MINING LEGISLATION: AN ANALYSIS OF LEGAL AND TECHNICAL IMPLICATIONS

Anna-Sophie Novak / Chibuzor Udokwu / Charalampos Alexopoulos / Michalis Loutsaris / Shefali Virkar

Research Associate, Danube University Krems, Department for E-Governance and Administration
Dr. Karl-Dorrek-Straße 30, 3500 Krems, AT
anna-sophie.novak@donau-uni.ac.at

Research Associate, Danube University Krems, Department for E-Governance and Administration
Dr. Karl-Dorrek-Straße 30, 3500 Krems, AT
cjobuzor@gmail.com

Postdoctoral Researcher, University of the Aegean, Department of Information and Communication Systems Engineering
Voulgaroktonou 30, 11472 Athens, GRC
alexop@aegean.gr

PhD Candidate, University of the Aegean, Department of Information and Communication Systems Engineering
Karlovassi 832 00, Samos, GRC
mloutsaris@aegean.gr

Research Associate, Danube University Krems, Department for E-Governance and Administration
Dr. Karl-Dorrek-Straße 30, 3500 Krems, AT
shefali.virkar@donau-uni.ac.at

Keywords: ***Text Mining, Legislation, Legislative Data, Copyright Law, Database Protection Law***

Abstract: *Text mining is the process by which information and connections are obtained from large amounts of text using algorithms. ManyLaws is a project that seeks to increase the accessibility of legal information by offering innovative services. Although several text mining techniques for extracting and storing information exist, an in-depth analysis of the legal and technical issues regarding text mining of legislative information systems is still missing. This paper identifies and examines the legal and technical implications associated with the mining of legislative data, with a focus on copyright and database law. Possible solutions to mitigate the identified issues are also discussed.*

1. Introduction

The aim of the ManyLaws project is to deliver a novel set of services for citizens, businesses and administrations of the European Union, built upon text mining, advanced processing and semantic analysis of legal information. To that end big legal data will need to be accessed, which is currently produced and published in multiple national or EU public databases (e.g. RIS, EUR-Lex). Those datasets will need to be extracted, linked and transformed into a structured relational, open database to prepare them for the mining process.

With regard to the legal field possible applications include the research through legal corpora, analyzing the alignment of national legislation with EU legislation, comparing national laws which target the same life events, analyzing the references to European legislation by national laws, analyzing related laws within the same Member State, timeline analysis for all legal acts, visualization of the progress and current status of a specific national or European piece of legislation and sentiment analysis towards new legislation.[1]

[1] Those are the services proposed by the ManyLaws project. The University of the Aegean, the Hellenic Parliament, Intrasoft, Danube University Krems and the Austrian Parliament are the Consortium of the ManyLaws project. The project runs for two years and is financed under the Connecting Europe Facility (CEF). The aim of this project is to deliver a set of novel services for citizens,

For the proposed applications, the standard Text Analysis pipeline performs several levels of analysis: morphological, syntactic, semantic, and discourse (Lacity/Janson 1994). The morphological and syntactic analysis is usually performed with a syntactic parser (Lagos et al., 2017), which recognizes the syntactic word classes such as nouns and verbs, and the syntactic dependency structure of the constituents of a sentence (main verb, subject, object, etc.). In general language processing, recognizing the basic semantic roles of a sentence constituents, i.e., the «who», «does what», «where», «when», and «how» constituents, is a well-established task for English. Co-reference resolution is identifying when two mentions of an entity or event refer to the same underlying person, place, thing or event in the real world. A layered approach supporting the data flow, from source data to visualized outputs will handle the large volumes of data. When it comes to the information processing layer, various text mining algorithms are applied in different processing tasks, relying on a supercomputing infrastructure. Service-oriented intermediate results are: Creation of reverse indexing, occurrence and frequency tables for millions of words, creation of various n-grams for the identification of important terms or phrases, semantic comparison of different law sets (e.g. EU Directive against national legal framework) performing full word-level and document-to-document comparison for billions of pages. Therefore, the power of thousands of processors is needed.

