THESIS ON INFORMATICS AND SYSTEM ENGINEERING C85

An Automated Legal Content Capture and Visualisation Method

ERMO TÄKS

TUT
PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Information Technology
Department of Informatics

Dissertation was accepted for the defence of the degree of Doctor of Philosophy in Engineering on April 18, 2013

Supervisors: Professor Dr. Rein Kuusik
Department of Informatics
Tallinn University of Technology

Professor Katrin Merike Nyman-Metcalf
Tallinn Law School
Faculty of Social Sciences
Tallinn University of Technology

Opponents: Professor Burkhard Schafer
School of Law
University of Edinburgh

Professor John Zeleznikow
Faculty of Business and Law
Victoria University in Melbourne

Defence of the thesis: June 06, 2013

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for any academic degree.

/Ermo Täks/

ISSN 1406-4731
ISBN 978-9949-23-472-1 (publication)
ISBN 978-9949-23-473-8 (PDF)

INFORMAATIKA JA SÜSTEEMITEHNIKA C85

Automaatne meetod õigusteabe hankimiseks ja visualiseerimiseks

ERMO TÄKS

TTÜ
KIRJASTUS

1 Contents

2 List of Publications

Liiv, A. Vedeshin, E. Täks, Visualization and structure analysis of legislative acts: a case study on the law of obligations. Proceedings of the 11th international conference on Artificial intelligence and law, Stanford, 2007.

E. Täks, A. Lohk, An alternative method for computerized legal text restructuring., in Frontiers in Artificial Intelligence and Applications. Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference, 2010.

E. Täks, L. Vohandu, A. Lohk, I. Liiv, An alternative method for computerized normative system, in ICAIL 2011.

E. Täks, L. Võhandu, A. Lohk, K. Nyman-Metcalf, A. Rull, A tool for exploring the hidden structure of legislation, in Proceedings of the fundamental concepts and the systematisation of law. Workshop at Jurix 2011 in Vienna, 2011.

K. Nyman-Metcalf, E. Täks, Simplifying the law - can ICT help us?, International Journal of Law and Information Technology, Oxford Journals, 2013; doi: 10.1093/ijlit/eat003

2.1 Author`s Contribution to the Publications

In abovementioned last four articles, main author was the author of this thesis.

The author's role was:

- performing a general analysis of on-going law simplification activities and IT related research
- developing and testing the main method of legal text capture and visualization,
- developing and testing the legal text similarity measurement method together with inexact graph matching method application;
- developing and testing the legislation restructuring approach.

3 Introduction

"Mastering the lawless science of our law,
That codeless myriad of precedent,
That wilderness of single instances,
Through which a few, by wit or fortune led,
May beat a pathway out to wealth and fame."
Alfred, Lord Tennyson, Aylmer's Field (1793)

A modern civilized society, like a functioning free market, develops spontaneously and is tremendously complex. Development of communication systems, increasing population and globalisation of markets create more and more opportunities and ways to communicate, but also brings along lots of risks and problems. In order to deal with such problems legislation is of great help. Mankind has historically developed a system of rules to deal with such kind of complexity. A set of written agreements has been used to reduce uncertainty and the possible negative impact such uncertainty may entail.

The legal system has many positive features, but inevitably has its weaknesses to deal with. Over time the amount of legal documents has increased a great deal. Just collecting and systematising a huge number of nationally developed legal documents might be quite a challenge. In addition, these acts tend to confront each other occasionally and this is bringing in a new type of confusion- a legal system internal one. These problems are qualitatively at a new level in the globally interconnected world of today and solving such problems needs an input from highly skilled experts. Due to globalisation, possible misunderstandings between participants having different cultural and traditional backgrounds pose special challenges linked to the growth of additional, supranational regulation. Such regulation strongly influences national legislation and has to be adopted by it.

However one thing remains unchanging- people desire to understand and relate somehow to the surrounding society. The human being is "constructed" in such way: the need for social security, to belong to something and be a part of something bigger is inevitable. Legislation should be the best place to start with but this citadel has high walls and strong barriers for newcomers. One has to know a lot in order to deal with the legislation. One has to know the principles of the build-up of the

legislation, the specific language, the way of applying the legal knowledge within specific situations, the way how to support argumentation based on legal findings. Too often this is too much for non-experts and their attempts to rely on the written law results in failure. John Sheridan from The National Archives, United Kingdom, who has been responsible for developing an electronic legal acts database, briefly commented in his invited speech during Jurix 2010 conference in Liverpool- these people will hurt themselves, when they do try to understand something about written law. The problem is that a very big part of any society is not educated enough in order not to hurt themselves in case of legal dispute. [1]

This notion opens for us actually a major problem buried somewhere inside the society. *Ignoratio juris non excusat* - ignorance of the law excuses no-one. By definition, a law is obligatory for all and should be followed by all the members of the society. Yet the specific answers about specific problems lay somewhere in the huge pile of documents, which is constantly growing and changing. How to guarantee that any member of society will get at least access to the right legal source when it is needed, not talking about the understanding and proper application of it?

We are not the first humans in the long history of society to face this question. There have been several ways found in the past how to deal with such shortcomings. Mainly it is built upon legal advisers who can even get governmental support for helping others in this delicate matter. Mostly the compliance with law rests upon the conscience of ordinary citizens and most of the time it works well. Common sense tells people to keep away from courts and lawyers as it is usually very expensive and very time consuming to deal with them. Often the justice is blind and in case of unavoidable court processes the outcome is unpredictable. This way the legal help turns out to be one of the last options instead of being the main and first one.

Interestingly the biggest problem here is not for citizens or legal experts. Both sides have developed sophisticated “survival kits” to overcome these shortcomings. The real problem is there for the government administration who tries to influence the development of the society. Legislation is created in order to regulate the relationship between the different parties of the society. How to influence their behaviour if the legislation is weakly implemented, the message is not received by the target audience and expected changes are not achieved? How to inject this vital part of information into the minds of millions of people who have limited

motivation for receiving it? How to achieve a real, 100% compliance with law in real life with the tools public servants have today at their disposition?

This doctoral thesis is dedicated to the abovementioned, not very well known aspect of the public administration. Namely, to the aspect of the public servant who has a responsibility to communicate the will of the government to the broad masses, to ensure that this message will be heard by each and every society member, loud and clear. This contains administrative and technological aspects. According to K. Merusk and R. Narits, two influential legal professors from the University of Tartu, **the justice is a normative communication and information system** [2]. This functionally ICT oriented approach raised an idea which forms the basis for the rest of the thesis. If we are talking about communication and information systems, then it is possible to apply methods and tools upon the legislation that have been well developed for computerised information systems. What if we try to analyse and manage the legal system, using an information engineering approach, would we benefit from that?

Due to the personal experience gained through actual work experience and thanks to the quick development of Estonian e-Government solutions during 1995-2001, this gave the author of this thesis a good overview of the problem and provided a most challenging research question, which the author has tried to solve ever since. Being an ordinary member of society, a professional politician, an information technology related project manager in the Ministry of Justice, a member of the Punishment Register Act drafting committee, IT manager of Communication Board and finally, a high level IT related public administrator in the Estonian State Chancellery, the author is familiar with quite some different aspects related to this field. This has provided with the motivation and also the responsibility to deal with this matter in a new, information technology enabled way.

3.1 Legislative informatics: setting the scene

Attempts to improve the legislation with help of information communication technology (ICT) solutions have a relatively long history. The importance of the subject, the amount of work invested into creation of the legislation, need for ultimate quality etc. have attracted the attention of the good new world avangarde- IT scientists- from the very beginning. The complexity of the problem was discovered soon enough and Artificial Intelligence (AI) type characterisation seemed to suite well to describe the domain. AI and Law is the general term often used to refer into this specific research domain.

According to McCarthy: ”Artificial intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.” [3] Intelligent machines should be able to support the decision-making of legal experts though by personal opinion of the author should never take the place of a human decision maker. This is an important prerequisite of such research- the machine should never decide over human beings but can always assist, thus increasing the human decision-making value and quality.

Legislation as a set of rules can be viewed as a big knowledge base. Knowledge is represented in the form of rules to follow and an ideal AI solution should be able to extract and apply this type of knowledge according to the specific situation. Such solutions are classified as knowledge-based and expert systems and much work have been done in order to understand these systems' potential and limitations. Expert systems derive their “intelligence” from human experts as a set of empirical associations and have to be encoded into machine readable format. This approach has several weaknesses:

- the choice of possible scenarios in the legal field is practically unlimited and therefore the full extent cannot be foreseen by the expert;
- it is difficult for experts to make their knowledge explicit;
- the knowledge is reusable with difficulty;
- systems can not sufficiently explain reasons of the choices made;

Therefore the modelling of the domain of expertise seems more promising. The domain should be conceptualized for modelling and this conceptualisation has to be made explicit. [4] Describing the problem like this by the definition one can call it ontology. [5]

By the definition ontology is a branch of metaphysics concerned with the nature and relations of being. [6] Ontology has a long history dating back to the great Antique Greek philosophers (Parmenidas). This was a philosophical attempt to describe the world and has been found to be a suitable approach in many fields, including computer science. It is based on taxonomy and shows the relation between objects (Figure 1). It is possible to build a solid reasoning solution upon it as it enables the generalisation or specification of the current subject.

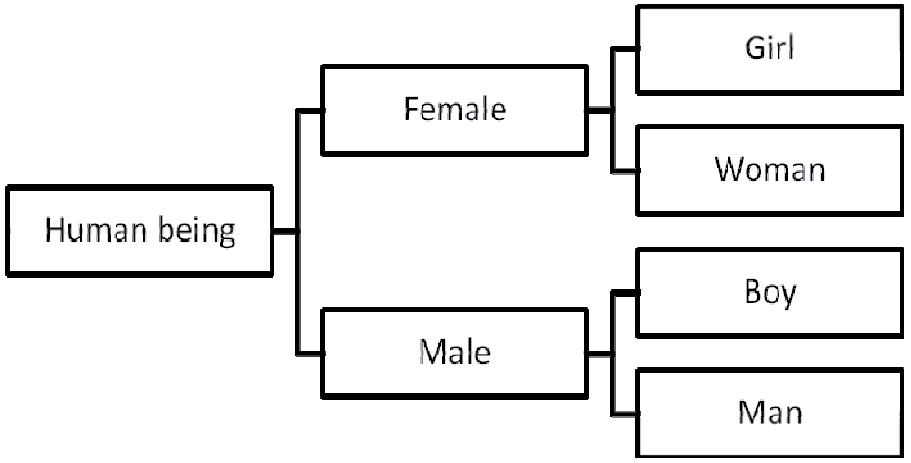


Figure 1. Example of simple ontology

Ontology has been chosen as one of the main tools for developing the next World Wide Web standard- Web 2.0, so it is quite logical that the AI and Law community is taking a step toward this direction. However, despite careful research the well-established connection between origins of the ontology and today's legal domain knowledge representation format was not found. The author finds that the ontological approach has strong practical usage within the field of AI and law but has a weak theoretical foundation.

Modelling of the law domain is an extremely challenging task. This involves many very important questions starting from the very basic, among others- how to make it possible, what method is suitable, how to perform it correctly. Early results of such approach indicated the following problems [4] :

- Legal texts are "open"- there are no adequately presented definitions with necessary and sufficient conditions;
- Legal rules are not static but dynamic. The construction of theory is more important than application of the theory;
- In process of theory construction there is no "right answer";
- Application of a concept to a new situation automatically modifies the concept.

Therefore the modelling of the law domain is twofold: how to capture and implement the legal knowledge in machine readable way and how to deal with the dynamics of the legal domain.

From the functional perspective Valente states that the legal system is used as an instrument to change or influence society in specific directions, determined by social goals. For applying this knowledge he decomposed this function into six primitive legal knowledge's [7]:

- *normative knowledge*: what is forbidden or obligatory;
- *word knowledge*: related to the legal terminology;
- *responsibility knowledge*: estimates responsibility for the violation of norms;
- *reactive knowledge*: related to the sanctions that can be taken;
- *meta-legal knowledge*: deals with how other legal knowledge should be reasoned with;
- *creative knowledge*: which states how items of legal knowledge can come into being and cease to be.

As the knowledge-based approach turns out to become more complex after each next step, a universal and widely accepted approach to the law domain modelling has not been found since, though there are quite some promising developments.

The author of this dissertation has been working in this direction for a couple of years and after describing all the relevant aspects of this approach a visualised picture applied to the local context is presented in Figure 2. Legal space is a term to bring together all the different aspects of such modelling approach. Legal space format deals with machine readable presentation of the legislation; legal space calculation methods are consolidating different knowledge extraction methods; the bottom region of the picture (LS theoretical foundations, LS Ontology) deals with theoretical foundations about accessible methods used for law domain knowledge description so far.

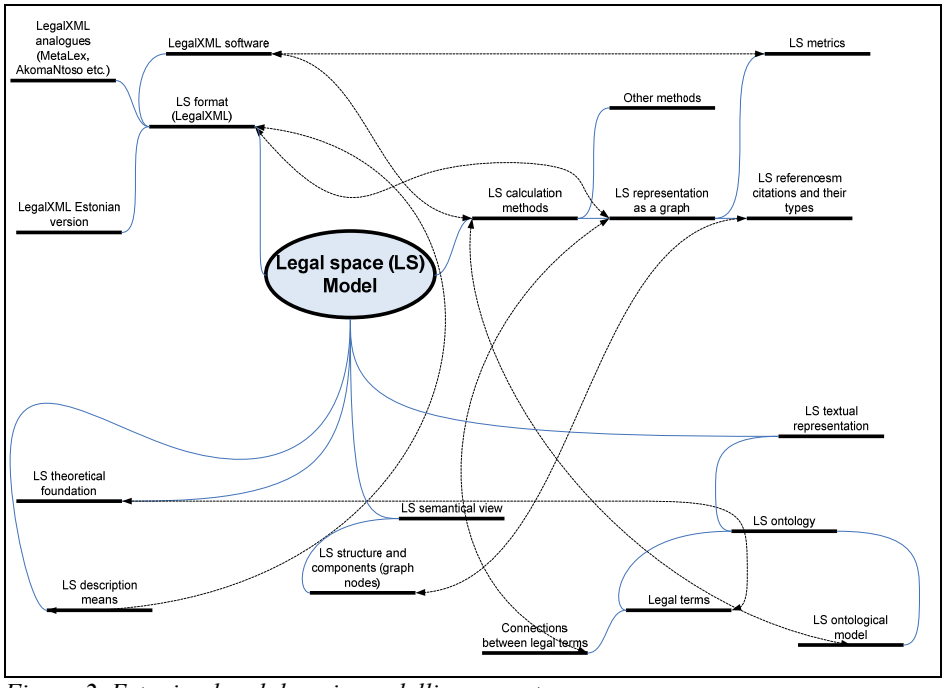


Figure 2. Estonian legal domain modelling aspects

Analysis results of presented approaches may conclude the following: the pathway from written rules into the well described knowledge usable for computers is long and contains many hidden risks. Legal texts are interpreted by humans quite controversially- three lawyers can present four opinions about the same legal question, as someone says. Transforming such widely interpretable content into ontologies is even more questionable. Ontologies are created by humans and no analyst will create exactly the same conceptualisation picture compared to another about the same issue. In order to get the best out of it, these mental pictures should be merged. There is no guarantee that the result will be encompassing and contain no errors. In the broader sense, it means a creation of a new speciality of legal practitioners- legal ontologists, who will "translate" the content of the legal domain for computers as the lawyers do translate the content of the legislation for people. Add one highly subjective layer (legal ontologist interpretation) upon another (legal expert interpretation) and then command computers to apply the result with extreme accuracy- this seems a quite unreliable scenario for the author.

The abovementioned reasons forced to find an alternative approach to the "translating the translation of legislation for computers" in order to achieve more objective results, but in the first stage just to find the location of relevant legal sources among a myriad of documents. Pinpointing the exact location of legal content will open a way for further automated extraction of it without human interference.

3.2 Motivation

Estonia got its sovereignty back in 1991 but being late to jump onto the train of the information revolution gave an unexpected advantage- there were no expensive legacy systems (mainly expensive mainframe solutions) to get in the way. This was the period when Estonia invested a lot into e-Government solutions. Client-server solutions and internet were the hot new keywords of this time and the advantages of such technical solutions showed their potential immediately. As Estonia was (and still is) an unattractive market due to the small size of it, there was a freedom to experiment without pressure from the side of the big international IT companies. During 1995-2001 the author of this thesis was serving in the public sector, responsible for developing several IT systems: prisons information system, bailiff offices information system, probation information system prototype, e-Cabinet solution for Estonian Government.

After the successful launch of the at that time unique Estonian Government e-Cabinet solution (10-th of August 2000) the main infrastructure of e-Government was successfully built and work was targeted toward more substantial issues. The Estonian Government itself was already working in a paperless way, but the ministries and other public offices were not. This created a huge task of scanning the paper documents into electronic ones inside the State Chancellery and quite a logical next step was to try to implement paperless solutions in the whole public sector. The e-Õigus (e-Justice) project team was created between different ministries in order to come up with an electronic document sharing system. Mainly it was meant for publishing projects of legal acts for a broader public and at the same time it could function as a specialised web portal between different agencies.

This gave the author a reason to look more seriously into the Estonian legal system- the previous work for the e-Cabinet was done only during four months and gave no time for such theoretical background research. Having a system analyst view and previous experience working with legislation, several promising new features of the legal system were revealed and

seemed to stress the importance of a network-like set-up of the legal system. Legal documents do not exist independently, but form a complicated set of mutually influencing clusters with no clear border between these clusters. As there is a strong internal logic of such legal build-up and there are clearly defined links between different legal acts, the first objective could be just visualizing these relations hoping that this will become a valuable input by itself. This was the general description of the idea presented to the e-Justice project team.

It quickly turned out that the presented proposal to develop such system in the framework of the e-Justice portal was too much out of scope. Although these ideas were abandoned quickly they served as a valuable input to proceed with doctoral research, which started some years later.

3.3 Theoretical background

"Written laws are like spider's webs;
they will catch, it is true, the weak and the poor,
but would be torn in pieces by the rich and powerful."
Anacharsis, Scythian philosopher (6th century BCE)

One important aspect of legislation has not been researched well enough-normative or legislative references within and between legal texts. There have been several attempts to handle this issue. Palmirani, Birghi and Massini defined in their work a normative reference as "*a linguistic expression that unambiguously identifies a specific act or a text partition referred to*" and created a framework for automated extraction of normative references in legal texts [8]. Winkels and Breebaart specify the reference as follows: "*a reference refers from a source – the referring source – to a(nother) source or a work – the referred source*". In their work they describe a semantic network, which is the document structure graph, visualizing the legal documents as nodes and references as nodes [9]. Some attempts have been trying to handle systems of legal documents, their inner structure and references as graphs, introducing basic graph measures to legislation [10] [11] [12].

References are the integral part of legislation, playing the important role of connecting different legal acts, made in different times and under different circumstances. References allow us to tie together two separated legal acts somehow regulating the same aspect of state or even changing another legal act. All the individual acts and many lower level general acts in Estonia should refer to the higher level legal acts in order to acquire the

needed legislative authority. Therefore it is possible to say that references sometimes transfer the legal power to the legal acts. Yet the simple recognition of the more complicated issue of legal references does not help us much. Plain graph visualizing the structure of legal documents is integral and gives us some hints about the importance of the legal act or the closest related acts, but does not open much from the legal point of view. The gained results are interpretable in many different ways and such kind of graphs give little indication about how to apply this or another linked law to a certain case. Therefore it is useful to enrich such graphs with some metadata in order to visualize more information.

There can be different types of references. Palmirani, Birghi and Massini had estimated only one type of reference, which unambiguously identifies a specific act or a text partition referred to [8]. Bourcier and Mazzega had found that there exist two kinds of links or edges: those corresponding to self or cross quotations and those induced by the codification process [11]. Brief research has shown that there are many possibilities to estimate the type of references: according to the legal acts these connect, legal acts position in the legal acts hierarchy and its structural elements referred and so on. Yet the limitations applied on the domain (will be discussed in the subdivision 0) made such an approach quite unpromising for this research and was left aside.

Previous unsuccessful attempts to “crack” the legal system still gave valuable insights into the problem and the breakthrough in this action research came from an unexpected aspect. Trying to apply different visualisation techniques upon legal texts and using existing language technology tools for this purpose, a system analyst approach seemed to fit well- extracting the objects, processes and events it is possible to present the core of the system under investigation with help of nice visual models or pictures helping IT experts to communicate with potential users. These pictures serve as simple mind maps for persons with non-technical background, provide them with the ability to guide the process in general and still enables system analysts to handle the smallest details in order to create incredibly complicated software. What if to try to apply the same method for the legislation, generate such easy readable layouts from the legal texts and look whether these pictures are able to communicate with non-expert users better than the text itself? If this is a good approach and if existing links between legal acts cannot be used to create an internally fully linked legal system, is it possible to find another, alternative way of referencing?

Solving such questions needed a simultaneous use of different disciplines: understanding the nature of the legislation in order to apply reverse engineering principles; natural language processing techniques to estimate the objects (represented as nouns) and processes/events (represented in texts as verbs). For the sake of clarity the found verbs had to be transformed into basic forms and their closeness established somehow. The frequency of use of found closest noun verb pairs were estimated with help of Excel VBA programming tools. The visualisation tool for generating visual layouts of the legal texts was mainly Google originated Gephi and graph theory was needed in order to estimate some features of such visualisations like closeness or the degree of the edges. The graph like layout opened a way to come out with a new way of measuring the similarity of the legal texts, based on inexact graph matching method and substantially measuring the co-use of the same objects in relation with the same activities within the legal texts. A program for implementing the inexact graph matching technique was created in Excel VBA, also the further data mining and data table seriation methods. A prototype for testing the ideas (http://data.ttu.ee/visuallaw) was implemented in SQL database, using JavaScript Infoviz Toolkit (http://thejit.org).

A short overview of used methods follows.

3.3.1 Reverse engineering

In order to create a new structure for the normative system a smallest, atomic unit had to be found. These legal particles can be used for restructuring the existing system without demolishing it. Alternating views, created in such a way can complement each other, open a number of hidden perspectives and interact successfully.

Reverse engineering is a process of analysing a subject system to:

- Identify the system components and their interrelationships;
- Create representations of the system in another form or at a higher level of abstraction. [13]

The normative system is a collection of norms. A legislative text may be seen as a vehicle that contains and transports rules; and the legal order as an organism of rules rather than that of laws - as a result enabling us to observe its contents better. Rules are, therefore, seen as the true foundations of normative systems, whilst laws are merely temporal. [14] Relevantly a norm can be used as an atomic unit, within the legal system, and also for its structural decomposition. It is interesting to remark that Japanese scientists

have also used a structural decomposition for legal text analysis, but used a word as atomic part of the norm. [15]

3.3.2 Natural language processing

Legal and linguistic aspects of a norm are tightly bound: the norm can be understood as "*thought (i.e. meaning) content expressed through language*". The norm as a rule receives expression in a norm sentence (norm formulation) and vice versa: the norm is a meaning content of the norm formulation. [16] The norms can be found or targeted with the help of linguistic or grammatical constructions. [17]

The clause is a minimal grammatical construction able to deliver the thought content of the norm, containing at least a subject (noun) and a predicate (verb). According to Fillmore's Case grammar ideology there is limited amount of abstract semantic relations between a verb and its nouns. [18] A short analysis of the usage of different types of words in Estonian legislation identified nouns in 47% of the cases and verbs in 13% of the cases, combining for 60% of the words in legal texts. It is important to notice that the informational value of verbs and nouns is significantly higher compared to adjectives and conjunctions, though it is hard to measure. Therefore the rest of the word types were left out from the scope of the current work, despite some very appealing features (negation, specification, role assignment etc.).

3.3.3 Business process modelling (BPM)

Modelling is a common task in BPM that aims to visualize the structure of complex systems. The method allows modelling any system using main activities, events and objects. The method was applied, declaring that verbs within legal text refer to events and activities, nouns to objects. Measuring the concordance - coexistence frequency of verbs (activities and events) and nouns (objects) within sub-clauses - it is possible to create a very rough BPM model of legal text, represented as a graph. This also allows us to compress the analysed legal text up to 53 times, still embedding, by rough estimations, 60-80% of the legal content. [19]

3.3.4 Graph theory

The visual representation of legal text as a graph makes a good formal comparison basis between legal acts and is a key to perform the quantitative analyses of the legal text, using different graph theory aspects and graph

mining methods. A data mining technique was applied to estimate the similarity of different graphs.

The process of evaluating the similarity of two graphs is commonly referred to as graph matching. The overall aim of graph matching is to find a correspondence between the nodes and edges of two graphs that satisfies some, more or less, stringent constraints. That is, by means of the graph matching process similar substructures in one graph are mapped to similar structures in the other graph. [20, p. 219]. Graph isomorphism calculation can be np-complete task, but as the graph edges have names (a basic form of filtered words due to Estonian language peculiarities) the similarity measurement is relatively simple. Therefore it is possible to perform graph structural similarity calculations and this notion forms a basis for further restructuring of legislation.

3.3.5 A deep structural analysis

However, to get an overview of such a huge amount of interconnections is impossible for humans. In order to get a better condensed overview of those similarities several different visualization methods were used and to present a complex connected system, it has been useful to use extreme representation principles. In the current research in the case of Estonian legislation, the principle of the maximal similarity spanning tree was used and for that the computed similarity matrix S of legal acts, the maximal connected path was calculated [21]. A symmetrical matrix of similarities was transformed into 0/1 matrix based on cutting level of lowest maximum similarity (37%) to guarantee get a single connected component of all acts. Last method of data reorganization was a ZODIAC method. [22]

3.4 Research questions

The research questions are divided into logical steps and starting from the top-level:

The main target is to increase the usability of legislation. The basis for this approach is the International usability standard, whereby the usability is defined as follows: ”*The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments*.“ [23]. The same principle can be used for assessing legal systems. Information technology is a discipline where the emphasis is on the word “information” whereby the word “technology” is merely a reference to the tools exploited for this purpose. Therefore the main purpose of this thesis is to find a

suitable information technology originated method for the legislation in order to open new ways to more efficiently manage legal information.

The second major target is to find and use a suitable visualisation in order to deliver the legal content to the non-specialised reader. Visualisation is one of the promising data mining approaches and due to the possibility to represent the different views about same objects, it has a potential to deliver different overviews of any complicated system, including legal systems. Visualisation can be done in different forms and formats depending from the task; therefore the current objective postulated from the perspective of the potential user is to get a better overview of the legislation systematic build-up and to pinpoint the location of the most relevant legal sources.

The third target is to find suitable methods and techniques in order to perform the visualisation approach. There is no benefit in visualisation if it cannot be generated in a trusted way; therefore the chosen visualisation method must be linguistically, legally and mathematically sound, properly applied and suitably adjusted in order to serve estimated purposes. These methods should enable the increased computerized processing of the legal texts to minimize manual work and increase the usability of the outcome. In case the method is new, new additional ways have to be discovered in order to prove the value of the results.

The fourth target is to apply chosen methods and techniques in order to test the chosen methods and get experimental results for further examination. As the result of this work is expected to be as practical as possible, it has to be tested on real-life situations and the gained results presented to a larger audience. This makes a basis for a motivation to build a suitable prototype based on applied methods and gained results. The method, test results and prototypes should be the main proof of quality for the described research.

The fifth target is to get an evaluation from legal experts in order to test the suitability of the chosen methods according to requirements of the legal domain. As the author of this thesis does not have qualified legal education but the research is targeting the legal domain, the results should be evaluated by legal specialists. Their critical input would be one of the main results of this research and one of the quality criteria to apply.

