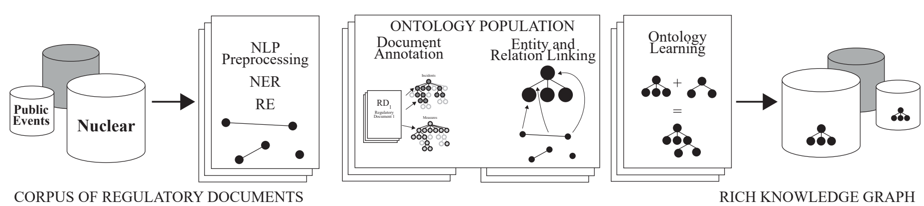

Fig. 2: Natural language processing pipeline that starts with a corpus of regulatory documents (safety instructions, compliance documents, manuals). Via various steps of information extraction and processing a corpus with rich semantic annotation is created. The original data is enriched by semantic information that can be exploited to support users in their search and decisions for instance by recommending and linking related semantic concepts described in other parts of the corpus.

2.2 Named Entity and Semantic Concept Recognition

For the task of named entity recognition there exist well proven methods [13]. The task of semantic concept extraction from text is more challenging. The difference between both is that named entities consist of some words but semantic concepts may consist of larger textual passages, may be composed from sub-concepts, and may be distributed in their textual representation. Initially we start to search for all known entities by matching their textual labels with phrase matching rules considering basic natural language operations like stop word removal and stemming. A gazetteer list is generated from the existing ontology labels and matched against the corpus. For instance the incident *fire* is defined in the ontology by the following code listing 1 in turtle syntax.

```
piri:fireIncident a piri:Incident ;
 piri:inScheme   piri:nuclearSafetyIncidents ;
 piri:broader    piri:incidentRootNuclearSafety ;
 piri:prefLabel  "fire"@en ;
 piri:altLabel   "fire incident"@en, "burning incident"@en ;
 piri:definition "Incidents caused by fire."@en .
```

Listing 1. Semantic concept of a fire incident.

Note that the PIRI ontology [15] aligns a number of sub-classes and subproperties of SKOS classes and properties (e.g. piri:prefLabel, piri:broader). Further exemplary instances of incidents and measures are listed in Example 1.

(1) "Exemplary entities of incidents are: fire incident, starting fire, spreading fire, fire alarm, burning vehicle, and technical failure. Exemplary measures are for instance: fire watch, fire briefing, reduce fire load, fire detection, and prohibit smoking." [2].

In the next step the available examples can be used to extract similar textual passages. Therefore, metrics are necessary that hold as a similarity measure and allow for the comparison of textual passages. One of the most popular approaches is to compare term frequencies of relevant domain vocabulary. More sophisticated approaches use vector representations of text for this task. The following Example 2 shows a textual elaboration for the measure *reduce fire load*. In can be observed that the phrase *"a risk of"* points to a following incident. Some examples of other indicating textual rules can be seen in Table 1. The mentioned score is a metric to assess how good the indication of a rule is, the higher the better. The scoring is part of the knowledge engineering process whilst handcrafting patterns. The recognition of concepts featured by a text is essential for the following efforts of finding relations between entities.

(2) "Procedures should be established for the purpose of ensuring that amounts of combustible materials (the fire load) and the numbers of ignition sources be minimized in areas containing items important to safety and in adjacent areas that may present a risk of exposure to fire for items important to safety." [2].

Table 1. Examples for patterns indicating incidents.

Pattern	Entity	Score
"a risk of X"	incident	3
"protection from X"	incident	3
"can result in X"	incident	1
"prevention of X"	incident	3
"X caused by"	incident	1

2.3 Rule-based Relation Extraction

Relation extraction is the task to automatically find relations between semantic concepts in a text [13]. We are interested in relations between the semantic concepts of *incidents* and *measures*. The following Table 2 shows these relations followed by the code listing 2 that gives an example how they are represented in turtle syntax followed by a textual example of this annotation.

Table 2. Examples for relations between semantic concepts of incidents and measures.