Although several methods and techniques for executing text mining processes exist, a detailed analysis on the implications and limitations of text mining, especially on legislative information systems, is still missing. Legislative information is usually stored in national databases, whose contents and outputs vary. Therefore, there is a gap in identifying and addressing legal and technical issues in executing text mining on legislative databases. To address this gap, this paper is guided by the main research question: What are the legal and technical implications in mining legislative information? Sub-questions, further derived from the main research question, are as follows: What are the existing information systems for storing legal information? What technical approaches can be used in addressing the identified issues?

The remainder of this paper is structured as follows: Section 2 discusses existing legislative databases within the context of the ManyLaws project and the selection of the databases that are relevant for this paper. Section 3 describes how the text mining process of legislative information is designed. Section 4 identifies the legal and technical implications of text mining and Section 5 provides a conclusion of the study.

2. Existing Legislative Databases

To identify the existing databases for storing legislative information, the following steps were completed. First, we performed a systematic search to identify databases that store legal information including public and private databases. Once this was completed, we consulted with legal experts to identify the main private databases. However, only legislative databases of the European Union, Austria and Greece were selected for this study since those are the pilot cases for the ManyLaws project. The results can be seen in Table 1.

businesses and administrations of the European Union, built upon text mining, advanced processing and semantic analysis of legal information. To that end we will access big legal data.

Platform	Country	Output type	Content type	Access
RIS (Legal Information System Austria)	Austria	HTML, PDF, RTF	Federal and State Legislation and Case Law	Free
Austrian Parliament	Austria	rss, xml	Government Bills, Legislative Initiatives, Explanatory Materials o	Free
RDB	Austria		Commentary, Doctrine, Legislation	Full access is not free
Lexis-Nexis	Austria		Commentary, Doctrine, Legislation	Full access is not free
Hellenic Parliament	Greece		Bills, Proposals, Legislation	Free
Greek National Printing House	Greece	Mainly PDF	Laws, Presidential Decrees, regulatory acts "	Free
NOMOS	Greece		Doctrine and Commentary, Law, Regulations	Full access is not free
EUR-Lex	Europe	HTML	Legislation, Case-Law, Treaties	Free

In the following sections, the mining process of the governmental legislative databases will be described theoretically. The analyses performed in these sections focus on governmental legislative databases as they provide the necessary data via open data portals. In addition, the legal and technical discussions focus on the Austrian databases.

3. The Text Mining Process

In order to design the appropriate text mining process, it is deemed necessary to first identify the service layer. Services are to be provided towards citizens, businesses and administrations, based on the most common needs of each user type. Through a user interface supporting simplification or advanced usage, these are some of the services to be provided at real time by the ManyLaws project: Research through EU member-state legal corpora, analyzing the alignment of national legislation with EU legislation, comparing national laws which target the same life events, analyzing the references to European legislation by national laws, analyzing related laws within the same Member State or with different EU-States, timeline analysis for all legal acts, visualization of the progress and current status of a specific national or European piece of legislation, visualizations of correlations, dependencies and conflicts between different laws and sentiment analysis towards new legislation.

Based on the above services and sub-products, a variety of add-on services can be developed after capturing new requirements from citizens, businesses and administrations. The proposed system consists of three service components:

User Generated Queries: The proposed system infrastructure stores search queries made by users for analysis and optimization purposes. The terms and structure that make up the user query are feeding into the semantic search engine to enhance the relevance of the results based on inferred concepts and semantic annotations.

Search relevant content based on queries: The search engine is used for searching through the system's triple store using a scalable Solr-based semantic search engine. Furthermore, the search engine is taking into account semantic relations between search terms and stored entities (e.g. synonyms). Best practices such as faceted search are also used to present the user with more search options, relevant to the search terms by semantic association.

Search results: This component is responsible for retrieving and presenting to the user the search results in an efficient and user-friendly manner.