3.5 References to previously published works

Several chapters in the thesis are based on previously published works. First article, published in "International Journal of Law and Information Technology*",* provides the reader with a general overview of the state of the

art in the field of law and technology, described in Chapter 5 [24]. This is the last part of the general analysis. The rest of the document describes the original research performed by the author in the frame of his doctoral study. Chapter 6 is based on delivering novel results of the research about use of the existing references within legal acts; a graph-like layout derived from it and is described with the help of some basic graph theory based measurements. The results were published at the ACM "International Conference on Artificial Intelligence and Law" [10]. A visualisation method described in Chapter 6, based on verb-noun concordance and first experimental results were published in the conference "The 23 rd. International Conference on Legal Knowledge and Information Systems", Jurix 2010 [19]. Estonian legislation related structural and experimental research results described in Chapter 7 were published in the "24-th International Conference on Legal Knowledge and Information Systems", Jurix 2011 [25] and another publication was published in the same conference in the frame of the workshop "Proceedings of the fundamental concepts and the systematisation of law". [26]. This last article described some results of the first analysis made by legal experts and presented some new considerations about the legislation structure used in Estonia.

3.6 Organization of the dissertation

This chapter presented the general introduction to legislation visualisation and highlighted the main motivations for undertaking the research presented in this dissertation. The rest of the dissertation is organized as follows.

A very short overview about two main areas of this interdisciplinary research, legislation and information communication technology, are provided in *Chapter* 4. The purpose of it is to create a common ground for further reading as the thesis is dealing with the humanities approach from one side and a quite technical approach from the other side. This chapter is meant for those readers who do not feel competent enough about one or another of these approaches and tries to reveal their complicated and manifold core. A third part of this chapter is dealing with a short overview over aspects of simultaneous application of both disciplines- the research object is the legal system and the method of research is derived from information science.

Chapter 3 analyses the related work also from two different perspectives- an on-going work of increasing the usability of legislation performed "internally" - performed in a very traditional manner by legal experts. Second perspective is about the information scientist's on-going

research about the same issue, trying to adapt the representation of the legal text in line with the challenges derived from the information revolution. Usually the legal documentation is "reflected" on the web, thus following the logic and build-up of the traditional "paper based" legal system. Such approach greatly underestimates the possibilities and services ICT systems are capable of offering. To understand the different possibilities that exist, to present the legal information better with help of computers and choosing the best option from these many possibilities, is a big and extremely challenging task. In order to succeed in the field of systematisation and simplification of the law, a combination of well-established traditional ways and promising new ones has to be explored.

The idea behind the overview of the EU initiatives is to see how well these can be supported by ICT. The idea behind the overview of research activities is to verify whether on-going academic research is adequate and addresses the same issues as the EU legislator has identified. Table 8 in the subdivision 8.1.1 summarizes this approach and fits the proposed method into same framework. Also an indicative potential of usability of different research approaches is described in this aspect in the last column.

Chapter 6 describes original results of the research, the method for visualisation and content capture of the legislation in detail [19]. *Chapter* 7 provides the reader with results of different test cases performed in order to reveal the value and the applicability of the proposed visualisation method [26, 25]. This is done from different points of view:

- Proving the method suitability through visualizing the legal content and testing it's scaling abilities in different generalisation levels together with suitable illustrations and interactive prototype;
- Testing the method driven and graph theory based legal act similarity measurement hypothesis and presenting the gained results;
- Performing a structural research of the Estonian legislation based on previously found similarity results and offering some novel aspects to the Estonian legislation built-up.

Chapter 8 describes the contribution and proposes ideas for the further research.

4 Background Information

This interdisciplinary research is mainly about two quite challenging fields- legislation and ICT. So far there has not been to too much inter-related activity taken place although ICT is one of the regulated fields within the legislation and legislation itself is presented for a wider public with the help of Internet. Legal informatics is a quite new field that emerged at the cross-road of these two disciplines, dealing with the legislation as a main research object and applying the novel methods derived from the computer science. Mainly these methods are related to the e-Government solutions as the main "client" of the research field.

In order to spot the opportunities for advancement hidden so far, one needs to have preferably good background information about both fields. Increasing complexity can be considered as one of the main weaknesses for both sides and before dealing with such complexity an important question has to be asked- can it be done simpler? Complexity is the price often paid for advancement in our lives and should give reasonable benefit in return.

The following chapter opens these disciplines, starting with a historical overview. The reason behind specific phenomena is perhaps the best justification. Some different aspects are added in order to give vital insights about the issue. Legal history is very long and therefore this overview does not pretend to be an exhaustive and detailed one, but is focusing on the aspects of legal history that are directly relevant for the main content of the thesis (legislation structure, referencing, legal text collections usage problems). ICT and e-Government are on the other hand too new phenomena to have good sources about emergence and development of it. Optimism related to the e-Government potential seems to be still overwhelming and perhaps more time is needed to get balanced surveys about the subject. Yet some generalisations can be made in order to introduce the third part, a merged approach to the subject, combing the different views into the legislation and revealing the challenges related to the logical and mathematical nature of the computers.

4.1 Legislation

This subdivision is meant to provide an overview of the legal system, mainly for non-experts in the field of legislation. It might attract some interest also from the experts due to the emphasis on the constant efforts regarding the systematisation and presentation of regulations.

4.1.1 Short overview of legal system history

The need for law in early societies emerged probably to maintain the "order" in the universe with the underlying notion that harmony must be maintained in the physical universe as well as in social life. [27, p. 345] The history of modern law is dated back to 3000 BC when ancient Egyptian law was formed as a civil code, consisting presumably of twelve books. The basis of it was the concept of *Ma'at* - meaning "Justice" or" truth" but more generally cosmic or earthly order, [28, p. 1093] and the civil code was characterized by tradition, rhetorical speech, social equality and impartiality. The sixth king of Babylon, Hammurabi (1792–1750 BC), is considered to be responsible for developing the first written code of law so called "The code of Hammurabi". It was written in stone and has preserved till our days. It consisted of his legal decisions, all together 282 case laws, and included several aspects like economic provisions (prices, tariffs, trade, and commerce), family law (marriage and divorce), criminal law (assault, theft) and civil law (slavery, debt). [29]. Based on this work, a conceptual model of legislation was created and has been generally unchanged for thousands of years.

The next important development regarding legislation took place in ancient Greece. Ancient Athens, a small city-state at that time, introduced the principle of citizenship (which however excluded women and the slave classes). Although this was an important development in that it created what became an important basis for any legal system, it has to be mentioned that ancient Athens did not have a special term or concept related to "law" and the origins of their legislation has to be found in a variety of literary sources. [30, pp. 5-7] Mostly the distinction between divine law, human decree and custom were retained. [27, p. 346]

Later distinctive developments in the field of legislation took place in different regions of the old world, in Asia (Arthashastra, 400 BC; Manusmriti, 100 AD; traditional Chinese Law and Islamic law for example) and Europe. A major step toward modern legal systems was Roman law, which had a very extensive influence on the formation of the legal systems of most countries in Europe and derivative systems elsewhere. The system was used by Ancient Rome starting from founding of the city in 753 BC until the fall of the Western Empire in the 5th century. [31]

Roman law changed over time quite significantly, from the Twelve Tables (c. 451–450 B.C.E.) to the *Corpus Juris Civilis* (Body of the Civil Law) of the sixth century C.E. It evolved rather randomly over a thousand years and involved many sources. Sources were divided into two groups:

unwritten law *(ius non scriptum)* and written law *(ius scriptum)*. Unwritten law handled the customs in Roman times. At times it was equated with written law. The written part was divided into several sections: acts, plebeian statutes, senate resolutions etc. As there was no coordinated approach to the legislation, nor were legal acts systematically collected, many contradictions occurred in the laws. [32]

Together with increasing amounts of legislation, the inconsistency of legal sources forced a reorganization of the existing legislation. A first attempt to make order from chaos is known as *Institutes of Gaius*. This was an introductory textbook of legal institutions in four books-"commentaris", written by Gaius AD 161. There is not much information about this influential Roman jurist available in these days. The system of Gaius was tripartite: the division was made between the law of person, law of things and law of action (civil procedure). [33, p. 504] *Codex Gregorianus*, the earliest known private collection of imperial laws was published not before AD 291. Unfortunately this codex has not been preserved to our days and is known only by excerpts. [33, p. 391] This work was followed by *Codex Hermogenianus*, containing laws from Diocletian from 291 until 294 AD. An official collection of imperial laws from AD 312 until 438, the *Codex Theodosianus*, was published by Theodosius II. The texts were divided into sixteen books, the books into titles. [33, p. 392]

A very important development for systematization of the Roman legal system took place around 528 AD when Justinian, the Byzantine Emperor, ordered the creation of a team of high officials and lawyers with the task to compile a collection of imperial laws. The team used previous codex's to publish a new code, *Codex Justinianus*. Due to political reasons, a new edition (*Codex repetitae praelectionis*) was published shortly after. The code is divided into twelve books, the books into titles. Within each title the laws are chronologically arranged and provided with information concerning the emperor, the destinatary to whom they were issued and date of issue. (33, p. 392) The compilers were authorized to make appropriate changes in the texts of the laws of the former emperors. Altogether the works are known under the name of *Corpus Juris Civilis* and consisted of four parts: the *Digest* with judicial interpretations of the laws; the *Code*, the elementary textbooks called *Institutes* for law students and the *Novels* (until 565) about new imperial laws. [32]. It collects legal texts from many sources in which the *leges* (laws) and the other rules were expressed or published: proper laws, senatorial consults (*senatusconsulta*), imperial

decrees, case law, and jurists' opinions and interpretations (*responsa prudentum*). [34]

In this way it is possible to say that the systematisation of the legislation is almost as old as legislation itself. As soon as the amount of legislative texts exceeded a certain critical mass and territories of the domains grew, an effective system for collecting and preserving laws was needed together with an effective system for correcting the errors found and contradictions revealed.

After the fall of the Roman Empire, its legislative heritage was abandoned for centuries in Western Europe although it was continuously taught in East. In the end of the eleventh century, an important part of the *Corpus* was found again in Pisa, Italy. This attracted a lot of attention, the missing part was soon recovered and studies related to Roman law spread over Europe again. A legal system based on the Roman origins was adopted by the Catholic Church and extended its influence to almost every country in Europe. In 1140 a *Concordance of Discordant Canons*, or *Decretum Gratiani,* the largest and best organized compendium of church law ever made was prepared. Its creator, the scholar Gratian, used *Corpus Juris Civilis* as a model and practiced the same methods as Roman lawyers. [35]. *Decretum* is a collection of nearly 3,800 texts covering all areas of church discipline and regulation. A treatise attempted to resolve some apparent contradictions and conflicts in the rules accumulated from different sources. Gratian also made extensive use of the works of the Church Fathers and of ecclesiastical writers. [36]

Roman legal texts were used for studies in medieval universities. During this time several approaches were developed. One group, called Glossators, wrote notes in the margins and sometimes between the Justinianic texts, called glosses. These ranged from simple explanations of words and alternative readings of the text to deep analysis of the legal content. Additionally they set up a system of references (allegationes), which helped link together same subject matters from different parts of *Corpus Iuris Civilis*. In good coherence with Justinianus instructions, they tried to interpret the text so that no contradictions remained. This was achieved by grouping together the texts in favour of and against a certain argument or rule, and then interpreting some texts as the rule and others as the exceptions to it by means of distinctions. Another type of writings worth mentioning in this work is literature about the law of procedure: *ordines iudiciarii*. This type of material tried to collect together all the relevant provisions on procedure in general and on specific actions, and provided

instructions on how to produce a writ. One of the most influential works from this range is *Speculum iudiciale* of Wilhelmus Durantis (± 1270). Between 1220 and 1250 the glossator Franciscus Accursius put together a collection of chosen glosses, known as the *Glossa ordinaria* (or *Magna glossa*). "With it, writing glosses had reached its zenith and the attention of legal scholars shifted to different approaches, made possible because the Corpus Iuris had by now been completely cross-referenced and provided with a full commentary." [37, pp. 5-7]

One of the important medieval scientific methods applied on legal texts is the scholastic method or dialectic: formal logic applied to an authoritative text. Attempts were made to handle the text as a logical unit. Applied to *Corpus Iuris* this involved finding a way of harmonising texts containing opposite opinions about, or solutions for the same legal problems. A pioneer for such formal logic was Abélard and he demonstrated this approach in his work *Sic et non*, written around 1120. He used formal logic principles introduced by Aristotle to texts of the Church fathers. Contrasting texts were handled in dialectic, critical manner and this raised several doubts over the texts: the text had been corrupted, the author had made a mistake or the reader did not understand the actual meaning of the writing. These potential issues had to be solved by the treatment of the texts. Texts were collected together based on their similarity (*similia*) or controversy (*contraria*) a "reasoning per analogiam or a contrario is applied, with distinctions (*distinctiones*) being established that explain the differences between the groups". [37, p. 8] This method is called scholastic and can be applied to any authoritative text in any field.

It is interesting to note that lawyers in the middle ages were not acquainted with the idea of hierarchy of sources of law. The civil law and natural law were treated equally until a civil lawyer Azo in his work somewhere around 1150–1230. Summa put some sources of law into higher level than others. The school of Orlèans, traced back until 1235, contributed to the development of laws. They treated Roman texts with more freedom and developed non-Roman areas of law like international private law and penal law. From the point of view of systematisation an important advancement was a new genre called *repetitio*: a separate lecture of especially difficult texts. During this event a central text (*sedes materiae*) was studied in close connection with other texts, which in the end gave rise to a complete and systematic treatment of a certain theme. [37, p. 9]

With economic prosperity the mass of authoritative texts increased significantly in Italy between the 11-th and 13-th century. With this added

to all kind of local customs, the number of legal sources was quite confusing and these were furthermore incomplete and inconsistent. Therefore the importance of Roman law decreased and in the future came to serve more as just a general framework. From the middle of the 13-th century lawyers increasingly based their argumentation on local law or even customary law, which were however analysed based on civil law methods adopted from Roman law. [37, p. 10]

Another advancement made during medieval times by the so called Post-Glossators, Commentators and Consiliators was to introduce a more practice oriented approach. The consilium – the academic advice of a law professor on a practical problem – was a new form of legal literature. Judges in many cases had to ask for a consilium before giving their decisions. Through this instrument Roman law as reflected not least in the *Corpus Iuris* become an important cornerstone of legal practice in medieval times. With a new type of literature - tractatus- the Commentators moved toward more systematic and synthetic use of *Corpus Iuris*. [37, p. 10]

Canon law, which was as important as civil law in medieval times, contributed to the systematic build-up of the modern law through adding an *ordines iudiciarii* – systematic overview of the law of procedure, which was found scattered all over *Corpus Iuris Civilis*. [37, p. 11]

Legal humanism, arising from moral theology and propagated with the help of Thomas Aquinas (±1225-1274), started a new era with a completely different approach to the *Corpus Iuris Civilis*. The critical approach from the Humanists slowly eroded the authority of Roman law in the end of the 16-th century. Humanists tried to restore the original texts of *Corpus Iuris Civilis*, and in this process cast a reasonable doubt over many alleged sources of Roman law as well as over the wordings of certain passages and so on. This made Roman law vulnerable in competition with other legal sources, especially with natural law. [37, p. 16]

The idea of natural law can be traced back to the ancient Greek philosophers. The basis of it is the understanding that law is equally observed by all people, thus having a divine origin. The Dutchman Hugo Grotius (1583-1645) developed from this basis a rational natural law, which was not based on God but rather on human intellect - *ratio*. He presented this position in his work *De iure belli ac pacis* (1625), stating that even if there would be no God, there would be valid natural law. A bit later Samuel Pufendorf (1632-1694) separated natural law from religion completely and changed the build-up of the legal system to make it more geometric, deducing its rules from a number of axioms. After that, the importance of

Roman law faded in practice during the coming some 150 years, giving a way to different designs of natural law. [37, p. 16]

Another important shift in legal understanding occurred mainly by introducing codification. In 1814 a Professor Anton Friedrich Justus Thibaut of Heidelberg proposed the idea that in order to unify national policy a general civil law codification should be carried out. His vicious opponent, Friedrich Carl von Savigny (1779-1861), was against the idea of law constructed in a mathematical way. Savigny argued that codification can be successfully carried out only if it is built up upon historical traditions of the law of the people and knowing the historical background is essential for carrying out the right codification. As one of the first examples of what the theories meant in practice, German codification, the Bürgerliches Gesetzbuch, was published in 1900. [37, p. 18]

The German traditions are important for Estonia, as the German legal system had an important influence on Estonia in the first independence period (1918-1940) and also as Estonian legislation was recreated starting from 1991 based mostly on traditions and principles inherited from Germany.

4.1.2 Overview of the current normative system used in Estonia and principles of interpretation of laws.

According to the classical approach, the Estonian legal system belongs to the continental European legal tradition, the Roman-Germanic family, and follows the classical division into private, public, and criminal law. On the other hand, Estonian law today is increasingly influenced by other legal systems as are legal systems of other developed countries. Different principles contribute to the development of law and each legal system is borrowing something from the other. As a member of the European Union since 2004, generally recognized principles of international law and EU law inject international treaties into Estonian law. Judicial precedent is becoming an important source of law in Estonia today. Case law of the Supreme Court is decisive when there is a need of interpretation of law or when gaps in legislation need to be bridged. The Supreme Court has the authority to interpret legal rules and its opinions are taken into account when dealing with similar situations in the future. Reflection of rapid changes and shifting values in society in court judgments also conforms to the spirit of the law. Interpretation of norms is necessary to allow the legal system to keep pace with a rapidly changing modern society. [38]

In the Estonian legal system, statutory laws are in the top of the so-called law pyramid. Under § 65 of the Constitution of the Republic of Estonia the *Riigikogu*, the parliament of Estonia, in which legislative power is vested, passes laws. To pass a law, the legislators come together, consider a bill and vote on it. These activities of the legislator constitute an external aspect of the process of passing laws, which only determines whether or not a bill will become part of the objective law. For proper application of abstract rules of law the applier should understand the meaning and the purposes of the law. [39]

According to the influential professor of law Narits from Tartu University a proper approach to interpretation of law originates from F. C. von Savigny, who stated the purpose of interpretation is to put oneself in the shoes of legislator and artificially repeat his or her actions. [40] To achieve this it is necessary to analyse the following aspects of the law [41]:

- grammatical;
- logical;
- historical;
- systematic.

The basis for legal interpretation is the text. The interpretation of a text of law does not involve the understanding of single words, but understanding [39]:

- the meaning of words as placed in a certain word order;
- the meaningful relationship between complete sentences.

This is called a circular structure of comprehension and it does not involve tautology. Circular movement does not bring the interpreter back to the starting point but helps to raise comprehension to a new qualitative level. The circular structure of comprehension allows one to return to previously read words in the text or its parts.

To grasp the accurate meaning of a designation the interpreter should [39]:

- know the actual role of a particular word or phrase in language;
- be able to understand the meaning of the word or phrase in the legal context;
- act as a "supplier", analysing different possible meanings and constantly asking which is the correct meaning in the context;
- rely:
  - on the text of law;
  - on the words contained therein;

- on the relations between different parts of the text;
- reach a decision which is not imposed logically but is based on a reasoned choice between different possible meanings.

It would be erroneous to think that texts of law require interpretation only if they are particularly unclear or contradictory. All texts of law should be interpreted, and the need for interpretation does not reflect a deficiency of the texts but is quite intentional. [39]

Prof. Narits assumes that no law can be created in such perfect form that it would instantly provide answers to different cases arising therefrom. Therefore, those who make legal decisions must abide by the law (ensure stability), allowing for a certain flexibility (interpretation). According to A. Aarnio, provisions of law are like pieces of elastic: the interpreter stretches them depending on the circumstances. If a law cannot be stretched any further, it is time to resolve the issue by passing a new law. [42] [39]

Language has constantly changing nature. The meanings of the word might change over time and give unexpected results in future. Sometimes the used words or phrases are not exact enough to deliver the actual meaning and therefore poorly or falsely applied. Sometimes a term might have many different meanings and then it is not known which one of them should be used within the given context. These problems can be classified into lexical or syntactic ambiguities. Lexical ambiguity expects users to use the meaning as decision ground, in case of syntactic ambiguity the syntactic structure should be analysed. [39]

4.1.3 The meaning of the law in modern society.

One can find many definitions for one word: law. For example the Oxford Dictionary gives this definition [43]:

"Law is the system of rules which a particular country or community recognizes as regulating the actions of its members and which it may enforce by the imposition of penalties."

Merriam-Webster gives in his online Encyclopaedia Britannica the following definition [44] :

"Discipline and profession concerned with the customs, practices, and rules of conduct that are recognized as binding by the community. Enforcement of the body of rules is through a controlling authority, such as a group of elders, a regent, a court, or a judiciary. Comparative law is the study of the differences, similarities, and interrelationships of different systems of law."

Another angle is represented by sociologists, who seem to be better able to formularize this phenomenon, seeming too big and complex for insiders [45, p. 3]:

"Law in its broadest sociological sense is a field of experience shaped and structured by problems of government, social control and social order."

The meaning of law can be viewed from different angles: for lawyers it is mostly a field of professional experience, for citizens an experience of power available for them. Mostly it is used in coherence with the ability to coerce, influence something, make something happen and get things done.

Sometimes a legal experience is about power over citizens, about the feeling of being subject of official control; it can be also a relatively unfocused feeling of general security, experience of safe streets and secure homes, reliable transactions with others, and of plans and investments in which reasonable expectations will not be frustrated. [45, pp. 3-5]

Law has several aims, four of them are central in this context. They are all concerned with making society more stable and enabling people to flourish:

- Set up an official framework of compulsion to forbid certain ways of behaving, like murder, libel;
- Provide facilities for people to make their own arrangements. With help of laws it is possible to guarantee to people who buy and sell goods, make wills etc. that the state will if necessary enforce these arrangements;
- Settle disputes about what law is and whether it has been broken;
- Settle what the system of government is to be. [46, pp. 1-2]

There are a variety of definitions of rule of law per se but there are elements which have to be recognized as necessary for a society aspiring to institute the rule of law [47]:

- Laws must exist and those laws should be obeyed by all, including government officials;
- Laws must be published;
- Laws must be prospective in nature so that the effect of the law may only take place after the law has been passed;
- Laws should be written with reasonable clarity to avoid unfair enforcement;
- Law must avoid contradictions;
- Law must not command the impossible;

- Law must stay constant through time to allow the formalization of rules. However, law also must allow for timely revision when the underlying social and political circumstances have changed;
- Official action should be consistent with the declared rule.

4.2 ICT development and potential to solve different society related problems

A civilized society, like a functioning free market, develops spontaneously and is tremendously complex. Development of communication systems, increasing population and globalisation of markets creates more and more opportunities and ways to communicate, but also brings along a lot of risks and problems for people involved. Therefore a set of agreements are used to reduce the uncertainty and possible negative impact. Mankind has historically developed a legal system to deal with such kind of complexity and development. Legal systems have many positive features, but inevitably also have certain weaknesses to deal with. Therefore it is interesting to analyse a bit the potential of modern technologies, like information and communication technologies to assess their applicability and suitability for solving such challenges that humankind is facing at the moment. As the history of this kind of effort is relatively short, the approach is bit different regarding the previous subdivision and describes the broader context at first and moving toward specific subfields later.

4.2.1 ICT potential in general

It is a widely held belief that ICT has an important role to play in national development. Many governments over the world have made significant investments in ICT and believe that it as an important contributor to industrialisation and economic development.

The emergence of ICTs seems to change the society even more and quicker than the industrial revolution. We are so used to the World Wide Web and mobile phones that it seems even impossible to return to the point before that. The main advantages (transmission of data at the speed of light, loss of regional barriers etc.) of this technology seems for us so convincing that we even do not ask whether and how useful it really is to apply this technology everywhere.

Whether this kind of assumptions and beliefs are justified is still under discussion. No good comes without bad, so what is the price we have to pay? In order to understand the issue better we have to dig a little deeper.

According to Sein and Harindranath there is the optimism that ICT will lead to development and pessimism that ICT by itself will not lead to development unless accompanied by social changes. The discussion involves three main points: modernization, dependency and human centeredness (human development). [48]

Looking at the developing countries they seem to have an advantage. In order to become developed, they need to utilize the power of technology, increase the availability of capital and skilled workforce and also promote an entrepreneurial class to achieve growth. ICT can be of great help here, but as they come late in the adoption process, they bypass the teething problems of the leading-edge technology and can benefit from declining costs. The weakness of this aspect is the fact that cultural and local context is not always taken into account.

Developed countries on the contrary have gone through a long process, partly started from exploiting the poorer countries through colonisation and trade. Poorer countries had to inline their production and services to benefit the richer countries. In this view the ICT is strengthening their position but there is also weakness: it treats underdeveloped countries uncritically.

Sein and Harindranath put up another alternative view, called human development or people centred development:

- Enabling choices in education, health, and standard;
- Building of living, a democratic society marked by involvement;
- Better participation, management and of transparency behaviour and customs, based on a better understanding of culture.

The main question for the cited authors regarding this view is whether and how ICT has been applied in these developments? The extent and nature of the impact needs to be investigated to estimate the role of the ICT in this. Therefore they used three different conceptualizations of ICT, specified by its subdivisions [48]:

- its use:
    - as commodity or product to earn foreign currency;
    - as support to development activities through planning, management, training and support to the development projects;
    - as a driver of the economy through products and service industries generated directly and indirectly etc.;
    - directed at specific development activities as having a developmental impact when it is used within the context of targeted developmental initiatives;

- how it is viewed:
  - Nominal View, where ICT is present only as the object of study and no specific view is articulated;
  - Tool View, where ICT is conceptualized as an engineered artefact, is a technical entity, and thus is a means to achieve something;
  - Computational View, where ICT is conceptualized purely as technology and the algorithms, codes, and models that comprise the system;
  - Proxy View, where ICT is conceptualized in terms of a surrogate for some other concept;
  - Ensemble View, where ICT is conceptualized as part of a bigger "package" going beyond the technology (hardware, software) to activities and interactions performed in specific social and cultural contexts;

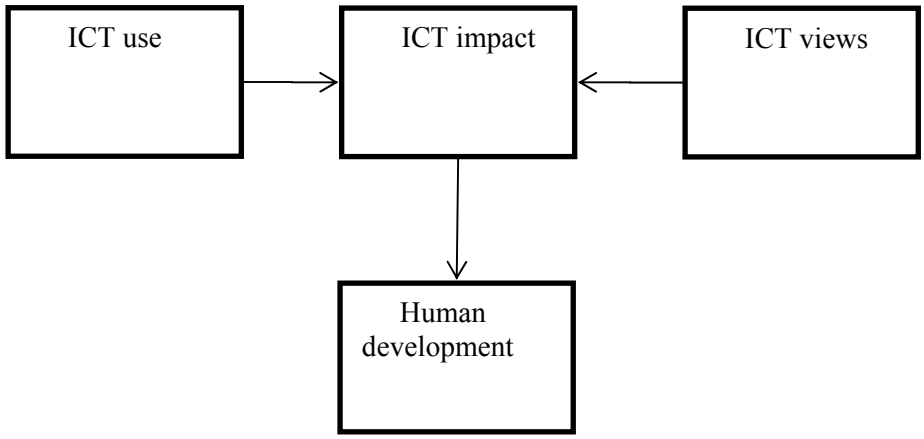


Figure 3. Integrative framework of ICT in development [48]

- how it impacts development:
  - primary effect - simple substitution of old technology by the new;
  - secondary effect - an increase in the phenomenon enabled by the technology;

- tertiary effect - the generation of new technology-related businesses and societal change.