Relation Type	Range
"hasMeasure"	incident *hasMeasure* measure
"hasPreventiveMeasure"	incident *hasPreventiveMeasure* measure
"hasReactiveMeasure"	incident *hasReactiveMeasure* measure
"narrower"	incident *narrower* incident, measure *narrower* measure
"broader"	incident *broader* incident, measure *broader* measure

(3) "Detecting and extinguishing quickly those fires which do start, thus limiting the damage;" [2].

```
piri:hasAnnotation
  [piri:hasIncident piri:fireIncident;
   piri:hasReactiveMeasure   piri:detectingMeasure,
                             piri:extinguishingMeasure,
                             piri:mitigatingMeasure ] .
```

Listing 2. Annotation with semantic concept.

To reveal such relations there exist different approaches that exploit for instance co-occurrence of entities, textual patterns, and syntactical patterns. These approaches face difficulties that are increased by the present scenario of having to handle generalized semantic concepts. The approach of co-occurrence is likely to have a high recall but will surely lack in precision. The usage of semantic patterns will be more precise but comes at the costs of creating the rules and will be most likely very domain dependent. For an approach of domain independent relation extraction we suggest a combination of different methods.

Lexical Analysis Using Phrase Patterns We previously extracted known semantic concepts using the gazetteer approach and candidates for semantic concepts using patterns. We use this preparatory work now for a co-occurrence analysis. We search the text for paragraphs where entities appear together; i.e., in the same sentence, the same paragraph, and the same chapter. This will discover relations from the type *hasMeasure* but can not distinguish from the sub-types *hasPreventiveMeasure* and *hasReactiveMeasure*. In this manner, it is also not decidable which of the co-occurring incidents and measures is the narrower and broader semantic concept, respectively. This problem is mitigated in some way by applying textual patterns that identify certain relations and classify the extracted textual passages. Examples of such indicating triggers are shown in Table 3.

Table 3. Examples of triggers indicating relations.

Pattern	Relation	Score
"protection from X"	hasMeasure	3
"prevention of X"	hasPreventiveMeasure	3
"examine for X"	hasMeasure	1
"to detect X"	hasMeasure	2

Syntactical Analysis using Dependency Structure Additionally to the lexical phrase matching rules, characteristics of speech can be exploited. This is for instance necessary when concepts are spread over longer passages and we need to discover which entity is actually in scope (co-reference analysis). Syntactical information is used to find negations, e.g. a certain measure is not suitable for an incident. This information supports the classification process. An exemplary dependency graph can be seen in Figure 3

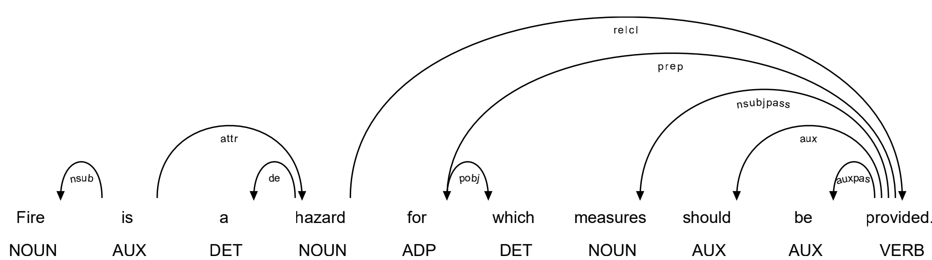

Fig. 3: Dependency graph that shows how a verb together with its auxiliary verb points to a measure to which an incident is syntactically related to. This example could hold as a rule as the incident *fire* and the *measure* could be replaced by their broader class.

Text Analysis using Document Structure We observed that the document structure is crucial for the assessment of relations between semantic concepts mentioned in a document. Sometimes the reference between entities can only be derived by taking for instance document title and headlines into account. Additionally, the textual analysis of such structural elements demands further considerations. The circumstances are clarified by the following Example 4.

(4) In the corpus of nuclear safety documents there exist two documents, one with the title *"Fire Safety in the Operation of Nuclear Power Plants"* and another having the title *"Protection against Internal Hazards other than Fires and Explosions in the Design of Nuclear Power Plants"* [2].

Only from the title it can be derived that measures mentioned in the first document are most likely related to a fire incident. The second title indicates that the document does not treat fire and explosion incidents but all other internal incidents which emphasizes the relation to the according measures mentioned in this document. It can also be observed in this example that there is the necessity of syntactical analysis. Just searching for mentioned entities would not reveal that the document excludes some of them by negation.