3.1. Design of the Text Mining Process

The information processing layer of the architecture deals with the text mining process of the identified legislative databases. These are the four necessary steps:

Data Preparation and Translation Services: In this stage, data from the identified legislative databases is acquired and prepared for the text mining tools to follow. This stage includes data reading and initial cleansing, semantic annotation and formulation for processing. Due to the diversified origin of the texts acquired, a large amount of effort and computational power is devoted to Optical Character Recognition (OCR) and translation

in English. Translation is based on EuroVoc and automated translation services, both for the complete texts but also for the various indexes and n-grams to be created at the processing stage.

Text Mining: Various algorithms are being applied in different processing tasks, relying on a super-computing infrastructure, in order to produce service-oriented intermediate results as described above.

Structured Data: This component represents the information collected from various sources and adhering to a common model and format that can then be used more effectively by the Visual Analytics Service. The data stored include any harvested and derived information that is necessary to realize the project's use cases.

Visual Analytics Service: The Visual Analytics Service provides the ability to access the entire data transformation pipeline from raw or semantic data to interactive visual representations. The main goal is to enable user-centred and comprehensible solutions for getting insights and knowledge about the entire domain.

4. Legal and Technical Issues in Mining Legislative Databases

After we have discussed the text mining approach in the sections above, we will continue below to identify the technical and legal issues in mining the governmental databases.

4.1. Technical Implications of the Text Mining Process

Table 1 shows that there are seven databases that store legislative data in the locations that were chosen for the ManyLaws pilot project. We consider the following as the technical issues when mining legislative text in Austrian databases. These include availability and accessibility of data, content type and content output type.

4.1.1. Accessibility and Availability of Data

The accessibility of information from the data sources depends on the data output of the data sources. Our analysis shows that, although most of the legislative data sources provide full access to the output data, this requires payment of a subscription service to access this information. For the other legislative databases that do not provide full access to their output information, it is therefore not possible to collect complete information from such data sources. The availability of text to be mined depends on the service level agreements entered with the data providers.

4.1.2. Content Type

The content type refers to the type of legal information contained in the output data from the data sources. As shown in table 1, our study shows that some of that data sources output data contain both legislative data and case law, while the others contain only legislative data. It is necessary to identify the type of legal information stored in each data source before a meaningful data analysis can be carried out on such data.[2] To analyze data, it is necessary to process the aggregated texts into structured data by classifying and organizing them into various groups and columns. In this case, if the content types are not properly identified in the processing stage, information generated from such analyses will be incorrect.

4.1.3. Content Output Type

As shown in table 1, all the databases analysed use different formats for displaying legal information to the users. There are technical implications related to the format of the law texts under examination in the different databases. For instance, in the process of transforming of legal documents to plain text (this applies mainly for information stored in the PDF format), the following issues can be experienced: identifications of columns in the law text, images included in the document, dashes used in continuing a word to the next row, types of article

[2] Talib, Ramzan, et al. Text mining: techniques, applications and issues. International Journal of Advanced Computer Science Applications 1.7 (2016): p. 417.

separation, information not related to law content, and character replacement step necessary for eliminating problems of using English language characters similar to other languages.[3]

4.1.4. Mitigation Methods for the identified Issues

To reduce loss of information when converting PDF documents to text, the original XML files can be requested from the database publishers before converting to PDF. This is a possible solution because of the limited number of databases involved. Another way to reduce the loss of information and maintain data integrity is to properly outline the contexts of the legislation documents and incorporate natural language processing capabilities of artificial intelligence (AI) tools.[4] For the content types, it is necessary to properly structure the information to be mined by properly categorizing the text outputs into various groups and columns before performing data mining analyses. The issue with the accessibility of data can easily be addressed by mining only databases that provide full access to the content in order to guarantee the authenticity of information generated.