However it is clear that any presented view cannot be evaluated independently to evaluate ICT effect on society development. The impact on human development is strongly related to the use of ICT and views on ICT. Therefore it should be handled inside the same framework, which is presented in Figure 3.

4.2.2 e-Government

The history of e-Government seems to be a too fresh issue at the moment to find many relevant sources in literature. The digital revolution has been with us only for a couple of decades and as this is still going on, the full extent and properties of it are hard to evaluate. The term e-Government emerged together with an Internet boom in the 90-s of the last century. It started as a practitioner's field who were struggling to meet new challenges of the Internet medium implementing new systems creatively.

e-Government is a complex issue to handle yet capable of changing government work to a great extent. According to researches quite many e-Government related articles in the academic world refer to the law approximately 7% in some or other way [49, pp. 178-185]. Therefore it is important to investigate this aspect of the e-Development.

Quick and thorough changes in ICT have created an information rich virtual environment, where the knowledge has a central role to play in gaining economic competitiveness. The challenge for political leaders of the different countries today is to provide opportunities for citizens and businesses to actively participate in the global knowledge based economy and maximize the country's benefits gained from ICT overall development.

In order to achieve these ambitious targets, the public sector faces a need to achieve greater operational efficiency and respond better to citizens demands for improved public services. This enforces governments to deal with economic and administrative reforms increasingly.

Despite the obvious orientation to promote the economy, the e-Government is considered to be a specific subdivision of the broad ICT application field. According to OECD definition e-Government "*focuses on the use of new information and communication technologies (ICTs) by governments as applied to the full range of government functions. In particular, the networking potential offered by the Internet and related technologies has the potential to transform the structures and operation of*

government." [50]. In the broadest sense it is referred to as the technology enabled transformation of government in order to, among others:

- reduce costs;
- promote economic development;
- enhance transparency and accountability;
- improve service delivery;
- facilitate the advancement of an information society.

According to a World Bank overview, usually e-Government relationships are characterized in three ways [51]:

- Government to Citizen (G2C): deals with the relationship between government and citizens;
- Government to Business (G2B): consists of e-interactions between government and the private sector;
- Government to Government (G2G): Governments depend on other levels of government within the state to effectively deliver services and allocate responsibilities.

e-Government uses different channels to deliver the services. This might include:

- Conventional telephone - call centres, automated voice responses, FAX on demand of forms and information;
- Personal Computers and the Internet - from simple display of information to highly sophisticated interactive services with credit card payment;
- In an assisted mode through integrated service delivery centres in convenient locations, rural Internet kiosks and cyber cafes;
- Mobile digital telephony and messaging technology;
- By any combination of ICT and manual procedures, for example providing information on the web, ordering and paying by mobile telephone, and delivery at a community centre.

United Nations (UN) Public Administration programme has developed a universal scheme to evaluate a country's e-Success. This model involves three main aspects with its subdivisions [52]:

- E-Government– index is used to measure the e-Government Development state; a composite measurement of the capacity and willingness of countries to use e-government for ICT-led development. It has three sub measures:

    - Web Measure- four-stage model of Web Measure Index to scale of progressively sophisticated citizen services within country;
    - Infrastructure- a composite weighted average index of six primary indices based on basic infrastructural indicators, which define a country's ICT infrastructure capacity, among others:
        - PC's/1000 persons;
        - Internet users/1000 persons;
        - Telephone Lines/1000 persons;
    - Human Capital- a composite of the adult literacy rate and the combined primary, secondary and tertiary gross enrolment ratio with two third weight given to adult literacy and one third to gross enrolment ratio;
- E-Participation - assesses the quality and usefulness of information and services provided by a country for the purpose of engaging its citizens in public policy making through the use of e-government programs;
- E-Inclusion- provides an assessment of the existing access-divide in the world, among others:
    - income divide;
    - telecommunication access-divide;
    - education access-divide.

4.2.3 Legal informatics

One of the emerging fields related to the Internet and e-Government is the discipline called legal informatics. As the importance of legislative informative systems has increased greatly in the last years, it has dramatically changed the way the information is produced and communicated [53]. Internet is widely used as a source of information by citizens but also from legal practitioners side and proper legal sources are moving into Internet. The Estonian State Gazette for example does not have a printed version any more (www.riigiteataja.ee). Legal documents are becoming a part of Web originated information overflow, being tightly interlinked with the rest of legal documents and also with other sources. This is bringing us to the emergence of a legal web as a subset of the broader Worldwide Web. Furthermore, this is bringing in new ways of communication between legislators and citizens, like interactive participation in legal discussions, litigation, etc. This opens totally new

perspectives for public debate and also for other important aspects of democracy like voting etc. [54, p. 9]

”Legal informatics is the discipline, which deals with the use of ICT to process legal information and support legal activities, namely, the creation, the cognition and the application of the law.“ [54, p. 10]

Legal informatics emerged in the 1960’s when the first legal document databases were created. In 1970’s the public sector started widely used large databases for administrative data (about population, fiscal info, etc.) In the 1980’s the era of office automation started, when the computer entered into most legal and administrative offices in the West. In the 1990’s the wider use of information systems to support the activities of public sector organisations, including legal offices, took place. This was the time of integration when different separated programs were compiled into organisation-wide information systems and it brought in such important changes as reengineering work processes to achieve rational workflows to enhance the organisation capabilities. At this time, legal informatics started to adopt new and modern ways of managing legal information, like analyses of legal reasoning and modelling of norms. Starting in the beginning of the 2000’s, the legal informatics profited from Internet and expanded the scope to the communication between legal organizations and their public- citizens, economic units and their consultants. In this way the legal informatics has become a significant aspect and important resource of e-Government. [54, p. 11]

A new and a bit different version of this discipline is legislative informatics, which has the following functions [54, pp. 12-13]:

- Providing information to all actors involved in the legislative process (legislators, citizens, experts, political parties, associations, organizations, lobbies, etc.);
- Enabling cooperation among such actors, so that each one can contribute to the process, according to his or her role;
- Ensuring efficiency, transparency, and control;
- Enabling access to the outcomes of each phase of the legislative process, and contributions to the next phase (managing the workflow);
- Ensuring the quality of the legislative outputs;
- Ensuring knowledge of the produced law texts and preliminary documents.

Some achievements made:

- in separate field:

    - legislation is now available in electronic format (in many countries);
    - it can be accessible online (in many countries);
    - the version in force can be automatically constructed (in some advanced approaches);
    - retrieval is facilitated by metadata and thesauri or ontologies (in some advanced approaches);
    - legislation produced by different bodies and stored in different databases is retrieved through a single Internet-based interface (in some advanced approaches);
- in workflows:
    - All information concerning a new bill can be preserved and updated during the process (in many countries);
    - The legislative process is monitored and supported by workflow facilities (to some extent, in many countries);
    - Amendments are updated, so that accurate information is provided on the state of the text as the procedure goes on (in some advanced approaches);
- in communication and information support:
    - IT supports interaction between legislators and other actors: between individual MPs and their constituency through e-petition, e-consultation, etc. (in some advanced approaches);
    - Relevant information is offered before any legislative choice (in some advanced approaches);
- in computable models of legislative knowledge:
    - Computer models are built of legislative concepts (in some advanced approaches, like the EU project Dalos and Estrella);
    - Consistency and completeness of legislation are checked (still experimentally);
    - Legislative impacts are being modelled (to some extent, in some advanced approaches, by using different techniques, from traditional statistical analyses to agent-based modelling);
    - Knowledge-based systems for the IT-supported application legislation are being developed (in some advanced approaches);
- in the area of legislative standards:
    - Standards are available for the structure of law texts (in Europe, Africa and other countries);

- Standards are available for references and modifications (in some advanced approaches);
- Standards are being developed for preliminary materials produced along the legislative process (in some advanced approaches);
- Standard-compliant software is being developed (in many countries).

4.3 Legislation related systematisation issues

Hereby the analysis of specific weaknesses of the legislation follows from two different levels (general and local), describing the problem mainly from the technical point of view and will offer some new aspects in accordance with a networked society perspective.

4.3.1 General problems with normative system globally and locally.

As, mentioned legislation plays an important role in the life cycle of a society, but has its weaknesses to be dealt with. Among others, the aspects of the general structure of the regulation and its simplicity are often under observation. The growth of national and international legal corpus and the rapidity of changes become more and more complex to manage for administrators. The impact that this complexity has on the intricacy of both legal hierarchies and legal contents are considered as causing other unwanted by-effects on the intelligibility of laws for citizens. [11] [55]

The quality of law is an issue, which has many different possibly problematic aspects and one can find a lot of governmental and non-governmental institutions addressing such problems [56, p. 23] [57] [58] [59, 60]. From the Artificial Intelligence (AI) point of view several researchers have found that the legal system usually suffers from many problems, like scarce transparency, lack of visibility and deficient structure [61] [62]. There are many causes to such problems, including the following:

- non-systematic organization of the legal order that arises from normative overproduction [62];
- fragmentation and overlapping of the modifications [62];
- emerging new soft law sources: -authority acts, new institutional bodies, etc. [62];
- unclean law-making and legal drafting processes, misleading about the range of competences assigned to national and regional bodies [62];

- Different legal sources belonging to different publishers, in different formats and with different methods for identification and referring to parts of the collections [63];
- Many different electronic formats available in the web (word, pdf, html, xml, etc.) [59].

Rules and norms have been created from the beginning of times and meant for human beings exclusively instead of for technical devices to implement it. Nevertheless, the increasing complexity of society creates needs for more complex legislation and this raises the relevancy of also developing assisting tools to deal with it.

The emergence of the World Wide Web as an enormously popular and information rich environment has created pressure to publish also legislation in it. This has been a big challenge for researchers and has been forcing them to re-evaluate the very basis of the legislation. Therefore the law is seen a normative and documentary unit of reference, hence the inability to obtain an analytical/ systematic vision of a legal order itself, allowing to query a legal information system according to the content of each norm. [14] This necessarily creates obstacles to the knowledge and upkeep of the legal order: from the uncertainty of the impact of new laws on the legal order in terms of coherency preservation, to the difficulties in norm-accessing by both citizens and legal experts [61]. Knowledge is not consistently documented, and documented knowledge is not always explicit [64]. All this makes the gathering and exploiting the extended legal knowledge feasible mainly to trained specialists and very unsuitable for AI related IT systems.

Well planned, easily searchable and user friendly presentation of legal documents improves greatly the legislation's simplicity. This is understood well by EU officials, declaring that "*Developing a user-friendly and easily accessible European law is a major concern for the Commission and the other EU institutions*". [65] Simplifying a bit one can say: information is best if it is presented in suitable form, in suitable amount, in suitable place and in suitable time. The information age is offering a new ways of doing this and there are many research activities going on at the moment, but legal work is still done mainly in a traditional manner and does not consider much the developments in this field. [66]

A good example of a well-accepted and widely recognized IT solution in legal field is web based legislation collections (EUR-Lex, Riigi Teataja, etc), making the legislation easily accessible for everyone. Usually the legal documentation is "reflected" on the web, thus following the logic and build-

up of the traditional "paper based" legal system. Such approach greatly underestimates the possibilities and services IT systems are capable to offer today. Due to the complexity of the issue it is not easy to understand how the electronic legal information can be presented and received in the best possible way. In order to succeed in this the traditional, legal methods and promising new, IT originated methods have to be explored.

4.3.2 Legislation as an influential part of network

The legislation is visualized according to popular interpretation as a pyramid of legal documents, forming a hierarchy based on the relative importance of legal acts. These acts do not exist independently but do have a strong internal coherency regarding to the higher level acts: in the case of contradiction between normative legal acts of different legal force, the normative legal act with superior legal force shall apply, *lex superior legi inferiori derogat.* Lower level acts derive their legal power from superior normative acts and the link to it should be clearly stated. Thus a set of legal acts and their interrelations could in ideal case be visualised as shown in Figure 4.

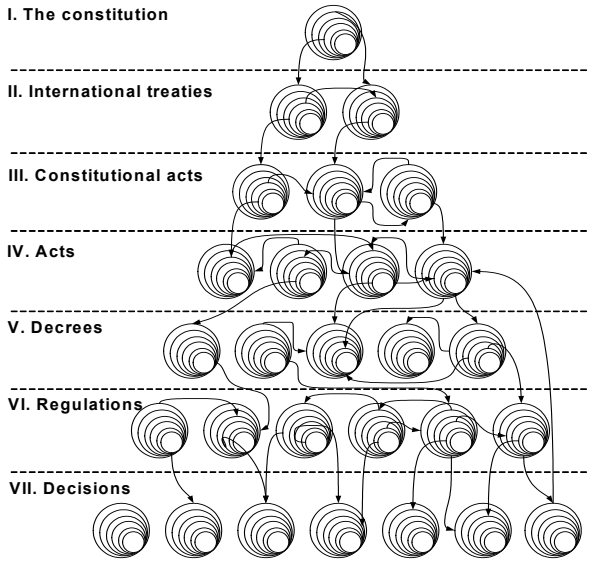


Figure 4. Ideal hierarchical build-up of Estonian legislation

Acknowledging the obvious need to remove contradictions between different legal acts and collect the data into one source in order for it to be

easily found, there are several other aspects developed over time to evaluate the quality of the legislation. Legislation is meant for society, supposed to benefit to the society and the legislative influence can therefore be measured differently.

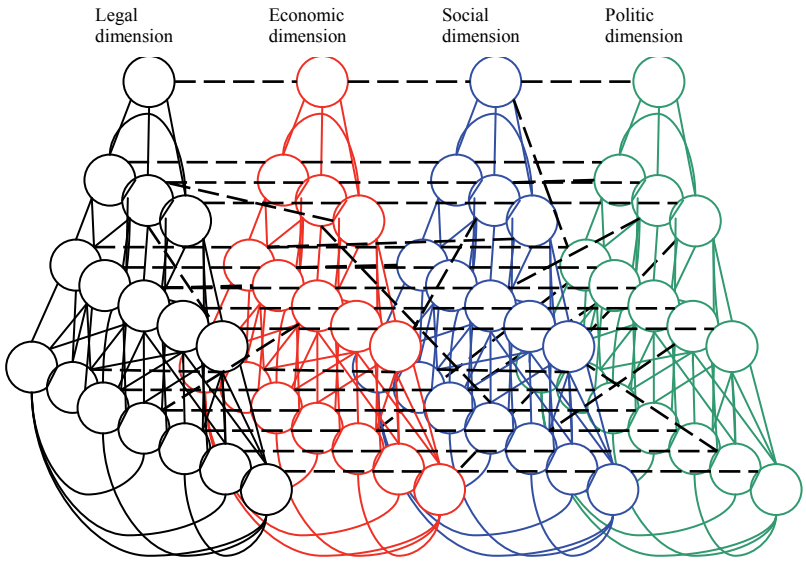


Figure 5. Legislation and its influence dimensions

Assuming that the legislation is a result of political will, consolidating the principles to guide the society into a prosperous future, the legislation is not an object standing outside the society but rather a tool for society. As a broad generalisation it can be said that this tool is used to shape the variety of decisions to maximize positive outcomes and minimize the negative ones. Each decision made in the highest legislative level is obviously important enough to enter into the litigation process and therefore meant to influence the society directly or indirectly. If this is not the case, the need for such legal act is questionable. If it is the case, this influence can, or in some cases, even must be followed and probably can be indicated or measured somehow. Society in turn influences the creation of new normative acts or changes in the existing legislation. Therefore the interaction of legislation and society can be viewed as a highly complicated set of links connecting different aspects of social life. From this angle the

legislation is not only a tightly interconnected set of hierarchically ordered documents, but a complicated model with a huge set of links connecting legislative documents to different other documents, aspects, layers and functions of the society. Having a good overview of these connections and being able to visualise the connections makes it possible to understand the functioning of the society better and perhaps even model it.

For the sake of clarity it is useful to draw a simplified picture of these influential dimensions: economic, political, social, environmental etc. These influence dimensions and connections can be principally visualised as presented in Figure 5.

Therefore a question arises: whether it is useful to view the legislation as a set of connections rather than a set of documents which has predictable outcome in society in different aspects? What could be the positive results of such an approach and can it be exploited for the sake of society?

4.3.3 Analysis of the existing legislation structure

In order to increase the usability of the legislation with help of enhanced structural properties an existing structure of the Estonian legislation has to be estimated and evaluated. Lately the visualisation of such networks is becoming more and more popular in the relevant academic literature. There are some novel researches available in this field, made on the basis of US legislation [67] [68] [69] [70] and French legislation [11]. US legislation is case based and therefore the court cases form an important part of such a legal system, the links between cases are relatively easily followed. One of such kind of impressive dynamic visualisations can be viewed from here: http://vimeo.com/9427420. The above-mentioned French legislation structural research approach estimates the links between legal acts and visualises these in order to estimate the "central players" and attempt to evaluate the complexity of the legislation as a whole.

In Estonia only two types of references are defined at the legislative level, "Technical Rules for Drafts of Legislation of General Application", §14, 15 [71] :

"Direct reference- a direct reference is a reference to a provision of the same draft Act, an Act or a regulation of the European Community or the European Atomic Energy Community.

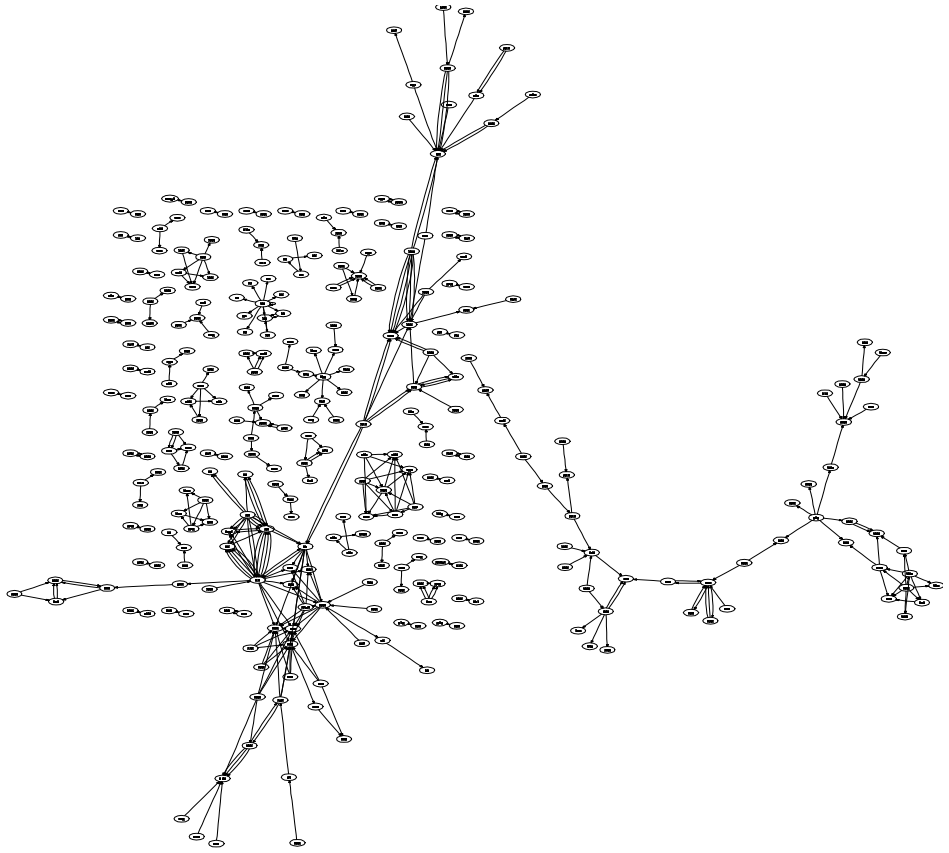


Figure 6. The visual representation of textual references between subdivisions within the Law of Obligations Act

Indirect reference - a reference to an applicable legal institute, an Act, EU legislation or an area within the competence of the EU."

For visualizing the existing structure of the legal act inside, more specifically between subdivisions of a law, an experimental research was conducted based on the Estonian Law of Obligations (Võlaõigusseadus). This legal act was selected for this research due to the large amount of sections (1075) and references between them (480). [10]

In the above-mentioned research direct links were analysed. Although we here consider one document from thousands forming the legislation, it gives a pretty good overview about non-systematic use of direct references

inside the legal texts. Hierarchical structure of the normative system is vaguely estimated- it has a well-developed referencing system, but does not form a systematic and well-structured connected graph (see Figure 6) [10].

This has several reasons, like the non-systematic usage practice, frequent changes in the existing legislation, big number of existing links hard to manage etc. The most important reason comes from the document "Technical Rules for Drafts of Legislation of General Application" [71]:

" §14. Direct reference

(2) No reference shall be made to a provision of legislation with lower legal force and to a provision of an Act which has not entered into force.

(4) Upon direct reference, the following references shall be avoided:

1) reference to the Constitution;

2) reference to the subsequent provisions of the draft Act;

3) reference to a referring provision;

4) repeated reference in one section;

5) reference to a directive of the European Community or the European Atomic Energy Community, unless the directive is directly applicable in Estonian law."

As a result of preliminary research, it is not possible to visualize the Estonian legislation as a connected graph due to inadequately determined legal references, as it was estimated in Figure 6.

Many research papers offer a graph-like presentation of the law, helping to visualize the connected parts of the legal documentation based on textual references [10] [11] [67] [68] [69] [12] [70]. These researches do investigate the use and visualisation of direct references which were considered as not applicable within the current research and were not used in further developments. Therefore a need for an additional, preferably computable layer to reveal the structure of the existing legislation emerged, to interconnect different yet similar parts of the legislative documentation.

A loss of systematic built-up of the legislation seems to be inevitable. The legal system complexity increases despite the constant adjustments with each additional legal act. Scientific and technological advancements (telecommunication, space exploration and so on) expect us to bring in more and more regulations in order to regulate emerging and qualitatively new relations within society. This makes it harder for ordinary citizens to use the legislation despite that they are obligated to so and this process never stops. Complexity might often challenge even specialists trained to overcome these obstacles.

On the other hand there are new tools and means offered by modern society to overcome such shortcomings. Governments, which are the main players in the information revolution, are capable of exploiting these advantages widely for public benefit including legislation related aspects. There is even a special academic research domain targeted to deal with legislation related challenges. So there is a long lasting pain and there are modern tools having potential to solve quite some of the abovementioned problems. Somehow the progress has been slow. On the one hand, computers capable of doing billions of precise mathematical calculations within seconds are not good at making generalisations and applying broader legal reasoning. On the other hand traditionally well developed skills and conservative nature of legal experts might cause this slowness. So it might be fruitful to analyse deeper how legislation changes in response to outside changes and how academic research helps to cope with this.

5 Review of the State of the Art

The usability of law is under more detailed consideration in this Chapter, where it is handled from two viewpoints as previously- traditional, legislation centric and ICT centric approach. A specific description about efforts from both sides is presented and a conclusion about joined approach potential is described. The legislation-related subdivision deals with different aspects of the work done in order to simplify the legislation from different levels – a national, federal and EU level. The relevant ICT research in support of EU efforts is described later. Table 8 is handling these issues from the joint viewpoint, but is presented in the end of the thesis (subdivision 8.1.1) to include also aspects from the current research. The major part of this chapter (subchapters 5.1.4 and 5.2) is published in K. Nyman-Metcalf and E. Täks article "Simplifying the law - can ICT help us?" [24].

5.1 Increasing the usability of law, a modern approach

There is no denying that legislation plays a big role in any society. From a sociological viewpoint it can be said to be a tool of government to organize society and protect citizens. Versions of the kind of legal systems that exist today have been used for thousands of years and have thus in essence proven their value, but inevitably the systems have organic weaknesses that have to be considered and overcome. Legislation is usually a huge collection of different normative documents, obligatory to know and follow for everyone yet effectively handled by a few. This huge collection furthermore tends to keep growing and changing with time. It is no exaggeration to say that finding a relevant norm may be a challenging task even for experts and all the more so for laymen

As for modern society a meaning for a good legislation, or the values to which a good legislation should implement, can be formulated as following [54]:

- Means-end rationality-regulations should effectively solve problems they address;
- Discursive rationality- debate over new legislations should be as open as possible, all interests should be considered, relevant arguments evaluated for and against the regulation;
- Responsiveness to citizens needs and preferences- regulation should be focused on the common goods;

- Progress (ability to change)- able to adapt to social change,
- Legal - should contribute to the certainty of the law, providing normative information which gives effective guidance to citizens and legal decision makers;
- Citizens' rights- should enable citizens to have a clear idea of what their rights are.

Different methods for legislation systematization, codification etc. are well developed, but there are more quality related aspects taken into account these days.

5.1.1 Motivation for systematization from a local perspective

The idea of systematisation of legal provisions has deep roots that go back thousands of years in history. As presented in the short history of the legal system, ruling the Roman Empire required an extensive amount of legislative acts, which eventually started to contradict each other at the highest legislative level. The same processes can be followed also today but problems seem to arise also because of the complexity of modern life. This is forcing also remarkably smaller countries than the Roman Empire to address this issue.

The systematisation of law is not only a technical task, but involves also a development of the systems of concepts. This helps to implement the rule of law as an idea in applied form. Just compiling the similar laws into one code is not enough; systematisation has to contain general content and be unified in order to answer the common task of legal science and all legislation. Systematisation is a tool or method of building a system from essentially uneven and irregularly linked provisions. Codification is a way to achieve legal certainty and clarity, by making it simpler for those applying the law to find the necessary regulation and providing a more general view of the applicable law. Doing so, the stability and unambiguousness of the legal order had to be maintained as the functionality and system of the laws not yet systematised should not be questioned. [72]

According to Narits the need for systematising is greater when the legal order has been developed over a great length of time and the body of law is large and even partly contradictory. On the other hand, the need may be great also when the legal order emerged over a relatively short span of time, in an extensive burst of activity. It might become difficult to find the necessary provision and this task might become much more complicated and problematic with time. Another challenge is to guarantee the

consistency in the situation when legal drafting is made in different contexts, like both the national and European Union context, as is the case for Estonia and all the other EU Member States.

In his article prof. Narits counts four traditional aspects of systematisation of legislation:

- the activities of state agencies, enterprises, and other agencies and organisations in collecting sources of law and filing them according to a system — this form could be called 'accounting for' the sources of law;
- preparation and publication of sets and bodies of different types of sources of law — this activity could be called 'incorporation';
- gathering of different provisions and sets of provisions of the same quality into a single compilation — this activity could be called the 'consolidation of law';
- preparation and adoption of new sources of law — this is the codification of legislation.

Prof. Narits also presents some important aspects of systematisation of law:

- the efficiency of the pre-project stage of legal drafting;
- the effectiveness of implementation of law;
- the legal educational element;

and some principles of codification to follow:

- the completeness of the body of information;
- the reliability of the information;
- the convenience of use.

He concludes in his work: "It is clear that the most modern way of collecting sources of law pertains to computers and relevant computer programs" [72]

5.1.2 US legislation Regulatory Impact Analysis (RIA) approach

The first steps toward Regulatory Impact Analysis (RIA) was done in the United States during the Nixon administration (1969-1974), when US companies started to complain about costs related to implementing the new legislation, mostly as a result of newly passed environmental legislation. In 1971 President Nixon introduced the Quality of Life review process to force agencies to consider various regulatory alternatives and costs when developing significant regulations. This first initiative was unsuccessful due to the aggressive pressure from well-organized interest groups [73]. Next

US president Ford promoted the use of cost-benefit analysis of proposed regulations with Executive Order 11,821, issued in 1974. In this document he mandated an Inflation Impact Assessment by federal agencies to be performed as an *ex ante* assessment of new regulatory measures on the inflation rate. The purpose was to ensure that proposed regulations likely to exert an effect on inflation could be rejected. President Carter then took over the initiative and put a group of economists called "Regulatory Analysis Review Group" to do the *ex ante* assessment of the 10 most relevant new regulations in each year's US government agenda. [74].