The next example shows how textual evidence for relational information can be found in a headline.

(5) The document for nuclear fire safety contains a section headed with *"FIRE PREVENTION AND FIRE PROTECTION"* [2].

From this headline it can be concluded that the following measures are preventive measures for a fire incident representing a more special relation between those entities. Without considering the headline it would barely be possible to classify the holding relations only out of the textual description of the measures mentioned in the chapter.

Additionally to relational information between incidents and measures the document structure encodes even more. Hierarchical information classifying incidents and measures into broader and narrower concepts can be extracted. This is obviously founded in the fact that most often documents are structured hierarchically. For instance all incidents mentioned in the document of nuclear fire safety are related to the fire incident most likely with a narrower scope.

3 Ontology Engineering Using Extracted Relations

Part of the initial data was a domain ontology basing on the SKOS standard providing semantic concepts and their relations engineered by domain experts. For applications like semantic search, document retrieval, and document generation we need a knowledge graph that codes the semantic information representing regulatory documents. Therefore, we now use the previously described methods and unite them in a coherent approach to enrich the existing domain ontology (ontology population) and give suggestions to improve it (ontology learning) [5].

The semantification of documents allows for the facilitated access to knowledge contained in documents, supports decisions and creation of new regulatory documents. For instance it is given a certain context for a new fire safety document. Depending on that context, relevant textual passages contained in existing regulatory safety documents can be proposed.

Ontology population is the process of creating and linking instances of existing ontology classes (by annotation) according to real life evidence, in our scenario textual evidence. This evidence is beneficial in two ways. First the instances of existing semantic concepts and proofs for their relations are created and are thus available for further usage. Second, the extracted data can be used to enrich and improve the existing domain ontology. This is done by suggesting new classes of semantic concepts derived from the discovered previously unknown entities and their relations. We have to decide whether an extracted semantic concept is actually new and how to integrate it into the available hierarchical structure. This task is currently left to domain experts. A strategy to support domain experts in that work is to present confidence values derived from the extraction rules and similarity metrics. The following algorithm shows the processing steps in pseudo code.

Algorithm 1: Algorithm for co-occurrence and pattern-based relation extraction from regulatory documents.

Data: Semi-Annotated corpus $\mathcal{A}$, Domain Ontology $\mathcal{O}$

Result: Set of new annotations for corpus $\mathcal{MA}$ with confidence level for each annotation. Set of new instances for $\mathcal{O}$ depending on confidence threshold.

Preprocess corpus (tokenize, lemmatize, part-of-speech-tagging);
Retrieve entity labels from $\mathcal{O}$ and map gazetteer list to $\mathcal{A}$;
Extend found matches with dependency rules;
Apply set of NER rules to $\mathcal{A}$;
Score NER matches;
Refine found matches with dependency rules;
Select all sentences with co-occurrence of incidents and measures;
Apply set of RE rules to $\mathcal{A}$;
Score RE matches;
Remove redundant annotations;
Construct new ontology with retrieved machine annotations higher than a selected confidence value;

---

4 Case Study

To evaluate the presented approach we applied different processing pipelines on a corpus of nuclear safety documents [2, 16]. The corpus consists of 140 documents with about 10.000 pages. The rules where extracted out of about 4000 manually labeled instances of incidents and measures. For the task of manual annotation we use the text highlighting and annotation environment ATHEN [17]. This work on creating a gold standard corpus is still in progress [24]. With ongoing evaluation efforts we manually add annotations as trustworthy standard annotations verified by human intelligence for the automatic evaluation of machine annotations. To have a coherent evaluation approach respecting the currently available quality of gold data we choose to compare the pipeline performances on the classification of whole sentences. A true positive match is found if the whole sentence contains at least the beginning of a gold standard entity. How well matched entities overlap with gold entities is currently neglected.

As expected, the matching of gazetteer lists is limited and performance differs strongly depending on the evaluated document but if available the information can be used. An interesting aspect of the gazetteer matching is how to interpret matches. For instance the example *"fire safety"* which represents a measure but also contains the incident *fire*. If we accept this as a revealed incident the gazetteer approach achieves a fairly high precision. If we want the system not to count *"fire safety"* as an incident because it represents a measure then the precision is really poor and drops. In general we observe that the classes *incident* and *measure* are overlapping and difficult to separate. Better results could be achieved with lexical rules to identify sentences that contain relevant semantic concepts as well as relational information. To extract fine grained information the application of methods relying on syntactical information was necessary. Syntactical patterns improved the recall at the cost of lower precision. For the joint approach we observe that the classification of sentences is sometimes contradictory. A strategy has to be applied to decide which classification has to be preferred. The analysis of the machine labeled documents showed that still many patterns are not covered by the current set of rules which leaves room for improvement. Facts about the performance are presented in the following Table 4.