4.2. Legal Implications of the Text Mining Process

Austria currently has no provisions specifically for the text mining process. Other countries have introduced explicit laws to ensure legal certainty. Germany regulates in section 60d German Copyright Act[5] that copying and publishing of protected works is admissible for the text mining process performed for non-commercial research purposes. The corpus may be published and shared with a distinguishable group of persons for joint research efforts or with third parties in order to review the research quality.

In 2010, then Prime Minister David Cameron commissioned a review of UK's intellectual property rights fearing those might not be suitable to enable innovation and growth.[6] One of the results was an exception from copyright law in Section 29A Copyright, Designs and Patents Act 1988, which states that copying of a work for text mining purposes is no infringement of copyright as long as the person has lawful access to the work. Furthermore, the sole intention of copying has to be research for non-commercial purposes and lastly it is to be accompanied, if possible, by a sufficient acknowledgement. Furthermore, a contractual exclusion of this right is unenforceable and there is no provision for remuneration. The copyright holders are entitled to adopt technical provisions, since the text mining process could overload their servers.

In 2016, France introduced two exceptions for text mining from its Intellectual Property Code (Code de la propriété intellectuelle). Article L122-5 paragraph 10[7] states that the author cannot prohibit the copying for research purposes from sources a person has lawfully access to, if the work has been rightfully published. Commercial research is excluded from this exception as well. Article L342-3 paragraph 5[8] covers the exception for databases. It states that the entitled person cannot prohibit the copying of the database for non-commercial research purposes by a person whose access to the database is lawful, should the database have been rightfully

---

3 Rajman, Martin, and Romaric Besançon. Text mining: natural language techniques and text mining applications. Data mining and reverse engineering. Springer, Boston, MA, 1998. 50–64.

4 Berg, Øyvind Raddum. High precision text extraction from PDF documents. MS thesis. 2011. p. 51–63.

5 § 60d German Copyright Act, Text and Data Mining, dejure.org/gesetze/UrhG/60d.html (accessed on 6 January 2019).

6 Hargreaves, Digital Opportunity. A Review of Intellectual Property and Growth, assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf (accessed on 5 January 2019), 2011.

7 Article L122-5 10° Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000037388886&cidTexte=LEGITEXT000006069414&dateTexte=20190106 (accessed on 6 January 2019).

8 Article L342-3 5° Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000033219347&cidTexte=LEGITEXT000006069414&dateTexte=20190106 (downloaded on 6 January 2019).

published. Any contractual clause contrary to this Article is void. In the course of the application of this exception, the normal operation of the database must not be interfered with.

Since Austria currently has no specific text mining exception, the extraction of legislative information from ris.bka.gv.at and parlament.gv.at will be investigated concentrating on copyright law and the legal protection of databases.

4.2.1. Copyright Law

The Austrian Copyright Act[9] protects individual and intellectual creations in the fields of literature, sound art, fine arts and film art. In the context of the text mining process, it depends firstly on the text mining technique used and secondly on the selected texts whether there is an infringement of copyrights of third parties. Most text mining techniques, as does the proposed technique above, rely heavily on copying the texts for them to be analyzed and annotated. The act of copying protected texts is covered by section 15 of the Copyright Act. It states that solely the author has the right to copy his or her work. The right to copy is understood broadly and includes even the technically required automated copying.[10] Therefore, the described mining process would be unlawful, if the used texts are protected and no exception applies. There are exceptions to Copyright Law that could find application to the text mining process. These exceptions are discussed as follows.