President Reagan forced agencies to adopt a fully-fledged regulatory impact analysis with his EO 12,291, issued in 1981 in order to perform a thorough reassessment of existing regulation in force to identify norms to be abolished or simplified. [75] For this purpose a special governmental unit, Office of Information and Regulatory Affairs was created within the Office of Management and Budget (OMB) under the White House, Presidents office. Also a special unit called Task Force on Regulatory Relief was created in order to deal with issues arising between OMB and agencies. According to this administration performed analysis, the yearly savings could reach up to 10-20 billion dollars. Due to this initiative the Reagan administration was able to drastically reduce the budget and personnel of federal agencies, achieving substantial decreases in the cost of regulation. From the negative side he managed to achieve also a dramatic reduction in the number of policy proposals. [76, p. 25].

US President Bush Sr. tried to abrogate all federal rules that could hinder the competitiveness of US firms. [74]. The main purpose was to reduce regulatory related costs for the economy as a whole. Still this initiative was heavily criticized and the importance of this initiative was underestimated. President Clinton addressed this issue from another angle in his EO 12866, National Performance Review and with the help of some adjustments managed significantly to increase the transparency of regulatory reforms in USA. [76, p. 27]

RIA is now commonly understood as a set of tools for better regulations, including the following general steps [76]:

- Problem definition- identification of the problem, e.g. a market failure or a regulatory failure, and the identification of the need to intervene;
- Identification of alternative regulatory options- light-touch regulation, regulation through information, principles-based

regulation, and alternative forms of intervention such as self- and co-regulation;
- Data collection- desk research, consultation of stakeholders, and commissioning of external studies;
- Assessment of alternative options – the most common being cost-effectiveness analysis (CEA), cost-benefit analysis (CBA), least-cost assessment (LCA) and risk analysis;
- Identification of the preferred policy option;
- In-depth assessment of the impact of the preferred policy option-subject to a more in-depth assessment, mostly aimed at quantifying the prospective impacts and associated risks, and identifying indicators for monitoring and evaluating the policy action over time;
- Input to proposal drafting- only as the result of an assessment exercise mostly aimed at capturing impacts;
- Provisions for monitoring and evaluation-how the impact of the selected policy action can be monitored overtime, and a clear and efficient time horizon for revision of the action in the future.

An important feature of the RIA is the limited scope in the US. RIA has been applied only to the secondary legislation, namely agency-made rules and is mandatory for government agencies only when they entail 'significant regulatory actions' [76, p. 31]:

- had an annual effect on the economy of $100 million or more adversely affect etc;
- Created a serious inconsistency or otherwise interfered with an action taken or planned by another agency;
- Materially altered the budgetary impact of entitlements, grants, etc.;
- raised novel legal or policy issues arising out of legal mandates.

5.1.3 EU legislation simplification approaches

Under the UK presidency a Business Impact Assessment (BIA) System was launched in 1986 in the European Commission. The initiative was similar to something called the British Compliance Cost Assessment procedure and tried to measure the regulatory impact toward businesses but not toward social welfare as a whole. Due to criticism received regarding this approach the European Commission decided to add new tools and to launch new projects like [76, p. 51]:

- SLIM project (Simplification of the Legislation on the Internal Market) with a purpose to simplify and update Community

Legislation. This programme included a range of initiatives (codification, consolidation, the removal of out of date texts and simplification) during two years, starting 1996 and had several iterations later; [77]

- BEST (Business Environment Simplification Task Force), in 1997. The aim of the Task Force was to see how the environment of businesses can be improved in order to create new jobs, by simplifying the regulatory and administrative burdens for particularly small and medium sized enterprises (SMEs); [78]
- European Business Test Panel (EBTB) in 1998 to systematically consult stakeholders and the general public in order to ensure that the legislative proposals are responding to real needs. Among the stakeholders are social partners, businesses, environmental and other civil society organisations. [79]

However, these attempts were considered unsuccessful and therefore the European Commission issued in 2001 a White Paper to mandate the creation of a high-level advisory group (the so-called 'Mandelkern Group') for the drafting of an "action plan for better regulation" and the definition of a new model of impact assessment to be implemented at Community level. This group was formed in December 2001 and included 16 experts from 15 EU countries and the Commission. The overall task was to develop a common method of evaluating the quality of regulation. Some additional guidelines were suggested in Strasbourg, on the 7th November 2000, when the Ministers in charge of public administration decided to set up a High Level Consultative Group on Better Regulation:

- The systematic use of impact studies;
- Transparency in the consultation process prior to the drafting of texts;
- Simplification of the texts adopted;
- Wider use of codification.

The expert group added another topic: the type of structures that would make sure that the procedures to ensure improvements in the quality of regulation were effectively implemented.

The result of this process was the Mandelkern Report, which proposes an Action Plan with deadlines, the implementation of which would contribute significantly to achieving the required improvements. It describes a comprehensive overall approach with a set of seven core principles: necessity, proportionality, subsidiarity, transparency, accountability, accessibility and simplicity [80]:

Policy implementation options. EU and national policymakers should always consider the full range of possible options for solving public policy issues and choose the most appropriate for the circumstances: though regulation is often the most appropriate option it should not be automatically the only choice in all circumstances.

Impact assessment. Regulatory impact assessment (RIA) is an effective tool for modern, evidence-based policy making, providing a structured framework for handling policy problems. RIA should be an integral part of the policy making process at EU and national levels and not a bureaucratic add-on. It does not replace the political decision: rather it allows that decision to be taken with clear knowledge of the evidence.

Consultation. Consultation is a means of open governance, and as such early and effective consultation of interested parties by EU and national policymakers is an important requirement. This does not usurp the role of civil servants, Ministers or Parliamentarians in the policymaking process but supplements the information they have to hand. Correctly done, consultation can avoid delays in policy development due to late-breaking controversy and need not unduly hinder progress.

Simplification. There is a constant need to update and simplify existing regulations. But simplification does not mean deregulation. It is aimed at preserving the existence of rules while making them more effective, less burdensome, and easier to understand and to comply with. This entails a systematic, preferably rolling and targeted programme of simplification, covering the regulation that impacts on citizens, business and the public bodies that have to implement it. Such programmes need to be established at both EU and national levels.

Access to regulation. Those affected by European or national regulation have the right to be able to access it and understand it. This means the coherence and clarity of regulations must be enhanced through consolidation (including codification and recasting) and access improved by better practical arrangements (especially using ICT). The former should be achieved through EU and national level programmes of consolidation and the latter through provision within each Member State and at European Union level of a public access service (either free or for a small fee).

Structures. Better regulation needs the appropriate supporting structures charged with its promotion to be successful. The best arrangement at EU or national level will depend on the relevant circumstances and charging a single unit at or near the centre with this should certainly be considered, but an effective solution must be found for each.

Implementation of European regulation. High quality regulation forms a chain from the earliest stages of its preparation through to its implementation. More attention should be paid at European level to implementation concerns to ensure that the full consequences are understood and considered. Member States should accord implementation of European regulation higher priority.

5.1.4 EU initiative "Better regulation"

Based on the Mandelkern report a series of initiatives begun in EU in 2001, aimed toward making the legislation better and this process is going on until today. During this time, the EU has developed ways and methods to simplify the regulation mainly based on administrative and political approaches. In the next paragraphs this will be described in more detailed way.

5.1.4.1 Impact assessment.

Impact assessment is designed to help in structuring and developing policies. It helps to identify the main options for achieving the objectives and analyses their likely impacts in the economic, environmental and social fields. The impact assessment system aims at helping the Commission to improve the quality and transparency of its proposals and to identify balanced solutions consistent with Community policy objectives through: [81]

- a coherent analysis of potential impacts;
- consideration of various policy choices (e.g. to use alternative instruments to 'control and command' regulation or non-intervention);
- consultation of stakeholders;
- enhanced transparency (IA roadmaps and IA reports published on the Impact Assessment website);
- executive summaries of impact assessments are translated into all EU languages.

5.1.4.2 Consultation.

Before making proposals and taking policy initiatives, the Commission must be aware of new situations and issues developing in Europe and it must consider whether EU legislation is the best way to deal with them. Therefore the Commission consults and is in constant touch with external parties when elaborating its policies. The dialogue between the Commission

and interested parties can take many forms, and methods for consultation and dialogue are adapted to different policy fields. The Commission consults through consultation papers (Green and White Papers), communications, advisory committees, expert groups, workshops and forums. [82]

5.1.4.3 Expertise.

The Commission maintains a high level of in-house expertise. As the expertise needed is quite wide and the knowledge required becomes increasingly technical and highly specialised, the Commission calls upon external specialists in different fields to provide advice. Expertise may take many forms, including both scientific and practical knowledge. The Commission has various ways of collecting expertise. Sometimes experts and interested parties are brought together in expert groups or the Commission may set up workshops, hearings, conferences or seminars. [83]

5.1.4.4 Administrative costs.

Implementing regulations and laws entails costs. They are called administrative costs. Some legal obligations to provide information have become needlessly time-consuming, excessively complicated or useless. By reducing unnecessary reporting requirements businesses can spend more time on their core activities which may reduce production costs and allow additional investment and innovation, which in turn should improve productivity and overall competitiveness. The benefits of the EU Standard Cost Model include [57]:

- bringing clarity about possible differences in procedures followed by the EU institutions and different Member States;
- facilitating cross-country or cross-policy area comparisons, benchmarking and the development of best practices;
- offering economies of scale in terms of data collection and validation.

5.1.4.5 Choice of regulatory instruments

Public authorities can regulate by setting standards, by levying taxes and charges, by financing specific actions and groups, or by providing information and advice. Forms of legislation in EU are [84]:

- Regulations and directives:
- Much of European law are directives which basically set out a result to be achieved but leave national authorities the choice of methods.

Now and then, they tend to go beyond what is required by EU legislation, adding obligations and procedures. This 'gold-plating' affects implementation of EU law as well as the quality of the regulatory environment in the EU;

- The European Parliament and the Council may also pass legislation through regulations which are directly applicable in all Member States. As these regulations require no transposition, there is no problem of gold-plating. Therefore, replacing directives with regulations can under certain circumstances be conducive to simplification.
- Review/sunset clauses. European legislation contains obsolete acts that no longer have real effect, but which remain in force because they have not been expressly repealed. As a means to prevent obsolescence of legislation, the Commission often introduces review, revision or sunset clauses in its legislative proposals, especially in policy areas of rapid technological development. Yet, alternative mechanisms, such as self- and co-regulation, may be more suitable in some cases.
- Alternative instruments. In its strategy for Growth and Jobs (2005), the Commission stressed the need to pay more attention to the use of self- and co-regulation.
- Private parties can make rules that benefit society. If for instance producers in a specific sector agree among themselves on a code of conduct to submit their products to extensive tests by approved laboratories, and label their products accordingly, they have created a form of self-regulation. Self-regulation is often developed between members of a specific profession (e.g. medical doctors, architects) but also between operators and their clients/consumers.
- Co-regulation occurs when EU authorities entrust the attainment of objectives to private parties recognized in the field, such as economic operators, the social partners, non-governmental organizations or associations.
- Monitoring and evaluating self-regulation practices and co-regulation mechanisms are essential to their credibility. The Commission and the European Economic and Social Committee (EESC) therefore have set up an internet EU self- and co-regulation database. In addition to listing EU wide self- and co-regulation cases, the database also facilitates exchange of information and identification of best practices.

5.1.4.6 Transposition and application of EU law

The timely and correct implementation of EU law by the Member States ensures that the results intended by EU policy are attained. Late or incorrect implementation can deprive businesses and citizens of their rights. The transposition into national law is done by national governments and parliaments sometimes involving regional and local authorities. At this stage, laws risk being 'gold plated' i.e. requirements or procedures which are not required by the initial directive are added. This affects both transposition and implementation of EU law as well as the quality of national and regional regulation. Monitoring of the transposition process relies on the correlation tables provided by the Member States, showing the link between the provisions in directives and national rules. [58]

5.1.4.7 Simplification

It is essential in a rapidly changing world to review laws, streamline and remove overlaps to ensure that EU legislation is clear and as unburden some for operators and citizens as possible. The EU has progressively developed a broad strategy to improve the regulatory environment and thus provide a more effective, efficient and transparent regulatory system for the benefit of citizens and reinforce competitiveness, growth and sustainable development. In its strategy to simplify the regulatory environment, the Commission uses the following methods [85, 20]:

- *repeal:* removes from the statute-book those legal acts which are unnecessary, irrelevant or obsolete;
- *codification:* contributes to the reduction in volume of EU legislation, and at the same time, provides more readable and legally secure texts, thus facilitating transparency and enforcement;
- *recasting:* is a simplification method as it simultaneously amends and codifies the legal acts in question;
- *co-regulation:* can be a more cost efficient and flexible method for addressing certain policy objectives than classical legislative tools. Standardization by independent bodies is an example of a well-recognized ‘co-regulation' instrument;
- *use of regulations:* replacing directives with regulations can under certain circumstances be conducive to simplification as regulations are directly applicable (i.e. no need for transposition into national legislation) and guarantee that all actors are subject to the same rules at the same time.

5.1.4.8 Accessibility and presentation of EU law

Two European Union websites offer free access to general public to EU law: [86]

- EUR-Lex website offers free access for the general public to the full range of EU law and treaties, including consolidated legislation, international agreements, parliamentary questions, case law, new legislative proposals and much of the EU Official Journal in all EU languages. EUR-Lex also contains a register of documents of the EU institutions.
- Pre-Lex database offers a possibility to follow the major stages of the decision-making process between the Commission and the other EU institutions starting from the Commission proposals. Commission communications are also accessible as well as various search possibilities.

5.1.4.9 Evaluation

Evaluation gives a judgment of interventions according to their results and impacts in relation to the needs they aim to satisfy and the resources mobilized. Evaluation can be carried out in a prospective (ex-ante evaluation) as well as a retrospective (ex-post evaluation) perspective, or in a combination of both. Evaluation generates relevant information that is essential for planning, designing and implementing EU policies [87].

5.1.4.10 Inter-institutional coordination

Better Regulation is a shared responsibility. The Commission submits proposals for adoption to the European Parliament and the Council. The EU laws are transposed into national law by national governments and parliaments and often applied at regional and local levels. The responsibility for regulating well is hence a shared one. The Inter-institutional Agreement (IIA) concentrates on [11]:

- improving inter-institutional coordination and transparency;
- providing a framework for alternative regulatory instruments (self- and co-regulation);
- increasing the use of impact assessment in Community decision-making;
- working methods for the adoption of proposals to simplify EU law.

A general overview of described activities is presented in Figure 7 [24].

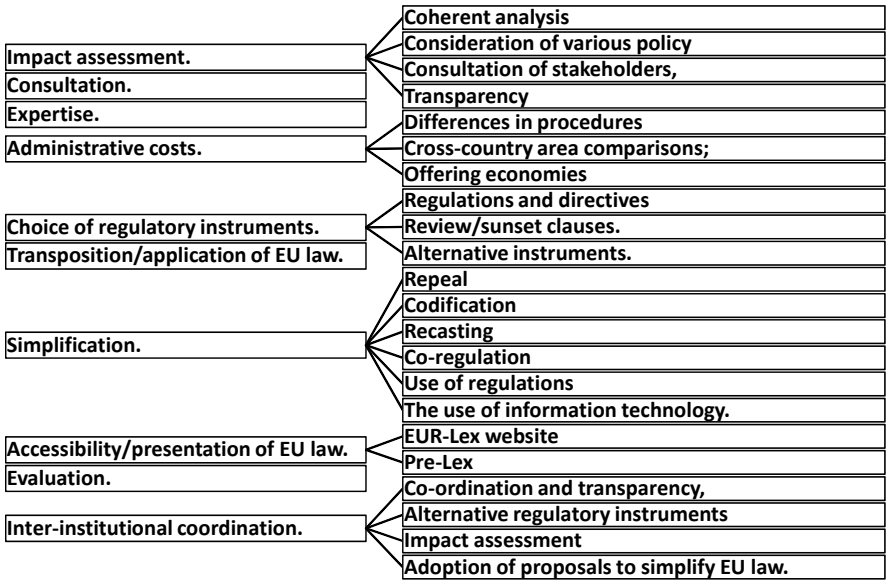


Figure 7. Overview of different EC activities within initiative "Better regulation"

5.2 How IT can help?

The past decade has shown an impressive increase of legal informatics-related works in science as well as in the public sector. Many problems can at the present time be solved more efficiently with help of computers and many promising research fields indicate the potential to offer even more solutions in the near future. The question is not any more whether ICT can be usefully exploited in the legal field but more where and how ICT can support or even replace existing approaches.

In this part of the thesis the short introduction of some of the most promising research in Europe is presented aimed to support legal specialists in their work for better regulation. One purpose of such an approach is to tie up both ends of the same issue, traditional and well proven legislative activities versus novel and innovative ICT centric aspects. Due to the complexity of the issue, these two aspects tend to drift apart so that it is hard for outsiders to keep track of the main issues. The area selected to be described in somewhat more detail is that which constitutes the main research topic of the current author, with other research alluded to briefly in order to keep the section in reasonable length.

Many efforts have been made in the last few years to investigate two problems in the analysis of legal texts: how to automatically identify structural portions of legal documents through their mutual references; and how to grasp semantic information of the legal text. Legislation usability for human beings and computers is equally under consideration. Human accessibility can be increased with help of additional structural analysis and a visualisation technique - computer usability is possible to increase thanks to the new standards and mark-up technologies.

Most of these approaches are based on XML and RDF, so it might be useful to give some background information about these acronyms.

5.2.1 XML, RDF and OWL

During the last couple of thousand years the usual, paper based document has evolved to the most convenient form considering human readers. Now, the digital revolution is introducing new challenges and one of them is to teach computers to understand open text. XML stands for eXtensible Mark-up Language. XML is designed to transport and store data and has a limited ability to add some meaning to text. It is done through text structuring and adding markings (tags) to these structural units. This forms a basis for a next generation World Wide Web- semantic web or WEB 2.0.

For example, a heading is easily recognizable for human readers within a text. It is usually separated from the rest of the text and is visually different. Despite that it is very complicated to create a good rule with help of which computers can independently recognize headings same ways as humans do. Even if it is possible to do so, it is impossible for computers to perform a reverse operation- decide which parts of text are suitable to use for headings. So the best thing to do is to divide manually text into smaller but meaningful parts- like heading, signature, date, addressee, etc.- and surround them with tags: *<heading> Chapter 1</heading>*.

RDF is used to define the structure of the data and is used to give additional semantic meaning to the structured text. If XML is a method to separate and mark textual parts, RDF provides a way to link these parts together in a meaningful way. The way how it is done is estimated like this: *Subject - Predicate – Object*. This is called triple and allows us to describe needed relationships between two entities: *Chapter 1 (Subject) is written by (Predicate) John More (Object)*. Technological presentation of it is much more complicated. For more complex relationships different ontology related methods are used like a Web Ontology Language (OWL) etc. This

kind of manually added information enriching information is called metadata.

The abovementioned method increases the usability of the legal text dramatically and therefore related technologies are one of the main topics of research in Europe and abroad. As the legislation within the EU is presented in different languages, it is structurally complex, constantly changing and sometimes internally conflicting; there is a reasonable hope that such novel technology could increase the usability of legal information a lot, not least in the EU.

5.2.2 Legislation related complexity and ways to measure it (short name: Complexity)

The research of Bourcier and Mazzega from 2007 is based on the assumption that there is a constant and accelerating growth of the national and international legal corpus and together with the rapidity of changes, the result is that it becomes more and more complex for anyone applying law to manage this process. Impacts of this complexity on the intricacy of both legal hierarchies and legal contents are considered as causing other unwanted by-effects on the intelligibility of laws for citizens, as mentioned above.

In their work they have analysed the French environmental code to understand how the drafters have organized the previous laws scattered in various fields of law into the structure of a hierarchical table of contents. Relying on graph representation, they have observed this architecture through the various levels of its organization and connections with other legal corpuses. They were also able to find some invariant distributions that shape various statistical distributions of the vertices and edges of the legal graph. [11]

5.2.3 Electronic presentation of law in METALex (short name: METALex)

XML is a next generation documentation standard created mainly for better computer processing. Information on a network that connects many different types of computers has to be usable on all of them. Especially public information cannot be restricted to one make or model or to give the control over its data format to one private body. Furthermore, public information must be possible to reuse in many different ways, which must be supported by the encoding system. This rules out proprietary data formats and this is what has led to XML use. The efficiency of managing

and processing information in legal documents can be dramatically improved by applying XML techniques. As a part of the more general idea of an integrated semantic web, documents are enriched with metadata to enable smart applications such as (intelligent) retrieval and reasoning. Various national initiatives have established XML standards for describing legal sources and documents, which have grown into projects aiming at integration and interoperability across all legal domains.

A necessary precondition for effective legal document management is the electronic availability of legal sources in a structured and standard format. Boer, Hoekstra and Winkels explain how the standard intends to provide a generic and easily extensible framework for the XML encoding of the structure and contents of legal and paralegal documents.

METALex is a generic open standard for legislative documents specifically designed to facilitate the maintenance of decision support software used by public bodies. In addition, it offers provisions for more or less traditional functionalities offered by publishers and search engines. The METALex XML schema aims to be a standard interchange format for legal documents for the purposes of presentation, description of the relations between legislative documents, search and filtering on meaningful levels of detail, and version management and file exchange.

As Boer, Hoekstra and Winkels say, the classification level presupposes that the user of the classification system can read the document to find out why the classification was attached. They say that even if such domain classification schemas, mainly predating the use of computers for storing legal information, they are not necessarily adequate for electronic use. In such classification the attributes used tend to be fairly traditional like author, creation, modification and promulgation dates, jurisdiction, legal status and language. Boer, Hoekstra and Winkels find such attributes rather crude in meaning, without a lot of relevant detail and the usefulness questionable for automated reasoning thus is questionable. Some issues they point to are that identification of documents by jurisdiction assumes that the user of a search service knows what jurisdictions he is in and the specific XML standard of that country. This standard comes from the legislative style and language of the country although in the EU there should be a common standard. It can be explained so that the EU provides the frame that can be filled by each country. [55, p. 2]

Boer, Hoekstra and Winkels as an alternative to rigid domain classification, suggest that statements can be directly identified in the contents of a document. Documents can make statements about other

documents and (fragments) of the document itself: the metadata on one document is distributed over different locations. For this, there is the Resource Description Framework (RDF). [55, p. 3]

5.2.4 Syntactic analysis

Brighi, Lesmo, Mazzei, Palmirani and Radicioni stress that hand-made annotations in law or other texts are time-consuming and error-prone, alternative tools for modelling and representing the structure and content of norms are needed and in the view of these authors such tools could greatly benefit from automatic approaches to extract both structural and semantic data from legal texts, conceivably generating XML output.

They concentrate on the annotation of modificatory provisions. In any legal system one norm refers to other norms, either for completing its own content or to change these other norms in some way through a modificatory provision. Even if legal language is stricter, or as Brighi et. al. put it, more controlled than ordinary language, tracking modifications requires considering the deep syntactic structure of sentences and encoding prior knowledge on possible modifications (content and how they practically may occur). For the automatic approach before any semantic annotations are added, the text is marked up with structural data as well as with normative references and "quotation mark" elements, i.e. pieces of text referring to a passive norm (such as any additional or replacement text, along with a string indicating where this text belongs in the passive norm). Semantic annotation enriches a text with the <mod> element, which delimits a modificatory clause and with the metadata that fully qualifies the modification and its attributes according to one of the following classes identified by the authors referred to:

"Type 1. A change made to the actual text or form of the norm (an integration, replacement, deletion, relocation) or to the meaning of the norm (an interpretation or variation of meaning or a modification of clauses);

Type 2. A change made to the range of a norm (an extension of its subject matter or range of application or a provision stating a derogation to it);

Type 3. A change made to the temporal parameters of the norm (the time of its entry into force, and the time when it becomes applicable or effective);

Type 4. A change made to the status of the norm within the legal system (a decree-law that is made into law, an international treaty that is transposed into domestic law)

Type 5. A change made to the powers conferred under a norm within the legal system (e.g. an EU directive transposed into domestic law". [88, pp. 203-204]

5.2.5 Multilingual drafting

The aim of multilingual drafting initiatives is to improve the quality of the legislative production, to enhance accessibility of legislation at European level and to promote awareness and democratic participation of citizens to the legislative process. This is essential in a legal system such as that of the EU with a large number of different languages but where still the law has to apply in the same way in all Member States.

This is the background to the project "Drafting Legislation with Ontology-based Support" [89], which aims at ensuring that legal drafters and decision-makers have control over the legal language at national and European level. What the project does is to provide law-makers with linguistic and knowledge management tools to be used in the legislative processes, more specifically in legislative drafting. The tools make sure the terminology is properly understood to its deep legal meaning. DALOS uses an ontological characterisation of legal language, giving conceptual meaning to the lexical units and providing connection with other terms. The combinations provided make the legal language easier to work with.

In legal language every collection of terms emanating from a specific language and a specific legal system is an autonomous vocabulary resource that can be mapped through relationships of equivalence with other systems. Words are the main tools for lawyers and the use of the correct terms as well as the correct combinations of them are of paramount importance. The best approach to map terms and term collections consists of developing parallel alignment with the same methodology and the same conceptual model.

DALOS shows how different methods may be applied, depending on the characteristic of the domain, the data structure and on the results sought. The project highlights how among structured data different degrees of formalization can be distinguished: controlled vocabularies (such as thesauri, classification trees, directories, keyword lists), semantic lexicons as well as foundational, core, and domain ontologies. Agnoloni and others show that the integration of lexical resources (heterogeneous because belonging to different law systems, or expressed in different languages, or pertaining to different domains) can lead to different solutions depending on the desired results:

- generate a single resource covering both (merging);
- compare and define correspondences and differences (mapping);
- combining different levels of knowledge representation, basically interfacing lexical resources and ontologies.

It is shown that the methodological approach chosen in the DALOS project is the third one: it requires the definition of mapping procedures between semantic lexicons. This is driven by the reference to an ontological level where the basic entities, which populate the legal domain, are described. Such an approach has been followed to obtain correspondence between terms of different languages as well to align corresponding terms towards a common conceptualization at a higher knowledge level. [66]

5.2.6 Legislative Meta-Drafting (short name: Meta Drafting)

The semantic mark-up of legal texts calls, first of all, for the development of suitable sets of meta-data, supposed to capture the formal structure of the legal text, as opposed to its content, as stressed by Biagioli and Grossi. Such meta-data need then to be systematically interconnected, to reveal the semantic structure underlying the mark-up.

Biagioli and Grossi start from the presumption that legal orders are perceived as accumulated sets of laws, created through a dynamic process. Legislative archives reflect the historical organisation of the legal order and the law is the documentary unit on which the archive is created. The lack of an analytical/systematic vision of the whole is an obstacle to obtaining information about and exercising control over the contents as a whole. [90, p. 2]

For Biagioli and Grossi the meta-data needed for legislative meta-drafting can be obtained from suitable theories of provisions. Such theories according to these authors make explicit the functional structure of the legal text. They divide provisions in the model into two main families: rules and rules on rules. They call the rules on rules "*a peculiar category that includes the provisions related to the dynamics of a legal system*". Further, the main family of rules is divided into a further two major classes, which are the subject of normative theories - constitutive rules and regulatory rules. Biagioli has thought up this approach, which has become popular. The text is not divided into paragraphs and sections but into so called provisions: small pieces of text, without a strict structure. [14]

5.2.7 Content management

Boer, Winkels, van Engers and de Maat have developed a content management system. Their starting premise is that governments as well as legal publishers usually make legislation available in some special purpose XML format or XHTML annotated with metadata describing what version of which legislation it is. As they put it, the electronic documents containing the legislation are taken from largely autonomous and heterogeneously organized repositories. Versioning metadata is closely tied to the organization of the repository it came from. [91, p. 19]

To understand the relationship among multiple metadata descriptions it is first necessary to understand the resources they purport to describe. Understanding the resource in turn requires a comprehension of its lifecycle including events and corresponding transformations of the resource that constitute this lifecycle. This leads to the conclusion that an electronic document and its metadata description represent a certain time point in this lifecycle. It also means that values of metadata attributes may change over the lifetime of a legal document, even if the document itself does not change. Metadata most often concerns information included in the document itself, or in another document that refers to it, so it is not added extra information. [91]

What is called Resource Description Framework (RDF) is used to store a description of the events that lead to the resource with the attached metadata. In this way RDF and related technologies helps the computer to make generalisations: it can understand that if the object is a girl it is also a woman, a human being and so on. Thus a wider range of norms can be applied.