Table 4. Performance of the approach analyzing the document *Fire Safety in the Operation of Nuclear Power Plants.*

Methods	Precision	Recall	F1
Entity extraction *"incident"*	0.78	0.76	0.77
Entity extraction *"measure"*	0.72	0.89	0.79
Relation extraction by co-occurrence	0.83	0.88	0.85

Additionally to the evaluation on the corpus of nuclear safety documents we started to transfer the processing pipelines to German language and applied them to a corpus of safety documents for public events in Germany like the Oktoberfest in Munich [14]. A German language model was used for natural language operations, a gazetteer list was created out of available German ontology labels, and the rules where manually crafted. First feasibility experiments are promising that the approach is transferable to different languages and different domains but the work is still too preliminary to present sound performance results.

5 Conclusions

This paper introduced an approach for the extraction of relational knowledge contained in regulatory documents. The strategy to describe semantic concepts by connected entities and a selection of their relations turned out to be useful for the access of textual information. A step by step presentation of convenient information extraction techniques was elaborated. With suitable textual examples from the domain of nuclear safety we explained different aspects of textual characteristics. It was shown how the extracted information can be mapped to an existing ontological structure for further usage. Finally, we presented the results of a case study using a corpus of regulatory documents in the domain of nuclear safety.

5.1 Related Work

The idea of mapping a textual description to a word that stands for a likewise semantic concept (reverse dictionary lookup) was given attention to by Hill et al. [11]. They presented a strategy to exploit these definitions for the semantification of a connected concept. In the present scenario a clear textual definition is often not available. Nevertheless, if it is possible to extract definitions distributed in textual data the idea of Hill et al. [11] can be picked up to support the semantification of concepts of broader scope. Having structured and semantified information available that holds as descriptions of concepts enables for their comparison and adaptation. This also helps to reveal tacit knowledge from textual data and subsequently facilitate the knowledge management in the domain. The extraction of complex events and their relations from science literature using diverse NLP methods was presented by Barik et al. [4]. An ontology and rule-based approach was presented in the thesis of Najihme Mousavi [19]. Even though dealing with the extraction of semantic concepts from speech similar obstacles were addressed by Ghannay et al. [9]. Atapattu et al. [3] focus on the extraction of semantic concepts as well es their hierarchical relations using classical methods taking into consideration the document structure. They provide a sound approach for the evaluation of the extraction efforts. To extract named entities and relations between them out of descriptions of music albums Oramas et al. [21] presented a rule-based approach. They address available methods and show up possibilities and limitations of state of the art relation extraction methods. The importance of the document structure for the information extraction task was emphasized by Furth and Baumeister [7]. An interesting rule-based approach was presented by Sadikin and Wasito [22]. They divide entities in main objects and supportive objects. We go a bit further by exploiting the whole available hierarchy of entities. A rule-based approach incorporating dependency parsing to simplify rules and distant supervision to automatically discover relations was presented by Garcia et al. [8]. Kamel Nebhi [20] uses the support of structural data to enhance the performance of his syntactical relation extraction approach. The concept of extracting textual passages that contain relations and then classifying them (distant supervision) was presented by Mintz et al. [18].

5.2 Future Work

Several tasks may be considered for future work. The most necessary is to improve the evaluation by extending the gold standard corpus especially to regulatory documents from different origin. Semi-supervised relation extraction approaches should come into focus as the manual hand crafting of rules is effective but at high costs. Methods that automatically learn rules should also be considered. Aspects of data augmentation are surely interesting to investigate as this would allow for the usage of neural approaches.

5.3 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 [1].