The **research exception**[11] includes solely non-commercial research. Further, the researcher has to attribute the source of the used data, if this is feasible. It is questionable whether a company's in-house research is also included in this exception.[12] Organizational structure and finances are not relevant for determining whether the research is commercial or non-commercial. Relevant is the research purpose.[13]

The **temporary copies exception**[14] allows the copying if it is an essential and integral part of a technical process and itself not of economic importance. As the name of the exception suggests, the copies are allowed to be saved temporarily as long as the technical process requires them.[15] The establishment of a permanent corpus with the original elements is therefore not possible under this exception.[16]

The **text mining exception** adopted by the European Parliament in the directive on copyright[17], states an exception to the reproduction right (copyright law) and the extraction right (*sui generis* database law) of the author if those were made for scientific research purposes, by research organizations that had lawful access to the texts. Unlike the exceptions mentioned above, the EU-exception does not require non-commercial research. However, further specifics on how to interpret the term «research organizations» are missing for now. According to a briefing requested by the JURI committee: «By ‹research organizations›, it is in fact intended universities, research institutes, non-profit or public interest research-intensive organizations.»[18] Public-Private partner-

---

9 Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz), BGBl. I 2018/63.

10 Dillenz/Gutmann, Praxiskommentar zum Urheberrecht. Österreichisches Urheberrechtsgesetz & Verwertungsgesellschaftsgesetz^2, Springer Verlag, Wien 2004, § 14 Rz 22.

11 § 42 (2) Urheberrechtsgesetz.

12 Zemann, in: Kuscko/Handig, urheber.recht. systematischer kommentar zum urheberrechtsgesetz^2, Manz Verlag, Wien 2018, § 42 Rz 23.

13 Dillenz/Gutmann, UrhG und VerwGesG^2, § 42 Rz 10.

14 § 41a Urheberrechtsgesetz.

15 Truyens/Van Ecke, Legal aspects of text mining, lrec-conf.org/proceedings/lrec2014/pdf/452_Paper.pdf. (accessed 2 January 2019).

16 Truyens/Van Ecke, Legal aspects of text mining.

17 Art 3 Text and Data Mining, Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0337+0+DOC+XML+V0//EN (accessed 5 January 2019).

18 Briefing requested by the JURI committee, European Parliament, europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf (accessed 2 January 2019), p. 8.

ships are not left out, as long as the commercial interests are not the decisive influence and control over the research organization.[19] Member states of the EU are required to implement this exception into their legislation and contracts that contradict the exception are unenforceable. Some criticize the unequal treatment of non-research organizations who have lawful access to texts.[20] Their text mining processes are not covered by the EU-exception.

As we have now discussed the general application of copyright law and the relevant exceptions on the text mining process, the following section will give an assessment of the applicable laws when mining Austrian legislative databases.

Article 2 paragraph 4 Berne Convention[21] states that «It shall be a matter for legislation in the countries of the Union to determine the protection to be granted to official texts of a legislative, administrative and legal nature, and to official translations of such texts.» According to section 7 paragraph 1 Austrian Copy Rights Act, laws, regulations, official decrees, notices and court judgements do not fall under the Copyright Act. Since the copying of legislative data falls under this exception, copyright law is not applicable to the extraction of legislative information. Concerning the explanatory materials from the website of the Austrian parliament one has to make sure that those are also covered by section 7 paragraph 1 Copyright Act. Explanatory Materials are of considerable importance for interpreting the law. Furthermore, they are attributable to the parliament, an authority with public authority tasks. As a result, copyright law is not applicable to explanatory materials, which can be subsumed as official notices.[22]

4.2.2. Database Law

Should the texts for the mining process be part of a database, copyright law as well as the *sui generis* database law have to be considered. Those two fields of law are independent of each other and can apply to the same database. According to the 17th recital of the directive on the legal protection of databases, the term «database» should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data, collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed. According to the European Court of Justice, databases «in any form», whether they are in electronic or non-electronic format, are covered by the directive.[23] Below we will discuss general database protection and then go on to our specific goal of mining governmental legislative databases.

Copyright Protection of Databases

The copyright protection of databases does not extend to their contents. The selection or arrangement of the contents mark the author's own intellectual creation and are protected as such by copyright. No temporary or permanent reproduction by any means and in any form, in whole or in part can be made without the consent of the author[24]. Art 6 paragraph 2 of the directive allows the member states to make certain exceptions. In Austria, section 42 paragraph 2 and 40 h paragraph 2 Copyright Act state that the reproduction for non-

[19] Briefing requested by the JURI committee, European Parliament, p. 8.