One main interest in proper content management systems is that commentaries from different sources should be available in internal applications for civil servants or others who need them in their work, but this is complicated because they are all organized in a different way and frequently updated, and sometimes even reorganized, by the publishers. Versioning, structuring, and naming practices are different between different publishers, but not because the content requires this. This makes it difficult to establish the exact identity of the legal sources contained in a product without continuous human intervention.

Among shared content elements of legislation in different jurisdictions in regard to which improved handling of metadata can be useful is that what Boer et. al. calls ‘time points’. This is a simple date, but one can distinguish

three kinds of timestamps used in legislation as outlined by Boer and others [91, pp. 20-21]:

Version management timestamps define the validity of the document for reference as well as the validity for application - the document can be applied by a competent decision-maker in the time-interval in which it is active.

Legislative Drafting timestamps relate to the procedures that have to be followed by the legislator including timestamps for certain events (e.g. signing) or minimum time intervals that must elapse before some consequence follows.

Application to cases: These timestamps define objects in the outside world that the legislation refers to, which can be immediate events that take negligible time (like traffic events); persistent 'objects' (mortgages, pension arrangements); or delayed payoffs of choices (financial products exempt from certain kinds of taxation). Transitory regimes are needed to minimize any damage caused by changes in legislation and legal principles like limits on retroactive application belong in this category.

Another important component of the structures developed by Boer et. al. is the lifecycle of legislation, which they divide into four phases [91, pp. 21-22]:

Fixed: This indicates the point in time when a design of legislation becomes an official proposal that cannot any longer be modified by the drafters in the regular way (by opening the document in an editor, changing it, and saving it). In the legislative process this will be an event linked to formal signing or confirmation, after which a new formal decision would be needed to change the text of the proposal. The authors (Boer and others) quote the example of the Netherlands, where this point is the date of signing of legislation by the monarch and where this point is furthermore characterised by auxiliary provisions like date of publication, date of enactment, official name of the legislation, official acronym of the legislation, and delegate legislative competence, becoming valid law.

Knowable: The date of publication is also the date from which the legislation is (presumed to be) publicly known. This date is usually prescribed by law either so that the legislation sets a date (specific or a general one – how many days after some event) or so that publication is relative to some other event (like the publication of another closely related law. This is an alternative date at which any such auxiliary provisions as those mentioned above may become valid law. From the date of publication,

legislation may be applied and only to events happening after this point in time (with some few exceptions for retroactive or delayed application).

Repealed: An end-point in the lifecycle of legislation is when legislation is repealed. This date is usually announced by law as a specific date or relative to some other event or after some fixed time interval. If legislation is repealed, for document management purposes it does not 'disappear': it can still be referenced by its official name and acronym.

Not all jurisdictions distinguish between the 'Fixed' and 'Knowable' dates, and this distinction is therefore not part of METALex XML. Events have input and/or output, and if they are actions they have an actor in a certain role (e.g. legislator) and optionally instruments (e.g. a legislative competence). Boer and others divide into five types of events that can cause a transition that bring a new version of (a part of) a legislative text into being:

Fix: Initiated by the legislator, based on legislation attributing legislative competence, having as its output fixed legislation.

Publish: Initiated by the same legislator, using a publication channel as an instrument and possibly legislation requiring the publication. Here one can see input (fixed legislation) as well as output (published legislation).

Enact: Initiated by the legislator, using as instrument legislation (or in all cases, a part if legislation) requiring enactment. Input - published legislation; output - enacted legislation.

Repeal: Initiated by the legislator, the instrument is legislation requiring cancellation of the legislation. Input is enacted legislation and output is repealed legislation.

Modify: Initiated by the legislator, the instrument is legislation requiring modification by replacing its text with text quoted in modifying (or possibly external, legislation). It is most common that parts of a legal act are modified, but it may be done also by inserting or removing an article. Input is legislation and output is amended legislation.

Boer, Winkels, Engers and Maat propose a novel use of such an event-based framework with legislation in an instrumental role. All relevant dates are attached to the event (if there is a corresponding source that contains the specific date, it is in external legislation in the instrumental role and not the legislation to which the date is usually attached as metadata). [91]

5.2.8 STIA

In research linked to semantic annotation in jurisprudence the starting point is the spread of norms and laws containing explicit cross references or

overlapping concerning the same or similar topics. This has entailed various actions for legislative simplification, not so much to reduce the amount of sources, but even more importantly a necessity in order to achieve the correct application of normative principles. These should be pronounced, discussed and dealt with as monolithic utterances instead of being spread across several distinct codes as stated by Pazienza, Scarpato and Stellato. [92, pp. 156-161]

These authors take as their example Italian legislation that is composed of more than 100 000 different acts, but this is not unique in any way for Italy but in any document collections in jurisprudence the cross referencing assumes huge proportions. The biggest problem in this case is to retrieve useful information in such enormous collections in relatively short time.

What Pazienza, Scarpato and Stellato suggest as a first step on the way to legal simplification is to identify relations of pertinence between distinct laws. Such identification allows for these laws to can be unified and reproduced in new synthetic codes. Information Retrieval is typically used to retrieve relevant information from document collections. Matching queries and documents is generally term-based: words within documents are used to describe the documents and to determine their relevance for a given query. This is the simple way of doing it that does however not show up other links than the simple correspondence between words. For a legal practitioner the real correlation between acts may not be shown in this manner. [92]

To introduce more information about the meaning of a document, semantic annotations can be added, containing additional information about the text or part of it, that are important to improve retrieval processes. In recent years, collaborative tagging systems have become very popular among users as a means for organizing their resources. These systems use semantic annotations taken by users to improve retrieval by using the information held into them.

The knowledge model of the framework extended by STIA offers two concept layers, consisting in: the application layer, containing ontologies from the project called Semantic Turkey and its extensions, which are necessary to drive the application, and the user layer, containing specific domain ontologies and allowing the user to add instance data. [92]

5.2.9 Legal Change Management

The final area of research to be described is what can be called legal change management. Palmirani and Cervone have pointed out that not only are

there many national and international XML standards for modelling and representing legal resources but furthermore for modelling norms by way of rules, there are very many different standards. They find a need to collectively manage heterogeneous legal resources that use different XML formats. [93, pp. 146-155]

There are good ways to answer this need (Metalex etc). Unfortunately they are not sufficient for managing a common query layer between heterogeneous collections of XML legal resources and providing the semantics with which to manage change over time. Therefore there is a need for developing a repository architecture, which is capable of managing heterogeneous documents through common document ontology and metadata mapping, while managing document change over time. A native XML document database of mixed resources has to be developed, which is able to maintain the flexibility and the expressiveness of the original local standard while also providing and interchanging a solution between several XML standards – all the while also building a solid common basis for making meta-queries between mixed documents. [93]

5.3 Conclusion

ICT in general can contribute to the usability of legislation in many different ways (in coherence with the criteria's stated in the beginning of this section) [54, pp. 5-6]:

- Means-ends rationality- to align legislation to its intended purposes by providing legislators with tools for anticipating the impact of new laws on the legal system, on administration, and on society or for monitoring and evaluating impacts of an existing law;
- Discursive rationality- to promote critical debate around legislation, by providing communication tools for promoting informed debate within parliaments and outside of them, by facilitating the preparation of legislative proposals, and by offering citizens and their associations new ways to participate in the legislative process;
- Responsiveness to citizens' need and preferences- to facilitate the contact between citizens and their representatives, providing citizens with new ways to express their views and with feedback about the choices of their representatives;
- Legal certainty- to enable citizens to anticipate legislation's impact on them, by providing access to laws and cases, by facilitating the drafting of more understandable regulations and the maintenance of the legal systems understandability as much as possible;

- Citizens' rights- to contribute to ensuring that laws protect citizens' rights by making knowledge about rules and remedies more accessible and by ensuring publicity of information about officials' behaviour.

There are some other aspects where legislative informatics can give a valuable input. As being a catalyst of many changes in society, ICT is a challenge in many ways:

- ICT is the infrastructure of globalization to enable economic and social networks transcending borders;
- ICT is the engine of economic and social development to increase the speed of change;
- ICT is the enzyme of complexity: by increasing knowledge and possibilities available to individual actors they increase the complexity of their interaction.

From the opposite side, ICT can offer many solutions for the society:

- ICT can provide an infrastructure for legislative networks in order to enable legislative coordination to address global issues;
- ICT can support dynamic drafting: ICT enables to control the mass of legislative material, and to intervene in it making the required changes;
- ICT can help legislators in modelling laws according to social complexities.

There are a lot of things happening in the field of legislation simplification and related ICT research fields. Two big powers in the world- USA and EU are seriously fighting back the increasing complexity with help of traditional (manual) and not so traditional means, reducing the legislative burden upon society, businesses and citizens. There are also a lot of activities in the European academic societies to support this. Nevertheless the research is still at an early stage and seems to focus on legislation presentation (XML, semantic or language related) and is not in a position yet to solve more substantial challenges, defeating the complexity in general for example. How to find efficiently the right document among the myriad of legislative acts applicable to the presented situation; how to make this process quick, easy and enjoyable as we expect it to be when performing internet search- this remains unsolved and this is the main research question of the present research.

If to try to conclude the long answer to the question in the heading of subdivision 5.2 "How IT can help?", then perhaps *Table 8* at page 122 would be the best source for that. Problems described in subdivision 5.1.4 and

briefly referred to in Figure 7 at page 64, form the left column of the table; research initiatives described in this subdivision form the heading of this table. The purpose of the table is to evaluate the potential of these research activities at different levels regarding the specific problem. These results were published in article [24]. As this table includes also analysis of the research approach described in next two chapters of this thesis, it has been moved to the conclusion part.

6 Legislation content capture and visualisation method description

In order to find a solution to the abovementioned broader purpose statement of the thesis:

- *to increase the usability of legislation;*

which is divided into next specific targets:

- *to find and use a suitable visualisation;*
- *to find suitable methods and techniques in order to perform the visualisation approach;*

an interdisciplinary method is introduced. It involves methods used in engineering, language technologies and information system analysis. Information system analysis approach formed the framework for combining these different methods. Language technology was used for information extraction, reengineering for splitting and restructuring the legal text. One of the purposes was to redesign the legal text to make it better suited for machine computing.

From this point forward the thesis is focused on original research [19, 26, 25, 24] though it is based on efforts of many researches from different fields (Aarnio, Biagoli, Aggarwal, Võhandu etc.). The first subchapters of this chapter (6.1, 6.2, 6.3, 6.4) are providing the user with presented method specific theoretical background while the subchapter 6.5 briefly concludes the main aspects of this method. The Chapter provides examples of use, applying the method on legal text to show the visualisation results from scratch.

Very briefly the main steps characterizing this method are:

- to find theoretical foundations for system decomposition;
- to find proper methods for system decomposition;
- to estimate the elementary, atomic unit usable for system decomposition;
- to estimate the way to capture the main content of the estimated elementary system unit;
- to estimate the way to visualize this content.

6.1 Reengineering, restructuring and business modelling

Reengineering and restructuring are necessary for decomposing the existing system without demolishing it. If we are able to do so, we are able to put it back together without loosing the valuable pieces and possibly get a new quality out of it. These new qualities usually reveal something new about the existing system and indicate the ways how it could be exploited further.

A need for restructuring is well known by specialists in the technical field but has been developed to high level applicability by computer scientists due to the extreme complexity of the software systems. According to Chiofski and Cross, the term "reverse engineering" has its origin in the analysis of hardware. Reverse engineering is regularly applied for improving the products, but also for analysing the competitors work. [13]

"Reverse engineering is the process of developing a set of specifications for a complex hardware system by an orderly examination of specimens of that system. It is a process of analysing a subject system to:

- Identify the system's components and their interrelationships;
- Create representations of the system in another form or at higher level of abstraction. " [13]

Reverse engineering usually involves a finding of small building blocks of the system in order to build or recreate solutions which do have different characteristics compared to the original one. It is tied only with observing or examination but not changing the investigated system. Restructuring is the term for creating a new system with qualitatively new characteristics, using the same building blocks but preserving the subject system's external behaviour (functionality and semantics). This makes the process suitable for applying on the legislation as the research object is a "living" system- a system which is in everyday use and quite literally the lives of people may be at stake.

A restructuring transformation is often about appearance, such as altering code to improve its structure in the traditional sense of structured design. However, the term has a broader meaning that recognizes the application of similar transformations and recasting techniques. There are many ways to perform the restructuring which can be performed with acknowledgement of structural form but without understanding the real meaning of it.

Reengineering is an analysis and change of the subject system to give a new form to it and implement it subsequently.

There are different kinds of purposes for applying the reverse engineering. The main purpose is to increase the possibilities for

maintenance and for new developments. There are six main objectives for reverse engineering [13]:

- *Cope with the complexity*- to develop methods to better deal with the sheer volume and complexity of the systems. A key to controlling these attributes is automated support;
- *Generate alternate views*- graphical representations have long been accepted as comprehension aids;
- *Recover lost information*- the continuing evolution of large, long-lived systems lead to lost information about the system design;
- *Detect side effects*- both haphazard initial design and successive modifications can lead to unintended ramifications and side effects that impede a system's performance in subtle ways;
- *Synthesize higher abstractions*- reverse engineering requires methods and techniques for creating alternate views that transcend to a higher abstraction level;
- *Facilitate reuse*- a significant issue in the movement toward software reusability is the large body of existing software assets. Reverse engineering can help detect candidates for reusable software components from present systems.

This method has been considered as a suitable approach for further application for the current research because it has many similar purposes- cope with complexity, generate alternative views, etc. **Therefore a question arises: what is the smallest part of the legislation, which could be used for coping with the complexity, generating alternative views and synthesizing higher abstractions?**

Modelling is a common task in Business Process Modelling that aims to visualize the structure of complex systems. The method allows modelling any system using main activities, events and objects. This method was applied, declaring that verbs within legal text refer to events and activities, nouns to objects. *Measuring the concordance - coexistence frequency- of meaningful parts of the system it is possible to create very rough business process models of legal text, represented as a graph.*

6.2 The norm

In order to create a better structure and knowledge extraction methods one must find a way how to reorganise or restructure the legislation itself. This task should be handled with extreme caution- legislation is used daily and cannot be changed due to technical demands without having tremendous negative effect on society. There has to be a way how to do this without

disturbing the traditional, well established and widely accepted organisation of current legislation.

As a result the search for atomic unit or the smallest unit of legislation began, hoping that a better solution based on building bricks of the legislation can be created. There are several possibilities of doing it; the most obvious one is the linguistic approach. ***It is rather logical to divide a legal text into paragraphs, sentences, words or even into letters***. Yet the natural language has the same origin as legislation- these means are specifically designed for human interaction and therefore hardly exploitable by machines. Human logic is often too loose to describe it with machine-oriented binary order. Nevertheless, this is an opportunity that has opened promising prospects, especially focusing on the words.

Law is often seen as a collection of norms, guiding everyday life. From this point legislation is consisted from individual small particles- norms, which are grouped based on similarity and combined together in the form of legal texts. According to Gibbs, there are different definitions of norms, illustrating the differences and certain points of agreement [94]:

"A norm, then, is a rule or a standard that governs our conduct in the social situations in which we participate. It is a societal expectation. It is a standard to which we are expected to conform whether we actually do so or not."

"All societies have rules of norms specifying appropriate and inappropriate behaviour, and individuals are rewarded or punished as they conform to or deviate from these rules. The norms are blueprints of behaviour, setting limits within which individuals may seek alternate ways to achieve their goals. Norms are based on cultural values, which are justified by moral standards, reasoning or aesthetic judgement."

"A norm is a statement made by a number of members of a group, not necessarily by all of them, that the members ought to behave in a certain way in a certain circumstances. "

"A norm is an abstract pattern, held in the mind that sets certain limits for behaviour. An "operative" norm is one that is not merely entertained in the mind but is considered worthy of following in actual behaviour; thus one feels that one ought to conform to it. This feeling means that one "accepts" the norm. "

It is obvious that the norm is a complicated term to understand and use. Norm lies somewhere between the ideal picture of the world -how it should be - and reality - what there really is. By some researchers there is no such thing as a norm as an ideal entity. There are only individuals who behave in

a certain way. According to Aarnio, a norm has an **"*ideal existence in the same way that general concepts exist in the ontology of conceptual realists*".** [16, p. 26]

In the field of AI and Law the use of norms for legal text analysis is widely accepted. The technological view of the norm is understandably a bit different, according to Boer, Winkels et al. : "*The norm is an epistemological concept identified by its role in a type of reasoning and not something that exclusively belongs to the vocabulary of the legal domain. A norm is a standard of performance, a measurement scale that defines what is normal or normative. It is used whenever we are assessing something, regardless of domain....).*" [95]

One well known approach, offered by Biagoli, uses norms as the natural part of the legal system and builds on it the functional micro-ontology of law [14]. According to author of this approach, a legislative text may be seen as a vehicle that contains and transports rules and the legal order as an organism of rules rather than of laws, enabling us to observe its contents better. Rules (norms) are, therefore, seen as the true foundations of normative systems, whilst laws are merely temporal.

Applying his proposed method, one can benefit from the following aspects [14]:

- Adopting an entity of reference more analytical than the law (rule);
- Viewing the legal order as a system of rules and the text as a set of rules;
- Describing the nature and functions of rules (types);
- Identifying significant and essential aspects of every type of rule (attributes);
- Identifying privileged relations between types of rules (nuclei of rules).

Biagoli argues that this perspective, inspired by analytic legal philosophy, permits us to perceive the rules as the true bricks in the legal system, and the laws as purely temporal events. ***Norms are considered to be an elementary unit of the legislation, helping to structure and understand the legislation better.***

6.3 Symbiosis of text and legal norms

For this research it is important to emphasise the textual representation of norms. It is valuable to distinguish the legal text and norms buried inside the text. In order to pinpoint the relevant part of the normative system, one has to locate the norms within texts. Usually the objective is not so much to find

the text but rather to find a rule or guideline to apply or follow. Therefore the text search includes the specification of the expected norm and locating the acts including the specific norm under the investigation.

It is important to emphasise also that legal text is a combination of sentences and words. Legal text is expected to be clear, simple and precise [96]. Yet these are demands, which are hard to follow. ***Norms are expressed in natural language***; the use of words is not as constant or systematic as wanted, these represent a general rule needed to be interpreted by individuals in certain ways etc. The process of catching the thought content of the norm is not straight-forward at all, including several steps of high-level human abstraction. After all, laws are written by humans for humans, therefore the use of machines is highly complicated. Therefore the beneficial services offered by computers have their natural limitations.

In order to deal with such ambiguity this research was focused on finding the features which make the norm somehow recognisable for computers. It is known that the norm is expressed in natural language, presented grammatically, semantically and correctly in order to be fully readable and understandable. ***So the norm is made from words and in order to recognise the norm, one has to recognise the words first***. It is quite logical, that similar norms are made from same words.

"Once a specific expression for a concept has been introduced, the same expression should be used throughout the text. Moreover, a specific fact should be described from the same point of view...." [96, p. 139]

Therefore it is correct to assume that the existence of the same words in the different norms can indicate the existence of the similar norm and this similarity can hopefully be measured.

In order to begin with norm location search within the normative system, the exact relation between norms and legal texts had to be specified. On the one hand, the legal text is a textual representation of norms. On the other hand, the legal texts and norms are not the same- according to Aarnio one is dealing with thought content of the norm, expressed in natural language in written format [16, p. 26]. This is important for us to acknowledge, that the written norm is a subject for further processing: decisions based on norms should be applied to the actual situation, verified for suitability and interpreted according to the thought content of the norm. In this, it is not possible to talk about fully computerised norm extraction or automatic legal reasoning until it is possible to fully copy human intellect by computers. So far it is possible to talk only about more or less computer assisted support for legal decision-making or legal search.

The norm is expressed in written format and using natural language. Text is used to express the product of the human mind and has its well-established mechanisms to do so. Text is usually constructed from paragraphs, paragraphs from sentences, sentences from clauses. Further decomposition is unnecessary, as words and letters do form the basis of the norm but are not capable of expressing the content meaning of the norm itself.

Legal text has its own specific limitations on their organisation and formulation because of the very function of such texts .They have to be structured in small units, articles and paragraphs, which can be connected by overt linguistic means only to a limited degree. Legal texts may consist only of normative statements and legal statements must be abstract and general. [96, p. 129]

It is justified to assume that the smallest meaningful representation of a norm in legal text is a clause. By definition the clause is a group of words containing a subject and predicate and functioning as a member of a complex or compound sentence. [97] By other definition the clause is a grammatical unit that:

- includes, at minimum, a predicate and an explicit or implied subject;
- expresses a proposition (meanings of sentences).

By first and second definition there must be a subject and predicate. The subject is a noun and the predicate is a verb [98]. The clause is containing a norm meaning, probably the shortest form for it. In special cases, the sentence is the clause.

6.4 Role of verbs and nouns

In order to recognise the norms within the legal text one can concentrate on finding the words, which carry the meaning of the content. Some words do not have a specific own meaning - conjunctions for example. Some carry it to a bigger extent- verb is a good example. There is little use in reading just the word “and”. Much more information is hidden in the single word like “fraud” or “manslaughter”. In suitable content last examples do not need much explanation at all- hearing the word “fraud” in courtroom or word “manslaughter” in the sight of crime scene. Therefore the contextual awareness enhances our ability to acquire and understand information.

According to this, one can use meaningful words for the automatic text content capture task. As the purpose is to find similar or related norms within the normative system, one can use these words for locating and representing parts of the normative system in connection with other parts and visualising a logic pattern of norms within the legal text.

Table 1 Frequency of word types usage in Estonian legal texts

Word type	**Frequency, %**
Nouns	47
Verbs	13
Numerals	12
Conjunction	7
Adjective	7
Pronoun	4
Adverb	4
ordinal numeral	3
Names	2
ad position	2

In order to check this theory a short analysis was done based on randomly chosen legal acts, all together 51036 words. Analysis revealed the following frequency of words usage (Table *1*):

For comparison reasons a Bible text was analysed the same way (569816 words) and the picture there is somewhat different (Table 2):

Table 2 Frequency of word types usage in Bible (Estonian language)

Word type	**Frequency, %**
Nouns	23
Verbs	21
Pronoun	18
Conjunction	12
Adverb	8
Names	7
Adjective	4
ad position	3
Numerals	1
ordinal numeral	0

Based on these findings only two types of words were chosen in order to use them for further content capture of the texts: nouns and verbs as these carry the “essence” of the norm. These word types are representing the legal content statistically 60% in sum, but one must consider also the informational “value” or “weight” of these words. Words and nouns do carry much more information in comparison with conjunctions or numerals.

Therefore it is assumed that relying only on these two word types it is possible to capture the content of the norm in range 60-80%.

6.5 Short summary of the method

To conclude all of this – legal system is the set of norms, the smallest presentation of a written norm is the clause and the clause consists at a minimum from nouns and verbs.

Yet the extraction of meaningful words is much more informative if it is possible somehow to relate these words to each other. Just a random choice of more frequent words does not give any picture about the regulation. ***Therefore the meaningful words are connected together by the common concept of the norm and the linkage between them should be shown.*** Two words- noun and verb in the same norm carry the same meaning; they are connected to each other through the norm content meaning. ***The clause therefore is acting as a natural language container, binding together norm specific nouns and verbs***. The relative importance of the link can be illustrated with help of connection "weights", which in the current case represent the frequency of use these word pairs within the clause. [19] Computationally this means a formation of two-way tables, consisting of nouns and verbs with frequency of found word pairs. ***Verb and noun concordance two-way table representation can be used as a way to show the content based on two types of words and their relationships.***

To illustrate all of this we take a norm from the Estonian Constitution § 20:"*Everyone has the right to liberty and security of person.*" [71]. With a use of suitable text analysis software it is possible to estimate automatically the clause (in simplest case after text mark "point" to the next "point"), the type of words (nouns, verbs, numerals, conjunctions etc) and count the frequency of noun and verb pairs presence as shown in Table 3.

Table 3. Two way table representing the content of the norm

Noun/verb	**has**
everyone	1
right	1
liberty	1
security	1
person	1

In order to enhance the further use of such tables and especially in case of Estonian language there is a good reason to establish the basic form of

the words and form the table based on basic form words. Estonian language uses a lot of different cases per word, depending form the situation (word "mees" (man) is represented as "meest", "mehele", "mehel", "mehest" etc) and therefore the use of basic form has a very practical, unifying reason. This does not reduce the usability of tables (on the contrary) but makes it more general and comparable to others. ***It is relatively easy to keep the two-way table connected to the original text to find the original presentation of the norm within legal text in case of necessity.*** Each number in the table refers to the actual norm where it was derived and technically it is easy to "remember" it (using unique id for clauses etc).

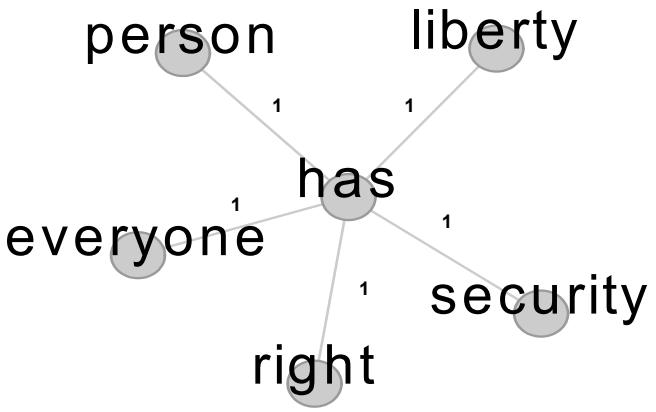


Figure 8. Sub graph of the Constitution norm "Everyone has the right to liberty and security of person"

As an analogy, this method is similar to the digitalisation of geographical maps, where the actual surface of the territory is turned into computer readable format (heights, lengths and so on) which makes connections between points of interest relatively easy to find, distance between them estimated etc. Though the landscape digitalisation process does not capture the information about geographical objects fully, it still allows us to navigate successfully through the landscape, plan buildings, gather necessary overview for further decisions, decide over ownership rights etc. A powerful moment within it is the possibility to find/define/add more layers in case of necessity, in order to add depth and different aspects to the analysis. This can be done very conveniently on the digital map applications where different layers (buildings, communications etc) can be added or removed in order to describe the landscape from the suitable angle.

Similarly to the map layout, such presentation allows us to form a visual backbone of the constitutional norm, a sub-graph of the example (Figure 8).