References

1. International Atomic Energy Association: https://www.iaea.org
2. Fire safety in the operation of nuclear power plants: safety guide. International Atomic Energy Agency, Vienna (2000)
3. Atapattu, T., Falkner, K., Falkner, N.: Automated extraction of semantic concepts from semi-structured data: Supporting computer-based education through the analysis of lecture notes. In: Liddle, S.W., Schewe, K.D., Tjoa, A.M., Zhou, X. (eds.) Database and Expert Systems Applications. pp. 161–175. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
4. Barik, B., Marsi, E., Öztürk, P.: Extracting causal relations among complex events in natural science literature. In: Frasincar, F., Ittoo, A., Nguyen, L.M., Métais, E. (eds.) Natural Language Processing and Information Systems. pp. 131–137. Springer International Publishing, Cham (2017)
5. Buitelaar, P., Cimiano, P. (eds.): Ontology Learning and Population: Bridging the Gap between Text and Knowledge, Frontiers in Artificial Intelligence and Applications, vol. 167. IOS Press (2008)
6. Ferrucci, D., Lally, A., Verspoor, K., Nyberg, E.: Unstructured information management architecture (UIMA) version 1.0. OASIS Standard (mar 2009)
7. Furth, S., Baumeister, J.: Semantification of large corpora of technical documentation. In: Atzmueller, M., Oussena, S., Roth-Berghofer, T. (eds.) Enterprise Big Data Engineering, Analytics, and Management. IGI Global (2016)
8. Garcia, Marcos, Gamallo, P.: A weakly-supervised rule-based approach for relation extraction. In: Jose A. Lozano, J.A.G., Pérez, J.A.M. (eds.) XIV Conference of the Spanish Association for Artificial Intelligence. Springer (2011)
9. Ghannay, S., Caubrière, A., Estève, Y., Camelin, N., Simonnet, E., Laurent, A., Morin, E.: End-to-end named entity and semantic concept extraction from speech. In: 2018 IEEE Spoken Language Technology Workshop (SLT). pp. 692–699 (2018)
10. Guha, R., McCool, R., Miller, E.: Semantic search. In: Twelfth International World Wide Web Conference (WWW 2003) (2003)
11. Hill, F., Cho, K., Korhonen, A., Bengio, Y.: Learning to understand phrases by embedding the dictionary. Transactions of the Association for Computational Linguistics **4**, 17–30 (2016)

12. Honnibal, M., Montani, I., Van Landeghem, S., Boyd, A.: spaCy: Industrial-strength Natural Language Processing in Python (2020)
13. Jurafsky, D., Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall PTR, USA, 1st edn. (2000)
14. 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)
15. 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)
16. Korger, A., Baumeister, J.: Construction of a corpus for the evaluation of textual case-based reasoning architectures. In: LWDA Bonn (2020)
17. Krug, M., Tu, N.D.T., Weimer, L., Reger, I., Konle, L., Jannidis, F., Puppe, F.: Annotation and beyond - using athen annotation and text highlighting environment. In: Digital Humanities (2018)
18. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2. p. 1003–1011. ACL '09, Association for Computational Linguistics, USA (2009)
19. Mousavi Nejad, N., Jabat, P., Nedelchev, R., Scerri, S., Graux, D.: Establishing a strong baseline for privacy policy classification. In: Hölbl, M., Rannenberg, K., Welzer, T. (eds.) ICT Systems Security and Privacy Protection. pp. 370–383. Springer International Publishing, Cham (2020)
20. Nebhi, K.: A rule-based relation extraction system using dbpedia and syntactic parsing. In: Proceedings of the 2013th International Conference on NLP & DBpedia - Volume 1064. p. 74–79. NLP-DBPEDIA'13, CEUR-WS.org, Aachen, DEU (2013)
21. Oramas, S., Sordo, M., Espinosa-Anke, L.: A rule-based approach to extracting relations from music tidbits. In: Proceedings of the 24th International Conference on World Wide Web. p. 661–666. WWW '15 Companion, Association for Computing Machinery, New York, NY, USA (2015)
22. Sadikin, M., Wasito, I.: A novel rule based approach for entity relations extraction. Journal of Theoretical and Applied Information Technology **74**(2), 249–261 (Jan 2015)
23. W3C: SKOS Simple Knowledge Organization System Reference: http://www.w3.org/TR/skos-reference (August 2009)
24. Wißler, L., Almashraee, M., Monett, D., Paschke, A.: The gold standard in corpus annotation. In: IEEE GSC Passau (2014)