[20] Provision Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules. Part B Exceptions and Limitations Chapter 1 Text and Data Mining, ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf (accessed 6 January 2018), p. 4.

[21] Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979) (Authentic text), wipolex.wipo.int/en/text/283693 (accessed 20 December 2018), p. 4.

[22] Katzenberger, Die Frage des urheberrechtlichen Schutzes amtlicher Werke, GRUR 1972, p. 692.

[23] Decision of ECJ of 9 November 2004 C-203/02 *British Horseracing Ltd*, Slg 2004, I-10415, Rz 5.

[24] Art 5 lit a Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal of the European Communities 1996 L 77/20.

commercial research purposes is admissible. It is doubtful that the text mining process includes the extraction of the protected selection and arrangement of the contents, as the mining process concentrates on the content.

***Sui Generis* Protection of Databases**

Section 76c of the Austrian Copyright Act protects investments made in databases. According to Article 7 of the directive, the maker of a database who has made qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents has the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of that database.[25] The extraction and re-utilization of an insubstantial part of such a database is therefore allowed, as long as they are not executed repeatedly and systematically. According to § 76d paragraph 3 number 2 Copyright Act, the extraction and re-utilization of a substantial part of a publicly available database is lawful for non-commercial research purposes, to a justifiable extent, as long as the source is named.

As we have now discussed the general application of database law on the text mining process, the following section will give an assessment of the applicable laws when mining Austrian legislative databases.

As neither the selection nor the arrangement of the contents of the Austrian legislative databases mark an intellectual creation, those are not protected as such by copyright protection of databases. Nonetheless, in the imprints on the website ris.bka.gv.at one will find a copyright notice in favor of the Federal Ministry of Digitalization and Economy and on the website parlament.gv.at one will find a copyright notice in favor of the Austrian Parliament. Concerning the *sui generis* protection of databases, one has to check if a substantial investment in either the obtaining, verification or presentation of the contents was made.[26] A qualitatively and quantitatively substantial investment was made especially in the presentation of the contents of the Austrian legislative databases. According to the CJEU, the investment going into the creation of the single entries is not taken into account.[27] According to the Austrian Supreme Court, the definition of an investment does not depend on whether the data is given to the maker or if law prescribes the presentation of the data.[28] Rather those expenses that run into the presentation and the updating of the database content are to be considered as investment.[29] Therefore, the maker of the websites ris.bka.gv.at and parlament.gv.at have the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of that database via the website. An application of section 7 Copyright Act on protected databases by analogy was denied by the Austrian Supreme Court.[30] Since both websites offer some of their data on the national open data portal (data.gv.at) this right does not constitute an insurmountable barrier for the mining process.

5. Conclusion

In this paper, we analyzed and discussed legal and technical implications experienced in mining legislative information. The legislative databases in Greece, Austria and Europe were chosen, as those are the pilots for the ManyLaws project. The ManyLaws project seeks to provide innovative services to increase accessibility of legal information. To address the gap identified by the main research question of this study, we systematically identified databases that store legal information and described text mining procedures for mining them. Then

25 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal of the European Communities 1996 L 77/20.

26 § 76c Urheberrechtsgesetz.

27 Decision of ECJ of 9 November 2004 C-203/02 *British Horseracing Ltd*, Slg 2004, I-10415, Rz 40 «However, such prior checks are made at the stage of creating the list for the race in question. They thus constitute investment in the creation of data and not in the verification of the contents of the database.».

28 OGH 12 June 2007, 4 Ob 11/07g, ÖBl 2007/65 S 291 (Dittrich) - ÖBl 2007,291 (Dittrich) = ecolex 2007/332 S 783 (Schumacher) - ecolex 2007,783 (Schumacher) = MR 2007,384 = Burgstaller, MR 2008,15 = jusIT 2008/43 S 94 (Mader) - jusIT 2008,94 (Mader) = RdW 2008/109 S 147 - RdW 2008,147 = SZ 2007/95 = Thiede/Schacherreiter, JBl 2015,287.