Yet it has to be assumed that the use of specific words in legislation is not constant or unified to the full extent. Sometimes the changes of the word usage or meaning are caused due to political reasons (political correctness, change of policy etc.) or words have been used in different times and conditions resulting in the shift of the actual word meaning. For example, the meaning of the word "marriage" is quite different than it was two hundred years ago. Today it can involve a couple of the same sex which was impossible before. ***Therefore the amount of the "actually captured truth" with the help of proposed method is not 100% and this method should be used bearing in mind this fact.***

7 Test cases

A next challenge for this research was to find a way to prove the previously described method of legal text content capture. If it is possible to extract up to 80% of information from the legal texts this can be used for further operations and derived results should be convincing. As the method is new and no analogue could be found, the author of this thesis had to come up with a way for verifying it. In order to test the abovementioned method and its practical applicability according to the fourth and fifth purpose statement:

- *to exploit visualization methods in order to test the chosen methods;*

a row of experiments were performed. As the author of this thesis is not expert in the legal field a critical response from legal experts was considered as necessary. This formed the background for the last purpose estimation:

- *to get an evaluation from legal experts.*

Therefore gained results were given to the legal scientist for further evaluation. These results were published also in Jurix 2011 conference in Vienna, in the frames of workshop on "Fundamental Concepts and the Systematization of Law". [26] The rest of the text within this chapter describes the evaluation feedback and some interesting outcomes of it. In the end of this chapter a large set of possible method application areas is described briefly.

In order to prove a proposed content capture method, several test cases were performed and analysed, based on Estonian legislation in Estonian language. These tests involved mainly top levels of the legislative pyramid (Figure 4), mainly constitution, constitutional acts and legal acts. These layers were chosen due to their clarity, as the role and place for other similar layers like international treaties, EU level legislation etc. is no yet very clear at the moment. The biggest test case analysis involved almost half of legal acts full texts, all together there are ~800 of them at the moment. These were chosen randomly and analysis operations applied "blindly" to secure the most objective outcome.

The following aspects were investigated for this reason:

- Information-enriched graphical representation to estimate the value and find preliminary properties of scalability, visualisation and mind mapping;
- A prototype was built in order to estimate the search computability, functionality and usability aspects;

- Similarity measurements of legal documents in order to perform the comparison between documents
- Experimental data mining approach based on similarity measurement results was performed in order to reveal the legislation structure and restructure the whole legislation.

The results described in this chapter were published also in the Jurix 2011 conference in Vienna [25].

7.1 Scaling and visualisation approach

First experiment was done in order to test method usability for different visualisations. As the method is supposed to help users to orient within legislation and give them some high level understanding about the content and location of it, the visualisation of the information was chosen as the best way of doing this.

Visualisation is an important part of this approach as it allows the user to get a high-level overview of the legal act; legal acts groups etc. before even starting to read the texts. Even for the experienced user finding a suitable legal act to start the search (or the entering point to the legal system) can be challenging, not talking about non-expert users. Therefore a representation capable of guiding the users, giving hints and suggestions where to look for and releasing them from the burden to read all the texts in very many legal acts could be useful. A vision of legislation as a connected set of different related documents and presenting important parts inside these could answer better to the needs of the modern networked society.

Visualisation method application can be briefly described by the following steps:

- legal text processing in order to:
- estimate the borders of the clauses;
- estimate the type of words within the clauses;
- estimate the basic form of the words;
- count the usage of verbs and nouns within the same clauses;
- form the two-way table, showing the concordance of nouns and verbs within clauses;
- transform the tables into visual layouts;
- reduce graph in case of necessity to get an overview about the most important part of it.

7.1.1 Visualisation of legal semantic network

The visualization approach is motivated mainly by the idea that in the legal domain there is a huge amount of explicit and hidden connections and relations, making all-together a huge semantic network. The lawyer usually has developed his own understanding of the network of relations in his or her mind, which he or she is using for the job. Such information is available only for this and is hardly reusable for other participants. Such a ”mental picture“ is not overwhelming and does not cover all the aspects of legislation because of the huge amount of relations and connections inside the legal domain. Simple visualisation can create a similar picture of semantic networks of legal objects, buried inside the legal domain. Adding relevant information to the picture (metadata like dates of validity, locator, type of relation, issuing authority etc.) allows us to enrich the picture to the level where the user can successfully browse through different levels of legislation, from top-down and vice versa, find suitable court cases, expanding or explaining the meaning of laws and acts and create semantic links to the outside world- adding relevant sociological and economic data for example. This should uncover even the most hidden relations and interdependencies between different laws, acts and amendments, but in the larger picture could give some hints and insights about interests to influence also the legislation drafting process. Last but not least- such visualisation could offer considerable help with planning changes in the legal domain, giving full overview about all aspects how legal objects are interrelated.

In order to capture the legal knowledge of legislation in a way, which leaves room for the widest interpretation and still allows users to find the most specific textual clarification, an effective visualisation is in our opinion the best solution. Visualisation should cover the existing relations (structural subdivisions, links in and out from statutes, implementation relations etc.) and add more relevant links by demand (within timeframe, issuing authority, sequence of related statutes acceptance etc) in order to give an overview of all relevant associations. Each node of the graph can act at the same time as an access to the lower level of the legal domain general structure or open direct access to the legal text behind this node, all describing relevant aspects of the legal domain. Connections between nodes carry certain information, used for estimating the relevance of the logical direction of the given analysis and showing preferably the shortest route between the general starting point toward the specific end point of the analysis. It is important to show the alternative routes and “loop-backs” in

order to open more hidden relations between the legal information objects (Figure 4).

7.1.1.1 Example of use

A part of Estonian legislation was chosen for analysis in order to test the current approach. This covers a subdivision of the legislation, grouped in Estonian State Gazette [99] under the name "State law" subdivision "Nation" and including only election related laws: European Parliament Election Act, Local Government Election Act, Referendum Act and Riigikogu Election Act. All together it covered a text massive with 2739 sentences and 32187 words.

Table 4. The two-way table, showing the concordance of nouns and verbs within clauses

Nouns\ verbs	be/is	submit	carry/ enter	do	write up	give	keep	become	organize	think
Selection Board	21	2	2	16	0	4	9	6	4	0
elector	34	3	5	0	9	10	3	9	0	0
list	6	3	3	10	34	0	2	3	0	0
appeal	4	20	0	0	0	0	0	0	0	0
person	9	11	14	3	4	4	5	0	0	0
day	10	0	5	8	0	0	0	2	1	27
party	10	8	12	0	0	4	0	2	0	0
division committee	3	4	0	4	0	0	0	0	0	0
member	4	7	7	0	0	14	4	0	0	0
candidate	4	1	0	1	0	2	2	2	0	0
voter	12	1	2	1	3	4	1	2	0	0
independent candidate	0	9	2	1	5	8	0	0	0	0
voting	8	0	0	1	0	0	2	0	24	0
envelope	0	0	0	0	0	0	0	0	0	0
decision	5	0	0	10	0	0	0	0	0	0
time	0	0	4	4	0	3	0	1	0	0
Family member	15	0	0	1	0	3	13	0	0	0

The consolidated text was then analysed automatically, based on language structure and using TAHMM morphological analyser. The borders of clauses and basic forms of words were estimated. After that the concordance of the verbs and nouns within the clauses were calculated, using Excel tools. As a result there was formed a table, showing how many times a found verb situated with found noun in the same clause (Table 4). All-together 396 unique nouns and 197 unique verbs were found, reducing

the amount of words compared to the original document more than 52 times. Additionally the table can be used for estimating the relative importance of verbs and nouns. This can be done differently, for example the word with highest concordance frequency is having the biggest importance, next smaller etc. Such lists can give to us the rankings of keywords, main objects and activities and so on. In the following example in Table 4 about election related laws, a most important object by this way is “elector” related to verb “be/is” or “list” related to word “write up”.

This way it is possible to get the compressed view of the legal acts and all the later analytical operations were based on this table. The cost of such condensation is a significant loss of information, but the suggestion is to keep this information attached to the original text to undo this operation in case of necessity. A positive feature of such approach is a relatively small need for computing power and possibility to get quick results.

This table was turned into an undirected graph (Figure 9). The number of words in this graph reached almost up to 600 words and the number of links was 1081. As a result it is possible to get a tense and almost fully connected network of keywords, a “skeleton” of the legal acts. Graph edges represent the words and the size of edges is showing the degree of it- to how many different words it connects. Arcs are showing the connections between words and the line thickness is showing how many times this pair of words did appear in one clause (max. 34 times).

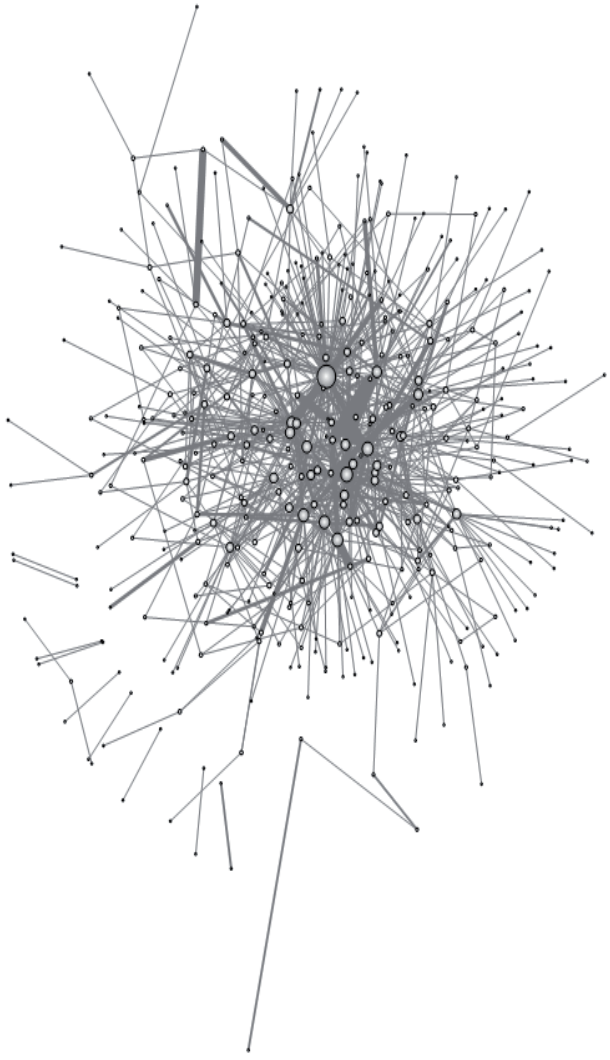


Figure 9. Full graph of connected keywords within selection of legislation

In order to reduce the complexity, the graph was reduced for more substantial analysis (Figure 10).

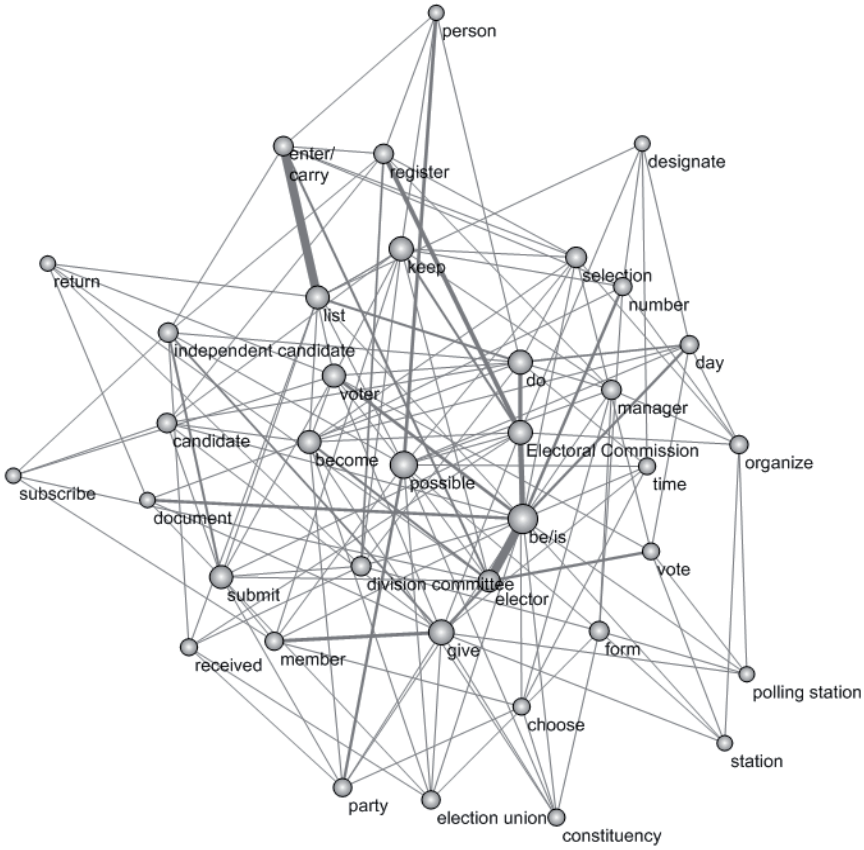


Figure 10. Graph derived from verbs and nouns concordance in the clause

This illustration is reflecting only the most intensive part of the graph, covering the most frequent words and connections between them. Sizes of edges are depending of the degrees, lines thickness between the edges reflect the connection frequency. Such graphical representation can be successfully used for brief inspection of the law. This helps to get a quick overview about the content of the legal act, showing the most important objects and activities. The Graph can be useful also for the text search (all activities regarding to certain object can be easily listed etc).

7.2 A similarity measurement

A second method to prove the legal text content capture method was an experiment to find similar norms within the legal texts. If the similarity can

be estimated based on verbs and nouns coexistence within different clauses a new similarity measure could be developed. In order to do the structural research of the legislation, a problem to measure the similarity of two different legal acts has to be solved. Data mining and graph theory aspects form a foundation for that.

The similarity measurement method can be briefly described by next following steps:

- find the same verbs within different clauses within different legal acts;
- find the same nouns "attached" to the previously found verbs;
- estimate the coexistence frequency similarity between same verb-noun pairs within different legal acts;
- calculate the similarity percentage (0% - no common pairs found, 100% - all the pairs from smaller legal act two-way table are present in bigger legal act one).

7.2.1 Data mining and graph theory

Visualisations are based on two-way frequency matrixes, which are one of the main research objects for the quickly evolving field of data mining. The aim of such an approach is to group and regroup similar items in different ways and will be discussed later. The visual representation of legal text as the graph makes a good formal comparison basis between legal acts and is a key to perform the quantitative analyses of the legal text, using different graph theory aspects and graph mining methods. A data mining technique was applied to estimate the similarity of different graphs.

The process of evaluating the similarity of two graphs is commonly referred to as graph matching. The overall aim of graph matching is to find a correspondence between the nodes and edges of two graphs that satisfies some, more or less, stringent constraints. That is, by means of the graph matching process similar substructures in one graph are mapped to similar structures in the other graph. [20, p. 219]. Graph isomorphism calculation can be an np-complete task, but as the graph edges have names (a basic form of filtered words due to Estonian language peculiarities) the similarity measurement is relatively simple. Therefore it is possible to perform graph structural similarity calculations and this notion forms a basis for further restructuring of legislation.

7.2.2 A test case based on acts layer of Estonian legislation.

To test this method and to demonstrate the broad applicability of such approach, an experimental fitness function was generated [25]. The purpose of this function was to measure similarity of two different legal acts. The function estimated the result in three stages- how many shared verbs were found, how many shared nouns connected to specific verbs and frequency of the use of each pair. The general weight is a sum of:

- Shared verbs account for 45%;
- Shared nouns connected to specific verbs 45%;
- Concordance frequency 10%.

Particular calculation weights were experimental, chosen willingly and are subject to further evaluation.

For the test case of normative system structural analysis:

- 386 Estonian legal acts were randomly chosen;
- Each legal act compared to each other legal act (148996 comparisons).

As a result a similarity table was created, consisting of 386 rows and 386 columns (see small piece of it in Table 5).

Table 5.Similarity table of legal acts, %

Act ID	13360508	13360557	13360674	13360776
1014238	27	27	10	11
12732218	38	23	22	15
12755417	64	36	37	19
12833542	39	28	29	14
12857270	64	22	32	14
12861944	64	36	31	21
12862402	15	30	0	0
12911615	64	20	21	18
12936755	0	15	15	15
12969907	42	20	27	15

As a result, some interesting general characteristics appeared. In medium, all legal acts seem to share the content roughly by 1/3 (see Table 6), but this is not a general rule for all. 167 cases (0,22%) two compared legal acts did not share any content. According to results (see Figure 17) it can be said that there is more likely overlapping between legal acts.

In 20 cases the similarity was measured 80% and higher. A quick control of two most similar legal acts (Estonian Parliament Election Act and European Parliament Election Act) showed a remarkable similarity of texts up to some parts exactly copied from each other.

Table 6.General characteristics of similarity table, %

Characteristic	**Value**
Lowest minimal similarity per act	0
Highest minimal similarity per act	33,1
Lowest maximal similarity per act	39,7
Highest maximal similarity per act	91
Medium similarity	29,9
Lowest medium similarity per act	12,6
Highest medium similarity per act	42,9

7.3 A prototype

A third way to test the legal text content capture method was to create an interactive software prototype in order to perform a high level analysis of the legal content, explore the functionality of the method and give broader access for interested parties to test the method. The prototype is universal and suitable for visualizing the legislation in full though it is experimentally applied only on some of legal acts. Software is built upon previously described two way tables and processed legal texts. The prototype allows performing different visualisations in different generalisation levels and is connected to the original legal text in order to get better understanding about reasons of usage of found noun-verb pairs. An intriguing feature of this prototype is the ability to show the common word pairs of two (or possibly more) different legal acts to help to reveal the origins of their similarity.

In order to test the visualisation properties and search potential of the offered approach, a prototype was designed based on the small part of the legislation, accessible from the link: http://data.ttu.ee/visuallaw. The prototype covered the following legal acts:

- European Parliament Election Act, (EUROOPA PARLAMENDI VALIMISE SEADUS)
- Local Government Election Act (KOHALIKU OMAVALITSUSE VOLIKOGU VALIMISE SEADUS)
- Referendum Act (RAHVAHÄÄLETUSE SEADUS)

- Estonian Parliament Election Act (RIIGIKOGU VALIMISE SEADUS)
- Political Parties Act (ERAKONNASEADUS)
- Citizen of European Union Act (EUROOPA LIIDU KODANIKU SEADUS)
- Citizenship Act (KODAKONDSUSE SEADUS)

These texts were analysed deeply and structural properties (chapters, subdivisions etc.) added in order to extract also the exact "location" of the norms within texts. Analysed and revised data was injected into the database and using JIT (Java Infoviz Tool) a web based user interface was created. The purpose was to build a web accessed dynamic query tool for browsing through the legislation, help users to find relevant norms quicker and free them from reading the complicated legal texts from top-down.

An opening window presents seven legal acts, connected together (Figure 11). Connection between legal acts and connection strength is characterized by act similarity, measured in percentages. This user interface also has a functionality to show the connected legal acts and list these up in the left side of the web page. The web page also has a filter to select legal acts with the strongest links between each legal act, in order to simplify the picture and reveal the strongest connected part of the graph (Figure 12). There is an option to delete nodes manually in case of necessity (clicking the crosses on the end of the legal act names).

A layout tool allows presenting the graph in a different format in order of convenience: force directed, circular and tree-like representation. These functions are available in sub-pages also.

In order to grasp the legal act content in a bird-eye view, a right mouse click has to be done on a suitable node. For example, clicking the node "RAHVAHÄÄLETUSE SEADUS", a connected word cloud or mind map are opened (Figure 13).

In order to capture the core of the legal act, a filter is useful and if to estimate the value of filtering on 15% (i.e. 15% of the strongest links are shown) the picture result is shown in Figure 14. The numbers on the lines show the actual frequency of use of specific word pairs within the clauses.

In order to find all the norms related to the organisation of voting ("hääletamine"-"korraldus"), one has to find the link between two nodes and right mouse click on it. As a result a left column will list up preview of all the norms consisting these two words and adds the info about structural location (chapter, subdivision etc. with the texts). Sliding over the norms with the mouse will open the full view to the norm (Figure 15)

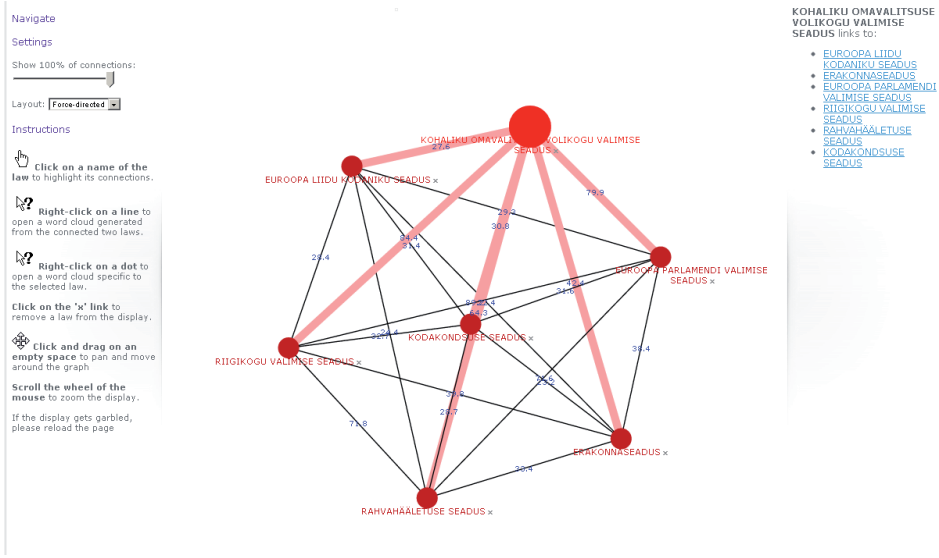


Figure 11. The prototype opening window

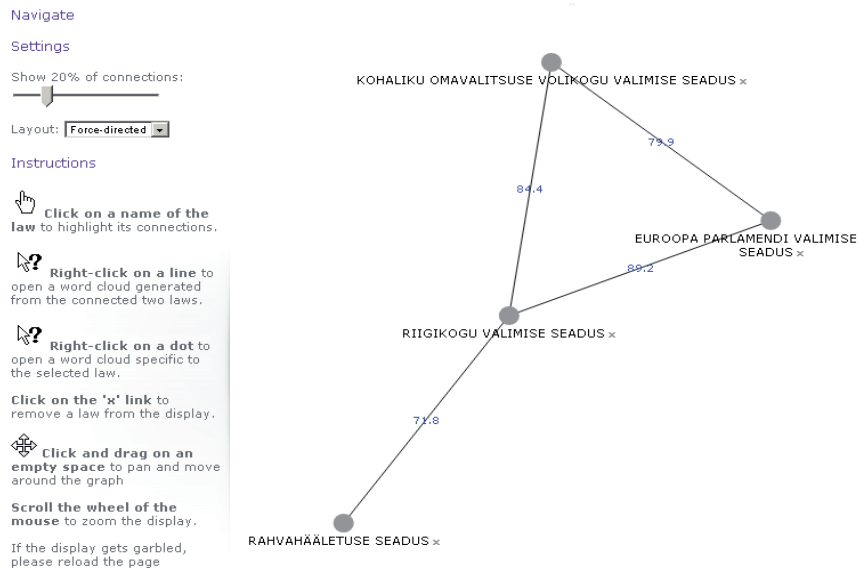


Figure 12. Filtered view to the opening window, showing 20% of the strongest existing links

A most innovative feature of this prototype is a possibility to find the common part of the two (or possibly more) legal acts. Right mouse clicking on the line connecting two legal acts in the opening window opens the shared part of the legal act, where two legal acts share the same nouns (objects) and verbs (processes, events). In Figure 16 the shared parts of the European Parliament Election Act and Citizen of European Union Act are presented as a single graph and a right mouse click on the line between “alus” (basis) and “olema” (is) allows us to filter out relevant norms from both legal acts at the same time.

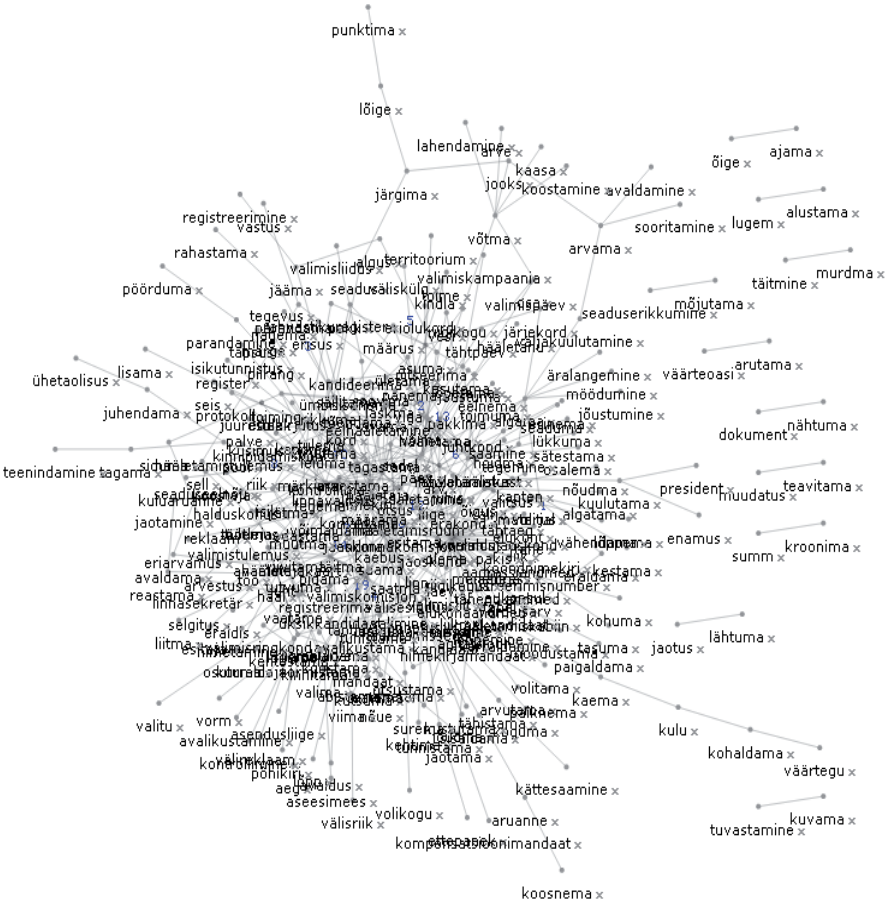


Figure 13. Visual layout of the legal act “Rahvahääletuse seadus”

Figure 14. Filtered view to the "Rahvahääletuse seadus"

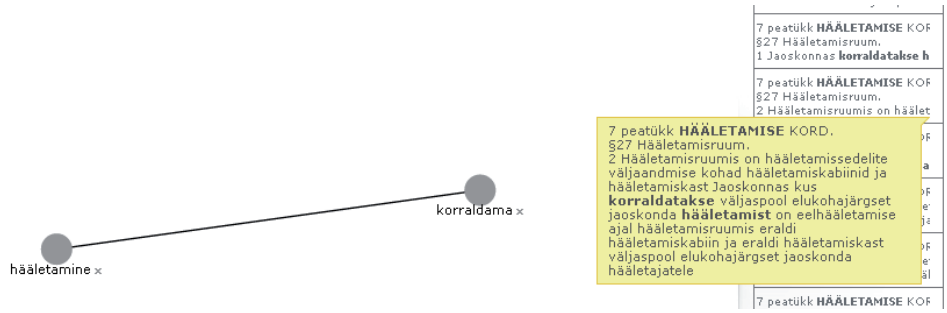


Figure 15. List of norms consisting words "hääletamine" (voting) and "korraldama" (organize)

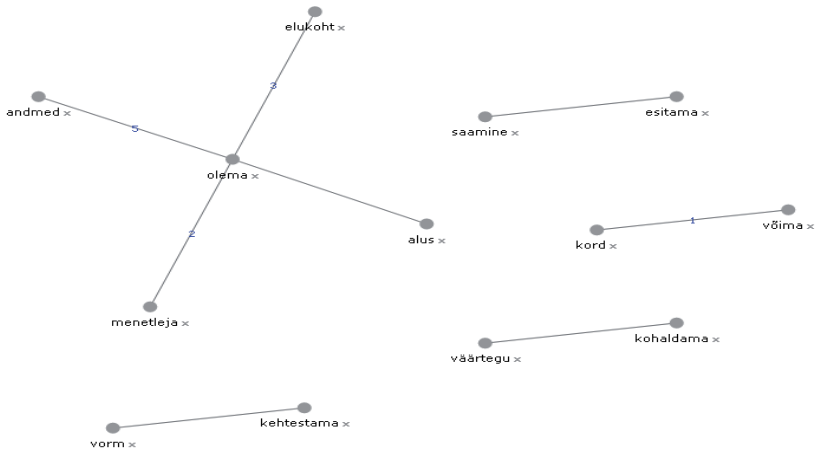


Figure 16. Shared graphs of two legal acts: European Parliament Election Act and Citizen of European Union Act

7.4 A deep structural analysis

The fourth and last step of proving the value of the legal text content capture method was an attempt to research the structure of the whole legislation, based on similarity measurements described in subdivision 7.2. The main reason behind this attempt was the desire to check the scalability of the proposed method and also to present more material for legal experts to discuss.

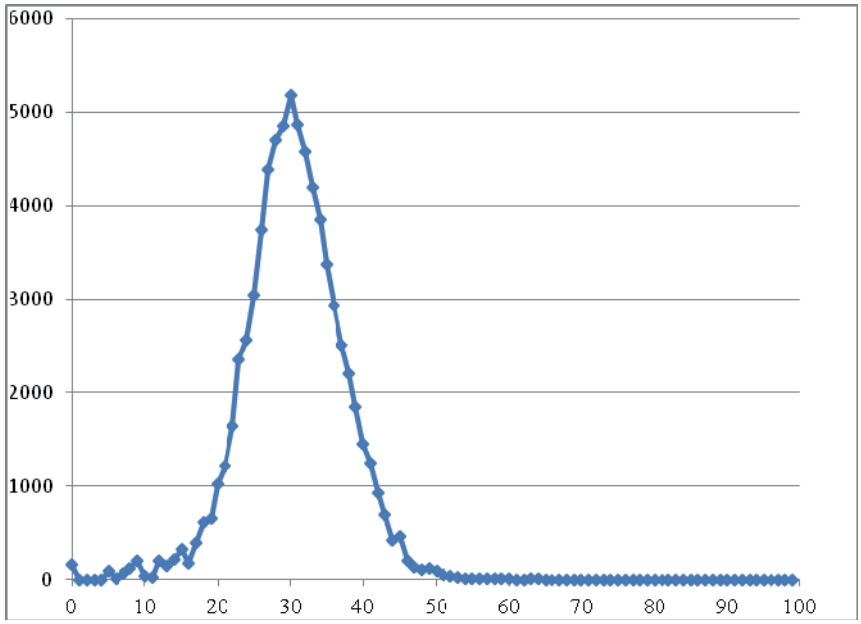


Figure 17. Frequency of similarity results

However, to get an overview of huge amount important aspects of similarity interconnections (as referenced in Table 5 for example) is impossible for humans. It is possible to get a better condensed overview of those similarities using different visualisation methods.

In order to perform the restructuring experiment, next steps were performed:

- the similarity table (*Table 5*) was analysed using maximal connected path (Figure 18);
- the most densely connected part of the legislation was visualised (Figure 20);
- the similarity table was turned into 0/1 table in order to simplify the matrix permutation operations;
- the most informative matrix permutation was selected (Figure 21);
- clusters found on the permutated matrixes were investigated (Table 7);
- common verb-noun pairs within clusters were estimated (Figure 22);
- norms (legal text clauses) containing found verb-noun pairs were estimated and presented to the legal experts.

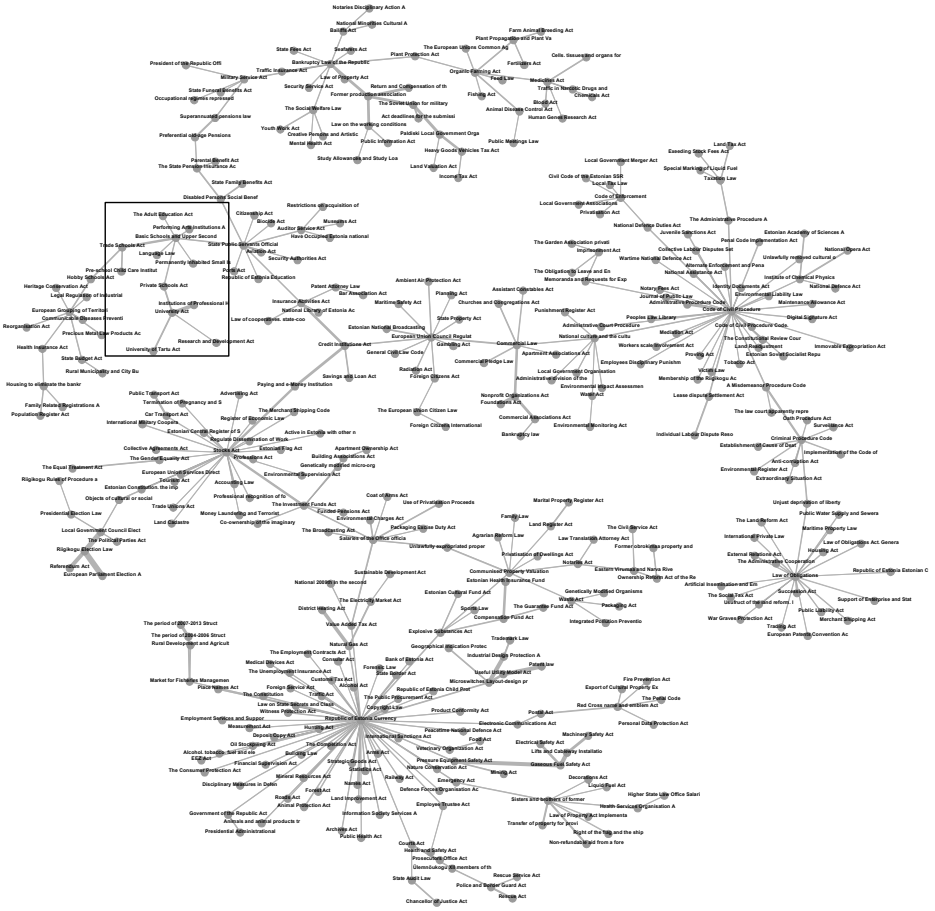


Figure 18. Similarity structure of the Estonian legislation, where each legal act is connected to the most similar one

In order to present a complex connected system, it has been useful to use extreme representation principles. In our case for Estonian legislation the principle of maximal similarity spanning tree was used and for that similarity matrix S of legal acts the maximal connected path was computed [21]. In Figure 18 every legal act is connected to its most similar legal act. A zoomed upper part of this graph is shown in Figure 19.

All 386 legal acts are represented in one tree, which represents the maximal similarities between acts.

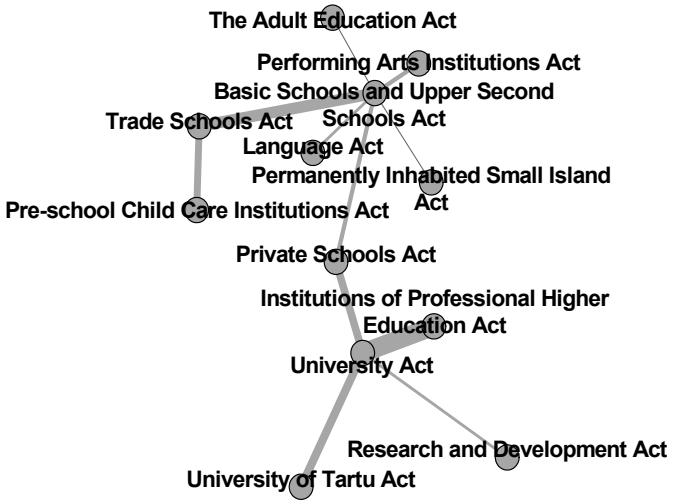


Figure 19. Extracted part of similarity structure

Figure 20. The densest part of the graph constructed from the strongest connections

A symmetrical matrix of similarities was transformed into 0/1 matrix based on cutting level of lowest maximum similarity (37%) to guarantee to

get a single connected component of all acts. After using the standard visualization method the densest path of the acquired graph is presented. (Figure 20) A matrix permutation was then applied on symmetrical matrix.

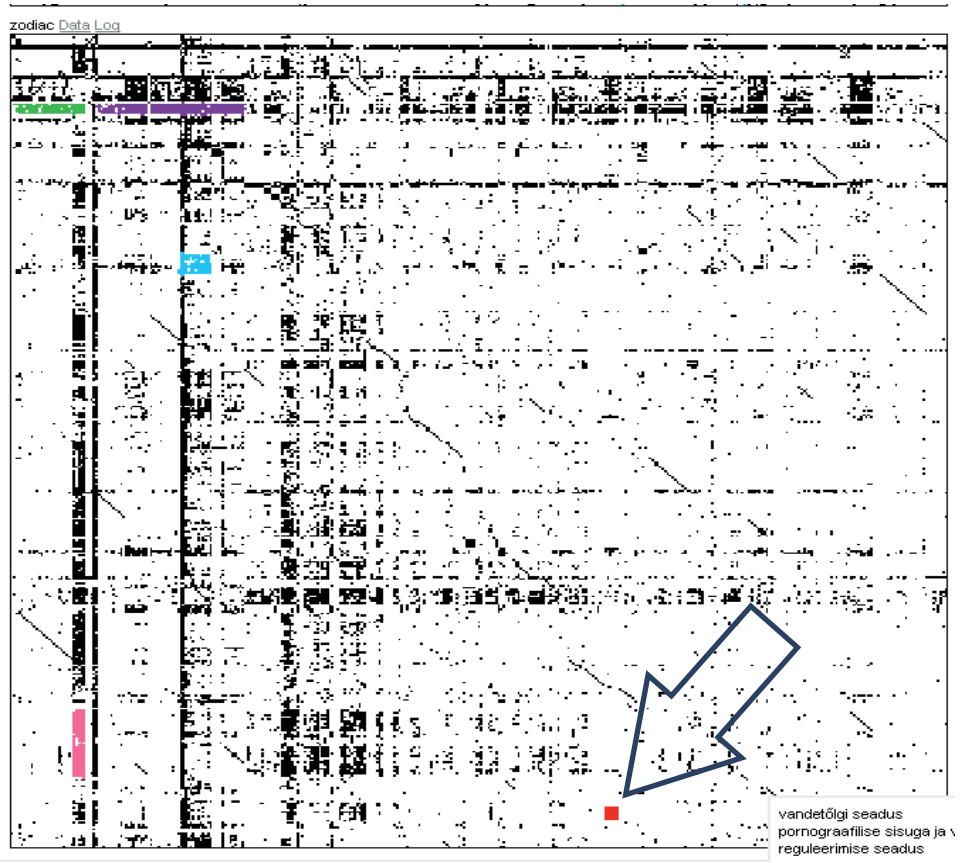


Figure 21. Reordered table of 0/1 matrix

Of the next methods tested [100]:

- fast O(n log n) heuristic for larger matrices, based on sorting by the frequency of 'ones' [100];
- conformity scale, algorithm from the Monotone Systems metaheuristic by Mullat [101] and Vyhandu [102];
- minus technique, algorithm from the Monotone Systems metaheuristic by Mullat [101] and Vyhandu [102];

- plus technique, algorithm from the Monotone Systems metaheuristic by Mullat [101] and Vyhandu [102];
- McCormick's Bond Energy Algorithm [103];
- ROC2-an enhanced rank order clustering by King [104];
- MODROC-an extension of the rank order clustering for group technology by Chandrasekharan and Rajagopalan [105];
- ART-a Carpenter–Grossberg neural network based clustering by Kaparthi–Suresh [106]and Kusiak–Chung [107];
- ZODIAC-ideal seed method for part-family and machine-cell formation in group technology by Chandrasekharan–Rajagopalan [22];

the ZODIAC method produced the most favourable results. [22, pp. 226-229] The method was chosen because of the dense black regions describing the subsets of strongly similar legal acts, shown in Figure 21. Other methods did not give such easily recognizable results but definitely deserve more careful analysis in future.

In order to reveal the meaning of clusters within a matrix, a reverse process was applied. Clusters were extracted and investigated carefully. One such example of a cluster is indicated by an arrow in Figure 21.

Table 7. The selected cluster of matrix on Figure 21.

Legal act name	Lifts and Cableway Installations Safety Act	Electric Safety Act	Mine Safety Act	Gaseous Fuel Safety Act	Machinery Safety Act	Pressure Equipment Safety Act
Lifts and Cableway Installations Safety Act	1	1	1	1	1	1
Electric Safety Act	1	1	1	1	1	1
Mine Safety Act	1	1	1	1	1	1
Gaseous Fuel Safety Act	1	1	1	1	1	1
Machinery Safety Act	1	1	1	1	1	1
Pressure Equipment Safety Act	1	1	1	1	1	1

7.4.1 Cluster analysis

The chosen cluster collects together most of the safety regulations within Estonian legislation, shown in Table 7.

This group of legal acts was analysed further to estimate the part of acts which were considered as similar ones. The analysis estimated the common pairs of nouns and verbs, being part of each chosen legal act (represented in Figure 22).

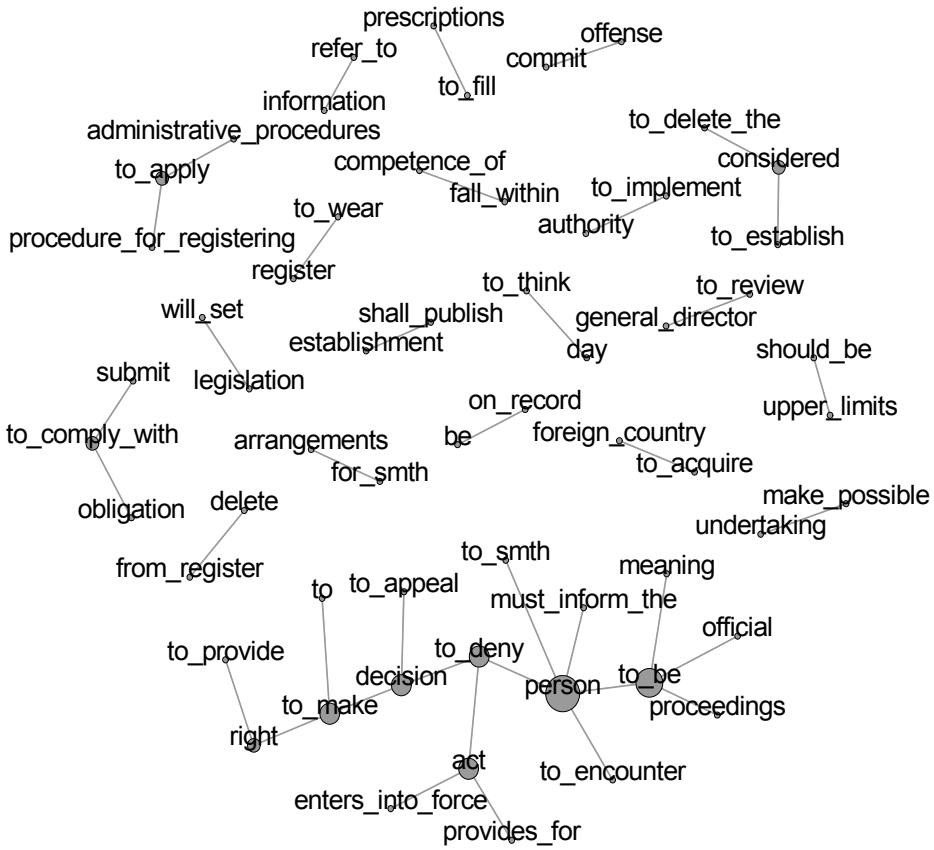


Figure 22.Common noun and verb pairs, found in each act within a cluster

A selection of common verb-noun pairs, found in each legal act within investigated cluster, was a basis for further analysis. Based on findings specific norms were selected from all of these legal acts and presented as full texts. The reason behind this text extraction was to understand the

broader context of the actual use of verb-noun pairs. In extreme cases, more than 6 000 sub clauses were found within one legal act, all together more than 17 000 sub clauses per cluster. This overview together with extracted texts was delivered to legal experts for legal analysis.

7.5 Legal analysis

The developed system of automatic systematization shows that a key issue is how to select the terms used for the analysis. An initial use of the method points to a number of connections between basic terms, which express the nature of normative acts rather than necessarily the specific content of one of the other normative acts. Examples include such connections as “law” and “prescribe”, “decision” and “make” or other similar connections. With an appropriate choice of terms to apply the system to, such overly general connections can be eliminated. At the same time, even such basic highlighting of terms used may be useful for the aforementioned task of determining if the legal language used is coherent and consistent throughout the legal system.

For example, consider legislation concerning health and safety requirements in different context, the systematization highlights the central role in the legislation for the individual. In many different pieces of legislation the same terminology is used, even word for word, when it comes to responsibility to ensure that conditions set out in law are met. This analysis, even on this brief and superficial level, points to a coherent legal drafting method, which is positive, as this should help those concerned to understand the requirements. Moreover, the background to similar terminology being used may be due to the fact that many laws were passed in the same time period and also because EU legislation may have been influential in drafting several legal acts (i.e. in the instance of implementing a directive or regulation). Certain terms referring to more specific content of obligations (to register, to fulfil etc.) also show a connection.

Although the coherence of health and safety legislation based on these connections may be assumed, this method may give us information, which can be used to further improve legislative drafting. For instance, the legal context of the words “person” and “undertaking” could be analysed in this particular field of legislation with the purpose of determining what is the difference in the use of these two words, because in legal terminology the word “person” can be a legal person or a natural person, and the word “undertaking” is something else, but formally it also means a legal person. It could be interesting to analyse what is the reason for using the word

"undertaking", but not a legal "person" instead. Questions can be raised as to whether the use of two words is necessary and justified or perhaps it could be reduced to one "person". It all depends on the context and purpose of using those words. If an "undertaking" was to be replaced with a "person", then the number of connections in Figure 22 would be reduced, and not all "undertakings" would have to be replaced to break this threshold.

Similarly, other areas of legislation could be analysed in order to identify whether the words "undertaking" and "person" are used in a similar way, and perhaps there are new words that can be identified, such as "entity" or "establishment", that are used in a similar manner. The correlation of the use of similar words in different areas of legislation may lead to the conclusion that the number of words used to denote the same can be reduced, and thus greater coherence can be achieved.

When navigating the various acts of commercial legislation the connections highlighted tend to be rather tenuous with only general words common to most laws showing a clear link. At the same time, the number of links is rather large, so a common terminology and common way of expressing obligations can be seen. This does show that there is a connection between the laws concerning a common topic as such similar links are not evident when analysing laws with different topics. Such connections are sparse when comparing public service, public procurement, and state property laws to each other. Here the number of links is surprisingly small, given that the subject matter of the laws is related. The reasons for this would necessitate a deeper analysis to eliminate such matters as: laws being drafted at different periods, thus affecting the terms used, rather than their substance. The systematization picture gives a useful clue to an area which would be relevant to analyse further.

7.5.1 Preliminary results of Estonian legislation structural analysis

Although the analysis process of experimental results is quite in the beginning it is possible to make some basic assumptions about the structure. Estonian database of legal acts www.riigiteataja.ee which is the only official means of publication of legal acts is based on a tree-like representation as in Figure 23. [99] Such structure divides legal act into branches and sub-branches, some legal acts can appear in different branches simultaneously.

There are other reasons than merely technical why legal acts are represented in this manner. A tree like representation of law follows the

traditional categorization method used to explain a civil law system. The classical division of legal acts between categories of private and public law is the starting point in the database. Both categories have subcategories based on subjects of law that traditionally fall under private and public law. For instance, private law has a subcategory called intellectual property law, which includes different legal acts such as Copyright Act, Geographical Signs Act, etc. There is nothing surprising for a lawyer who has been taught about the structure of a legal system in a law school (see Figure 23).

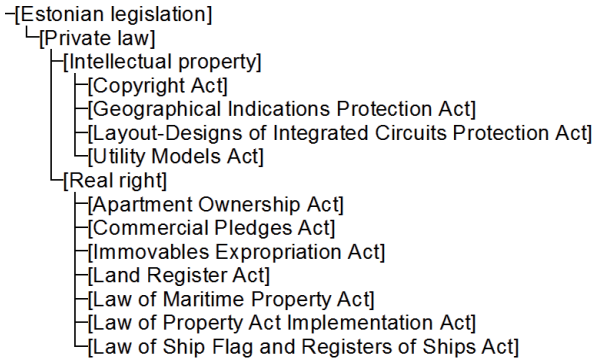


Figure 23. Abstract of the tree like representation of Estonian legislation

Yet, the ordinary person does not understand a legal system in the same way. He or she has not been taught about the difference of private and public law, and why some areas of law fall under both categories. For instance, consumer and employment laws intercept with both categories of private and public law. In Estonia consumer contracts are regulated in Law of Obligations Act, which is of private law and consumer rights are more specifically regulated in Consumer Protection Act, which is of public law. Furthermore, in a case of a consumer dispute it may be necessary to look for relevant provisions in General Part of the Civil Code Act, which is of private law. If an ordinary person is looking for information about consumer provisions and types in the word “consumer” to search for relevant legal acts by a title, then the result is the Consumer Protection Act and other legal acts, but not the Law of Obligations Act or the General Part of the Civil Code Act. If the search is performed by looking for the word “consumer” in a text, then also the Law of Obligations Act is included in the result, but not the latter.

The reordered table of 0/1 matrix (Figure 21) allows us to offer an alternative view, more suitable for further analysis. The table reveals some

strong lines which are more and less dense regions of strongly connected legal acts. Key players within these clusters tend to be the same ones and in our case some known repeating legal acts are: Commercial Code, Auditors Activities Act, Insurance Activities Act, and Credit Institutions Act. In many cases other legal acts tend to group around these regulations, forming a certain pattern, but sharing common components. Therefore it is suggested that a more correct approach to the Estonian legislation structure should take a flower like shape as shown in Figure 24. Key acts are located in middle and petals are formed from related acts having less central role. The full number of petals and their content are under research at the moment.

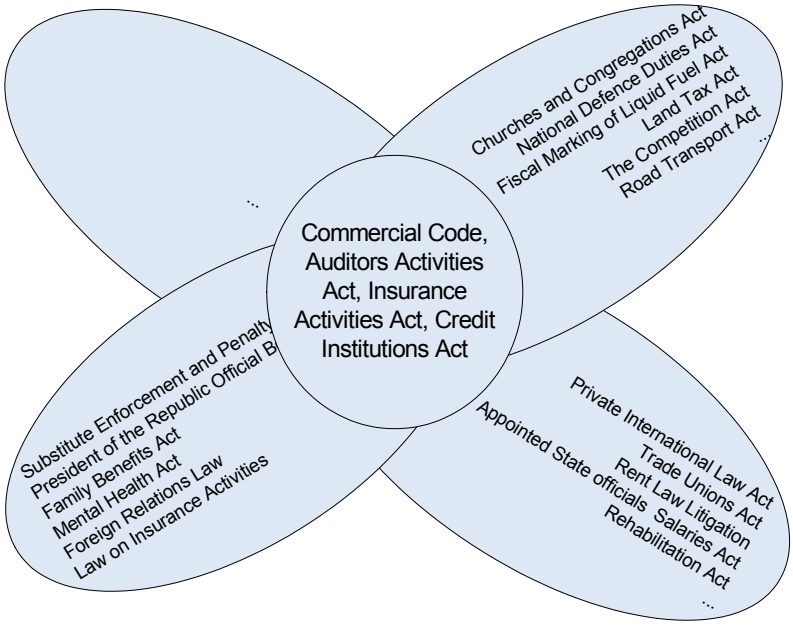


Figure 24. Flower like representation of Estonian legislation

The approach taken by the method presented here is different from the traditional categorization of law. The method under research disregards categories and subcategories altogether. Figure 21 shows a result that leads us to believe that further research on this method may lead to an alternative search methodology that could be applied to databases of legal acts. In the context of the illustration of the consumer law one of the aims of further research would be to find out whether the search by inserting the word "consumer" could reveal all relevant legal acts concerning consumer law,

i.e. also legal acts that do not contain the word “consumer” in a title nor in a text, but have a connection. If this proves to be possible, then it would enhance the possibilities of searching for topics of law based on connections between provisions that are scattered in different legal acts, yet are relevant to a particular topic of law. This may render improved possibilities for an ordinary person to study about topics of interest in the labyrinth of legal acts as it would contribute new possibilities to the legal education and practice.

8 Conclusions

8.1 Research results

According to the research questions estimated in subdivision 3.4, the following results were gained:

Increased usability of legislation: an alternative guiding layout, which is closely connected to the written legal texts, is developed with a potential to lower the barriers for non-educated users and easing access to the sources of law.

Suitable visualisation approach found: a graph layout was chosen, allowing scalable visualisation of the legal text, enabling the on-spot access to the legal sources at different generalisation levels and opening a way for many different information enriched representations of the legal texts.

Suitable methods and techniques found: necessary linguistic, legal and mathematical methods for creating visualisations were estimated and applied successfully, using among other system analysis modelling, re-engineering and graph theory.

Selected methods exploited and tested: the method was applied on legal text and the text was successfully turned into graph. In order to test the value of such layout itself, a big part of Estonian legislation was selected and the method was applied on it. Several other methods (graph similarity measurements, data mining) were exploited on gained results in order to get the results and test its usability for further developments. The prototype for testing was built and opened to the public (http://data.ttu.ee/visuallaw/).

Objective evaluation from legal experts gained: in cooperation with legal scientists the first evaluation about the proposed approach was performed and the method was considered suitable for the chosen purpose. Some consequent observations about legislation build-up were made and introduced at a high level conference.

8.1.1 Method usability

In order to present the findings and conclusions in a more elaborate manner and to give some additional insights to this complicated issue, results of the research will be described in the following subdivisions from different angles.

Extracting selected words (nouns and verbs) from the text and preserving their connections to each other allows us to:

- capture the norm content (60-.80%);
- compress the content (more than 50%);
- achieve qualitatively new level for automated analysis of legal texts.

The two-way table forms a layer above actual legal texts, reflecting the content of it at a very high level, which is better suited for computerized tasks compared to natural language. Therefore such presentation of information is:

- *Scalable*- it can consolidate the norm, a subdivision, a chapter, legal act or whole legislation
- *Computable*- with help of math tools (data mining, graph theory etc.) it can be easily analysed by computers;
- *Visual*- it can be presented visually for human users and deliver high level overview about the content of the legal document;
- *Information enriched*- the picture of most frequent keywords (like Wordle), it is giving also some information about connections between different words and characterises

Based on the previous, several other features can be moved out in order to describe the value of such a method.

Information-enriched graphical representation of the law

The graph is representing a content-rich environment, opening new ways of treating legal documents. Built upon the existing texts, the graph can be directly related to it and does not conflict with any part of it. A visual layout of the graph is formed from the keywords, which are considered to be important within the text. The cost of such condensation is a certain loss of information, but the generality and visibility of it is greatly enhanced and one should use this graph together with legal text in order to get a complete picture.