29 OGH 12 June 2007, 4 Ob 11/07g.

30 OGH 9 April 2002, 4 Ob 17/02g.

we analyzed the properties of these databases and discussed technical and legal implications that could arise in mining these databases.

The main limitation of this paper is that the study focuses only on the technical and legal implications of text mining legislative databases in relation to the ManyLaws project. This limits the scope of the paper. Also, as a result of the page limitation of this paper, the study does not fully describe the necessary steps in addressing the technical issues identified in this paper.

6. References

Heyer, Gerhard/Quasthoff, Uwe/Wittig, Thomas, Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse, W3L AG 2006.

German Copyright Act, Text and Data Mining, dejure.org/gesetze/UrhG/60d.html (accessed on 6 January 2019).

Hargreaves, Ian, Digital Opportunity. A Review of Intellectual Property and Growth, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf (accessed on 5 January 2019), 2011.

Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000037388886&cidTexte=LEGITEXT000006069414&dateTexte=20190106 (accessed on 6 January 2019).

Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz), BGBl. I 2018/63.

Dillenz, Walter/Gutmann, Daniel, Praxiskommentar zum Urheberrecht. Österreichisches Urheberrechtsgesetz & Verwertungsgesellschaftsgesetz^2, Springer Verlag, Wien 2004.

Zemann, Adolf, in: Kuscko, Guido/Handig, Christian, urheber.recht. systematischer kommentar zum urheberrechtsgesetz^2, Manz Verlag, Wien 2018.

Truyens, Maarten/Van Ecke, Patrick, Legal aspects of text mining, http://lrec-conf.org/proceedings/lrec2014/pdf/452_Paper.pdf (accessed 2 January 2019).

Art 3 Text and Data Mining, Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), http://europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0337+0+DOC+XML+V0//EN (accessed 5 January 2019).

Briefing requested by the JURI committee, European Parliament, http://europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf (accessed 2 January 2019).

Provision Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules. Part B Exceptions and Limitations Chapter 1 Text and Data Mining, http://ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf (accessed 6 January 2018).

Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979) (Authentic text), http://wipolex.wipo.int/en/text/283693 (accessed 20 December 2018).

Katzenberger, Paul, Die Frage des urheberrechtlichen Schutzes amtlicher Werke, GRUR 1972.

Lagos, N., Gallé, M., & Chernov, A. (2017). U.S. Patent Application No. 14/850,060.

Lacity, M. C., & Janson, M. A. (1994). Understanding qualitative data: A framework of text analysis methods. Journal of Management Information Systems, 11(2), 137–155.

Baroni, M., Bernardi, R., and Zamparelli, R. (2014). Frege in space: A program for compositional distributional semantics. Linguistic Issues in Language Technology, 9.

França, M. V., Zaverucha, G., & Garcez, A. S. D. A. (2014). Fast relational learning using bottom clause propositionalization with artificial neural networks. Machine learning, 94(1), 81–104. Brunschwig, Colette (2001): Visualisierung von Rechtsnormen, Legal Design, Zürcher Studien zur Rechtsgeschichte, hg. von M. T. Fögen [u.a.], Zürich: Schulthess Juristische Medien.

Röhl, K. F., & Ulbrich, S. (2007). Recht anschaulich: Visualisierung in der Juristenausbildung (Vol. 3). Herbert von Halem Verlag.

Talib, Ramzan, et al. Text mining: techniques, applications and issues. International Journal of Advanced Computer Science Applications 1.7 (2016): p. 417.

Rajman, Martin, and Romaric Besançon. Text mining: natural language techniques and text mining applications. Data mining and reverse engineering. Springer, Boston, MA, 1998. 50–64.

Berg, Øyvind Raddum. High precision text extraction from PDF documents. MS thesis. 2011. p. 51–63.