A search tool

The data and graphical representation of important nouns/objects and verbs/activities in the legal act supports a specific search within the legal text. Such document-specific listings support the text search, but can also be used for broader searching tasks. Untrained users learn the specific language; trained users stay better on track and do not fall accidentally out from the legislation-specific frames. Relative significance of different terms can be estimated with the help of different quantitative parameters (number of connections, frequency of co-occurrences); less important terms can be filtered out and this way concentrate on the most central terms. In case of necessity the search can be done at a more detailed level in order to find the smallest relevant detail.

A mind map: giving valuable insight into the most important aspects

The networked presentation of the matter, especially in a reduced graph, forms a good visual picture of the main regulated issues under the regulation. This way the graph layout is a kind of a roadmap or a mind map for human users, guiding them quickly through the legislation, providing them with a general understanding of it and revealing groupings of keywords to reflect main issues. Scaling is easily achievable and it is a purely technical question whether the object is part of the legal act, part of the legislation or the legislation in its full extent. The method can be applied outside the legislation, still keeping good contact with the legal content-evaluating random texts in order to match them with the legislation content for example.

Entry to ontology

The presented graph layout is developed independently from ontology, but can be used as an entry to ontology. By definition, ontology is an explicit, formal representation of the entities and relationships that can exist in a domain of application. Usually entities and relationships are estimated manually, the current method offers a way how to find entities and create relationships automatically. Again, the more valuable entities can be estimated with help of quantitative weighting (number of connections, frequency etc.), offering this way a set of statistically weighted keywords and their weighted relationships. The main advantage is a quick, objective and fully automated process against the slow, subjective and manual process used so far.

Comparison basis

The graph layout is suitable for the comparison of legal documents or its parts. Most frequently used keywords and their relations form a sufficient basis for quick and rough comparison of different legal documents. This again opens these results widely for data mining operations, like grouping, reordering, filtering and so on. Based on experience gained so far, it is a rather promising approach and does bring out the similarity in a humanly recognizable way.

A basis for restructuring legislation

The ability to compare graphs is a key to the restructuring of the legal domain, taking into consideration the graph-based similarities of documents and the existing hierarchies in the legal domain. This opens the way to the restructuring of the legislation and visual presentation of its structural relations based on the probability of the similarity. This could be the case when one needs to understand the legislation better or has to reveal the

hidden relationship between different parts of the legislation. Based on actual, measured similarity, it is possible to build up better approaches and ways how to find the needed legal sources.

8.1.2 Test results

In addition to previous findings, the test cases proved the following:

The proposed method allows creating an easily readable compressed visual picture (a graph) of the legal act specifying concordance of verbs and nouns within clauses. Such a representation has a good potential to free the user from the obligation to read huge amount of full texts in order to estimate the most relevant legal act among thousands, suitable for his or her situation. When the required legal source is estimated, it also helps to locate the exact position of the needed legal norm within the legal act. This way such graphical layout reduces the barriers to use the legislation:

- *easing feature*: the need for previous, high level knowledge about the build-up and general principles of using the legislation is significantly reduced;
- *releasing feature*: necessity to read carefully all seemingly relevant or questionable legal texts is reduced to a large extent;
- *training feature*: often the legal text includes specialized terms unfamiliar for non-expert users, here the graphical layout becomes also a tool for recognizing the gaps of viewers´ knowledge and creating prerequisites for better understanding of legal texts (one has to understand specific terms in order to understand the meaning of this norm); if the list of terms does not contain the specific word in the users mind, one has to find the closest match to it
- *uncertainty reducing feature*: list of keywords presented in graphical layout are derived directly from the legislation and therefore any search is limited by and falls within the frames of the legislation;
- *guiding feature*: a connected set of legal terms is providing users with hints to find relevant connections with other legal norms in- and outside of the legal act under investigation.

Still at the end the actual, original legal text in full must be analysed in order to understand the real meaning of it, therefore it is useful to link graphical layout back to the original text. This is easily achievable and this concept is developed and researched by means of a prototype, described in subdivision 7.3.

Using these graphs to measure the similarities of different legal acts, it is possible to analyse the whole normative system (for example the maximal

connectivity graph, Figure 18). This kind of results are useful for, among other things:

- *enabling the big picture*: understanding the general build-up of the whole legislation;
- *pointing out the big players*: locating the key legal acts mostly influencing others;
- *estimating the range of power*: estimating broadly the range of influence of these key players;
- *measuring the cost of changes*: estimating broadly the range of related legal acts needing change if some of key acts will be changed.

The main purpose of similarity measurements is to present an alternative, fresh view to the existing legislation and to legal acts grouping principles. Therefore it helps develop users´ understanding about the legislation and their view of it as a highly connected system.

- The method is mathematically well justified and is based on objective, quantitative measurements opposed to the subjective, intuitive normative system classification used so far. As the similarity measurement is based on co-existence of word pairs within the clause, their frequency within the legal act is counted and presented as cross-tables. Such result is a good ground for mathematical operations at different levels. During the process, the content of the legal acts is preserved this way up to 80%, therefore we are dealing with unique semi-textual, semi-numerical data layers closely tied with actual legal texts. This allows us to perform numerous data mining operations or other mathematically founded operations in order to reveal some quantitative characteristics of the legal system not accessible so far. Most challenging could be the interactive support for legal analysis, armed with sophisticated aiding instruments and reacting flexibly according to the users´ needs.
- Qualitatively different new applications can be derived from presented graphs to perform a normative system analysis. The dual face of the data layer- cross table and its graphical layout allows us to combine different analysis techniques in order to gain some practical results:
- Minimal connected path could reveal the norm specification pathways starting from the constitution down to the court decisions. This way it is possible to follow the application of the most general

norm in the most specific situation and reveal all its intermediate statuses;

- Most densely connected parts of legislation can be separated and investigated as independent parts of the legislation in order to understand its function and purpose better;

Hereby only a small set of possible applications derived from network analysis was presented. A good analogy can be borrowed from geographical information systems- a map. The map is a just a small glimpse of the reality- it is impossible to mark on it every tree and stone on the road. Maps have high strategic value for each country. Military, public service, business- maps are used everywhere and are able to free the users from the necessity to actually visit the part of the land of interest. Digital maps are even better- with help of different tools (GPS, WIFI, mobile networks and so on) one can in real time accurately measure the distances, find the location of point of interests, closest place for eating and so on.

Similar services can be built upon the legislation. Legislation is a kind of map, guiding through the deep forests and deep valleys. What if it would be easy to zoom in and out from the legislation, mark the points of interest, calculate the closest path between them, plan the voyage and get the feedback about the progress made meanwhile as we do it on GPS devices enforced with digital maps and data transfer options today?

The presented approach offers a method to create a systematic alternative structure to any naturally evolved normative system. The method developed within the framework of this thesis is not limited to legislation although derived from it, nor is it limited to a culture or language. Whenever there is a possibility and need to analyse a text and separate different type of words from it, it is relatively easy to create a lightweight digitalised layer over it and make it easily computable. Any type of text collection- document database, literature, even World Wide Web can be represented as two-way tables and derive quite some practical and general conclusions from that. The method does not need a lot of computing power, which makes it accessible and relatively cheap for applying and consuming as the operations and decisions mostly use only digitalised, condensed layer and the actual text is accessed only in the final phase.

8.1.3 Conformity with the EU initiatives

Another way to assess the results is to put it online with similar research going on in the field of legislative informatics from one side and considering the legislation specific developments (described respectively in

subdivision 5.2 and 6) from the other. This leads us to describe the possible positive aspects of this research a bit deeper, linked to the simplification procedures outlined above:

Impact assessment: A visualized layout of the legal act presents a suitable tool for problem estimation, description, decomposition and negotiations between parties. This helps in the process of consultation with stakeholders, creating a good basis for comparison with which to measure the achieved transparency.

Consultation: A visualized layout of the legal act presents a suitable tool for problem spotting, description, decomposition and negotiations between parties. Visual layout is good support for creating a common understanding about the issue and the ways to analyse the influence of decisions.

Expertise: Visual layout gives a bird-view of the legal domain, offering an overview of relations and locations affecting the legal question.

Administrative costs: The structure of legal acts and connections of different legal acts form a good starting point for a cross-nation area comparisons and making savings in data handling processes.

Choice of regulatory instruments: The location and role of the proposed legal act can be tested in visual layouts to estimate the potential effect of different legal acts and for the legislation as whole.

Simplification: Different viewing and representing options based on different connections between legal acts allows broadening or narrowing the scope of comparison, to represent a collection of connected documents and parts of it, to merge or decompose parts of existing legislation and so on, to provide an overview of the legislation based on which simplification can be decided.

Accessibility/presentation of EU law: Visualized and compressed layout of the legal text helps to access and present the legal content more efficiently and it can be read quicker.

Inter-institutional coordination: Visualized and compressed layout of the legal text helps to get a quick overview of the content and applied changes.

Table 8. A cross table reflecting the on-going research in the field of subject in comparison with EU activities for better regulation

EC Better Regulation activity		Legal informatics research field									Compute-rizable?
		Metalex	Complexity	Visualisation	Syntactic analysis	Multilingual drafting	Meta drafting	Content management	STIA	Change management	
Impact assessment	Impact coherency analysis	1	0	1	0	0	0	0	0	1	1
	Consideration of various policies	0	0	1	0	0	0	0	0	0	1
	Consultation of stakeholders	0	1	1	1	0	0	0	0	1	1
	Achieved transparency	1	1	1	1	1	0	0	0	1	1
Consultation		0	1	1	0	0	1	0	1	1	1
Expertise		0	1	1	1	1	0	1	0	0	1
Administrative costs	Cross-country area comparisons	1	2	1	0	1	1	1	0	1	2
	Procedural differences	1	0	0	0	1	0	1	0	1	2
	Offered economies	2	1	2	0	1	1	1	0	1	2
Choice of regulatory instruments	Regulations and directives usage	1	1	1	1	0	0	0	0	0	1
	Reviewing/sun-setting clauses	1	1	1	1	1	1	0	0	0	1
	Alternative instruments usage	1	1	1	1	1	1	0	0	0	1
Transposition/application of EU law		2	0	1	0	2	1	2	0	2	2
Simplification	Repeal	1	1	1	1	1	1	0	1	1	1
	Codification	2	1	2	1	2	1	0	1	1	2
	Recasting	2	1	2	1	1	1	0	0	1	2
	Co-regulation	1	1	2	1	1	1	0	0	1	1
Accessibility/presentation of EU law	EUR-Lex website	2	1	2	2	2	1	2	2	1	2
	Pre-Lex database	2	1	2	2	2	1	2	2	1	2
Evaluation		1	0	1	0	0	1	1	0	0	1
Inter-institutional coordination	Impact assessment	0	0	1	1	0	0	0	0	0	1
	Co-ordination and transparency	2	1	1	0	1	1	2	1	1	2
	Alternative regulatory instruments	1	1	1	1	1	0	1	0	1	1
	Adoption of proposals to simplify EU law	1	1	1	0	0	0	0	0	0	1

0 No support
1 Work process support
2 Content handling support

Table 8 gives a general picture about specific research within similar research activities and how these can contribute to increasing the quality of EU regulation in the broad sense. The table represents the indicative potential of chosen researches to add value to the problem solution. The ability to contribute is described by three different values, where meanings of the digits are:

0 –no or very little value

1 –ability to support problem solution process somehow

2 – ability to deal with problem through the use of novel solutions and IT support.

A column "Computerizable?" is summarizing up the potential of all the researches going on and uses the same indications to describe the extent and ability potential of IT support for problem solution in close range future. As it appears all the above mentioned issues are under development and have a potential to pay back the expected results to the society during next 10 years.

8.2 Contribution

Main contributions of this thesis:

- A novel approach for legal text visualisation and increasing the computational linguistic support for legal texts provided with experimental results;
- A prototype for testing proposed legal text visualisation approach;
- A novel method for measuring the similarity of legal texts provided with experimental results;
- A novel approach for legislation restructuring provided with experimental results.

8.3 Future Research

Several aspects of this action research need some additional research. In order to open some perspectives for the future:

The existing textual references within legal texts form a natural part of the legislation. Combination of textual and similarity measurement related references could give valuable insights and enrich greatly the visual presentation. Therefore an algorithm combining indications from both types of references should be developed and applied for getting more targeted results.

- Presented graph expresses the hidden structure of only one layer, acts layer. Similar structure can be found in each level of normative

system and it is possible to connect these levels to other levels. Forming a connected graph of the whole normative system is the next target.

- The proposed approach is not quite suitable to take into account very short or very specific (state rewarding related etc.) legal texts. This is a problem because these acts form a natural part of the normative system and should not be excluded. A suitable method to create readable visual layouts including very small graphs should be developed.
- Word usage within the legislation is regulated but not to the full extent. How to take account of variations in meanings and ways of application?
- If the sources of law are located with required exactness how to extract the content of it in an automated way? How to use it for reasoning? How to assist users to plan their further activities in good accordance with a law?

9 References

[1] J. Sheridan, Written Laws are Like Spiders Webs, in Proceedings of the 2010 conference on Legal Knowledge and Information Systems: JURIX 2010: The Twenty-Third Annual Conference , Amsterdam, 2010.

[2] K. Merusk and R. Narits, Õiguse olemusest ja seaduse võimalustest, Riigikogu Toimetised, vol. 2, 2000.

[3] J. McCarthy, What is artificial intelligence?, 12 November 2007. [Online]. Available: http://www-formal.stanford.edu/jmc/whatisai/whatisai.html. [Accessed 20 October 2012].

[4] T. Bench-Capon, P. Visser, Deep models, ontologies and legal knowledge based systems, in Legal knowledge based systems. JURIX 1996: The nineth annual conference, 1996.

[5] T. R. Gruber, The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases, in Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, 1991.

[6] Ontology, Merriam Webster, Incorporated, [Online]. Available: http://www.merriam-webster.com/dictionary/ontology. [Accessed 20 October 2012].

[7] A. Valente, Legal Knowledge Engineering; A Modelling Approach, Amsterdam: IOS Press, 1995.

[8] M. Palmirani, R. Brighi, M. Massini, Automated Extraction of Normative References in Legal, Proceedings of the 9th international conference on Artificial intelligence and law, Edinburgh, 2003 .

[9] R. Winkels, M. Breebaart, Constructing a Semantic Network for Legal Content. Proceedings of the 10th international conference on Artificial intelligence and law, Bologna, 2005.

[10] I. Liiv, A. Vedeshin, E. Täks, Visualization and structure analysis of legislative acts: a case study on the law of obligations. Proceedings of the 11th international conference on Artificial intelligence and law, Stanford, 2007.

[11] D. Bourcier, P. Mazzega, Codification, Law Article and Graphs, Proceeding of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference, Amsterdam, 2007.

[12] P. Mazzega, D. Bourcier, R. Boulet, The network of French legal codes, in Proceedings of the 12th International Conference on Artificial Intelligence and Law, Barcelona, Spain, ACM, 2009, pp. 236-237.

[13] J. E. Chikofsky, H. J. Cross, Reverse engineering and design recovery: a taxonomy, Software, vol. 7, no. IEEE, pp. 13-17, January 1990.

[14] C. Biagioli, Towards a legal rules functional micro-ontology, in Proceedings of the First International Workshop on Legal Ontologies, Un. Of Melbourne, Law School, Melbourne, Victoria, Australia, 1997.

[15] M. Nakamura, S. Nobuokau, A. Shimazu, Towards translation of legal sentences into logical forms, in Proceedings of the 2007 conference on New frontiers in artificial intelligence, JSAI'07, Miyazaki, Japan, 2008.

[16] A. Aarnio, The Rational as Reasonable. A Treatise on Legal Justification, Blackwell Publishing Ltd, 1988.

[17] E. de Maat, R. Winkels, Automated Classification of Norms in Sources of Law, in Semantic Processing of Legal Texts, E. Francesconi, S. Montemagni, P. Wim and D. Tiscornia, Eds., Springer Berlin / Heidelberg, 2010, pp. 170-191.

[18] J. C. Fillmore, The Case for Case, vol. Universals in Linguistic Theory, New York: Holt, Rinehart, and Winston, 1968, pp. 1-88.

[19] E. Täks, A. Lohk, An alternative method for computerized legal text restructuring., in Frontiers in Artificial Intelligence and Applications. Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference, 2010.

[20] H. C. Aggarwal, H. Wang, Managing and mining graph data., vol. Volume 40 of Advances in Database Systems, Springer, 2010.

[21] L. K. Vohandu, Analysis of Biological Systems with many Variables., in III Conference of Mathematical Methods in Biology., 1961.

[22] M. P. Chandrasekharan, R. Rajagopalan, ZODIAC - an algorithm for concurrent formation of part-families and machine-cells, International Journal of Production Research, vol. 25, no. 6, 1987.

[23] ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability, International Organzation for Standardzation, 1998.

[24] K.-M. Nyman-Metcalf, E. Täks, Simplifying the law - can ICT help us?, International Journal of Law and Information Technology, Oxford Journals, 2013; doi: 10.1093/ijlit/eat003

[25] E. Täks, L. Vohandu, A. Lohk, I. Liiv, An Experiment to Find a Deep Structure of Estonian Legislation. Katie M. Atkinson (Ed.). Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference, IOS Press, 2011, pp. 93-102.

[26] E. Täks, L. Võhandu, A. Lohk, K. Nyman-Metcalf, A. Rull, A tool for exploring the hidden structure of legislation, in Proceedings of the fundamental concepts and the systematisation of law. Workshop at Jurix 2011 in Vienna, 2011.

[27] J. P. Mallory, D. Q. Adams, Encyclopedia of Indo-European Culture, London: Fitzroy Dearborn Publishers, 1997.

[28] K. A. Bard, Encyclopedia of the Archaeology of Ancient Egypt, London: Routlege, 1999.

[29] Code of Hammurabi, [Online]. Available: http://www.britannica.com/EBchecked/topic/253710/Code-of-Hammurabi. [Accessed 2012 August 15].

[30] J. M. Kelly, A Short History of Western Legal Theory, New York: Oxford University Press, 1992.

[31] Roman law, [Online]. Available: http://www.britannica.com/EBchecked/topic/507759/Roman-law. [Accessed 2012 August 15].

[32] K. A. Parrow, Roman Law. Europe, 1450 to 1789: Encyclopedia of the Early Modern World. 2004. Encyclopedia.com., 2003. [Online]. Available: http://www.encyclopedia.com/doc/1G2-3404900622.html. [Accessed 2012 August 16].

[33] A. Berger, Encyclopedic Dictionary of Roman Law, vol. 43, Philadelphia: American Philosophical Society, 1953.

[34] N. W. E. contributors, Justinian I, 2 April 2008. [Online]. Available:

http://www.newworldencyclopedia.org/p/index.php?title=Justinian_I&oldid=680773. [Accessed 17 August 2012].

[35] Roman Law. West's Encyclopedia of American Law, [Online]. Available: http://www.encyclopedia.com/doc/1G2-3437703852.html. [Accessed 17 ugust 2012].

[36] Gratian's Decretum. Encyclopaedia Britannica, [Online]. Available: http://www.britannica.com/EBchecked/topic/242263/Gratians-Decretum. [Accessed 2012 August 20].

[37] T. Wallinga, The Common History of European Legal Scholarship, Erasmus Law Review, Vol. 4, No. 1, 2011 , 2011.

[38] J. Kuusik, K. Miil, Guide to Estonian Legal System and Legal Research, Hauser Global Law School Program, New York University School of Law, 12 01 2011. [Online].

[39] R. Narits, Interpretation of Law in the Estonian Legal System, Juridica International. Law Review. University of Tartu, vol. I, pp. 11-16, 1996.

[40] F. C. v. Savigny, System des heutigen Römischen Rechts, vol. I, Berlin, 1840, p. 213.

[41] R. Zippelius, Juristische Methodenlehre, München, 1990, p. 36.

[42] A. Aarnio, Laintulkinnan teoria., Juva, p. 19, 1988.

[43] Law, [Online]. Oxford Dictionaries, Available: http://oxforddictionaries.com/definition/english/law?q=law. [Accessed 23 September 2012].

[44] Law, [Online]. Available: http://www.merriam-webster.com/dictionary/law. [Accessed 23 September 2012].

[45] R. Cotterrell, Law's community: legal theory in sociological perspective, Oxford: Oxford University Press, 1997.

[46] T. Honoré, About law, New York: Oxford University Press Inc, 1996.

[47] L. L. Fuller, The morality of law, Yale University Press, 1964.

[48] M. K. Sein, G. Harindranath, Conceptualizing the ICT Artifact: Toward Understanding the Role of ICT in National Development, The Information Society, vol. 20, p. 15–24, 2004.

[49] Å. Grönlund, State of the Art in e-Gov Research – A Survey, Electronic Government, vol. 3183/2004, 2004.

[50] e-Government, [Online]. Available: http://www.olis.oecd.org/olis/2001doc.nsf/c5ce8ffa41835d64c125685d005300b0/0b677ed527d35bc0c1256b21004f4b6a/$FILE/JT00118445.PDF. [Accessed 7 September 2012].

[51] Introduction to e-Government: Overview of key concepts, [Online]. Available: http://www.worldbank.org/egov. [Accessed 17 September 2012].

[52] E-Government Overview, [Online]. Available: http://www2.unpan.org/egovkb/datacenter/CountryView.aspx?reg=EUROPE. [Accessed 25 September 2012].

[53] M. Castells, The Rise of the Network Society, Oxford: Blackwell Publishers Inc, 1996.

[54] M. Biasiotti, E. Francesconi, M. Palmirani, G. Sartor, F. Vitali, Legal Informatics and Management of Legislative Documents, Global Centre for ICT in Parliament, 2008.

[55] A. Boer, R.Hoekstra, R.Winkels, METALex: Legislation in XML, in Proceedings of JURIX 2002: The Fifteenth Annual Conference, Amsterdam, 2002.

[56] V. Perton, N. Cole, F. Andrighetto, M. Delahunty, C. Furletti, M. Gould, N. Maughan, A. Paterson, T. Robinson, Technology & the law, Government Printer, Parliament of Australia, Law Reform Committee, Melbourne, 1999.

[57] Communication from the Commission to the Council and the European Parliament - Better Regulation for Growth and Jobs in the European Union {SEC(2005) 175} COM/2005/0097 final, 2005. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0097:FIN:EN:PDF. [Accessed 5 3 2011].

[58] Communication from the Commission - European Governance: Better lawmaking COM/2002/0275 final, 2002. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0275:FIN:EN:PDF. [Accessed 5 3 2011].

[59] OECD, Improving the Quality of Laws and Regulations: Economic, Legal and, 1994.

[60] OECD, Regulatory Impact Analysis: Best Practice in OECD

countries, Paris, 1997.

[61] C. Biagioli, A. Passerini, S. Montemagni, C. Soria, Automatic semantics extraction in law documents, Bologna, Italy, 2005.

[62] M. Palmirani, R. Brighi, Time model for managing the dynamic of normative system, in EGOV'06 Proceedings of the 5th international conference on Electronic Government , Berlin, 2006.

[63] R. Winkels, A. Boer, E. d. Maat, T. v. Engers, M. Breebaart, H. Melger, Constructing a semantic network for legal content, in ICAIL '05 Proceedings of the 10th international conference on Artificial intelligence and law , New York, 2005.

[64] L. D. Edwards, E. D. Mahling, Toward knowledge management systems in the legal domain, in Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge, Phoenix, 1997.

[65] European Commission, 25 02 2011. [Online]. Available: http://ec.europa.eu/governance/better_regulation/. [Accessed 5 March 2011].

[66] T. Agnoloni, L. Bacci, E. Francesconi, P. Spinosa, D. Tiscornia, S. Montemagni, G. Venturi, Building an ontological support for multilingual legislative drafting, in Proceeding of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference, Amsterdam, 2007.

[67] M. J. Bommarito II, D. M. Katz, A Mathematical Approach to the Study of the United States Code, Physica A, vol. Vol. 389, March 25, 2010.

[68] M. J. Bommarito II, D. M. Katz, Properties of the United States Code Citation Network, Working Paper Series, no. SSRN eLibrary, November 2009.

[69] F. B. Cross, T. A. Smith, The Reagan Revolution in the Network of Law, no. http://ssrn.com/paper=909217, 2006.

[70] P. Zhang, L. Koppaka, Semantics-Based Legal Citation Network. ICAIL '07: Proceedings of the 11th international conference on Artificial intelligence and law, Stanford, Californi, 2007.

[71] Technical Rules for Drafts of Legislation of General Application, Government of the Republic Regulation No. 279 of 28 September 1999, [Online]. Available:

http://www.legaltext.ee/et/andmebaas/document.asp?ptyyp=RT&q2=normitehnika&order=TA&tyyp=X&query=&display=1&nupp=Otsi! . [Accessed 30 September 2012].

[72] R. Narits, Systematisation of Objective Law: From Codification to Reformation of Law, Juridica International. Law Review. University of Tartu, vol. X, pp. 161-167, 2005.

[73] L. Rodriguez, Constitutional and statutory limits for cost-benefit analysis pursuant to Executive Orders 12,291 and 12,498, Boston College Environmental Affairs Law Review, Spring88, Vol. 15 Issue 3/4, p. p 505, 1988.

[74] Regulatory Process Reform From Ford to Clinton, [Online]. Available: http://www.cato.org/pubs/regulation/regv20n1/reg20n1a.html. [Accessed 22 September 2012].

[75] Executive Order 12291, [Online]. Available: http://www.thecre.com/pdf/EO12291.PDF. [Accessed 22 September 2012].

[76] A. Renda, Law and Economics in the RIA World: Improving the Use of Economic Analysis in Public Policy and Legislation, Intersentia, 2011.

[77] Simpler Community Legislation, [Online]. Available: http://ec.europa.eu/internal_market/simplification/index_en.htm#slim . [Accessed 24 September 2012].

[78] Press release, [Online]. Available: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/97/792&format=PDF&aged=1&language=EN&guiLanguage=en-HZT24QTE0oDAAg&usg=. [Accessed 23 September 2012].

[79] European Business Test Panel, [Online]. Available: http://ec.europa.eu/yourvoice/ebtp/index_en.htm. [Accessed 23 September 2012].

[80] D. Mandelkern, E. A. Garcia-Manso, K. Burger, M. Cabane, L. Carbone, B. Chavanat, C. Church, D. Curran, E. D. Pue, P. G. Michael Fruhmann, R. Hayder, L. Hjortnæs, P. Karkatsoulis, E. Milleker, Mandelkern Group on Better Regulation., European Commission, 2001.

[81] Communication from the Commission to the European Parliament,

the Council, the Economic and Social Committee and the Committee of the Regions - The Commission's legislative and work programme for 2003, 2002. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0590:FIN:EN:PDF. [Accessed 5 March 2011].

[82] Communication from the Commission - Towards a reinforced culture of consultation and dialogue - General principles and minimum standards for consultation of interested parties by the Commission. COM/2002/0704 final, 2002. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0704:FIN:EN:PDF. [Accessed 5 March 2011].

[83] Communication from the Commission on the collection and use of expertise by the Commission: principles and guidelines - Improving the knowledge base for better policies COM/2002/0713 final, 2002. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52002DC0713:EN:NOT. [Accessed 5 March 2011].

[84] Interinstitutional agreement on better law-making (2003/C 321/01), 2003. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2003:321:0001:0005:EN:PDF. [Accessed 5 3 2011].

[85] Communication from the Commission - Action plan Simplifying and improving the regulatory environment COM/2002/0278 fina, 2002. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0278:FIN:EN:PDF. [Accessed 5 3 2011].

[86] Commission Manual on Legislative Drafting, 1997. [Online]. Available: http://ec.europa.eu/governance/better_regulation/documents/legis_draft_comm_en.pdf. [Accessed 5 3 2011].

[87] Communication to the commission from ms Grybauskaité in agreement with the President. Responding to strategic needs: reinforcing the use of evaluation. SEC(2007)213, 2007. [Online]. Available: http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_comm_sec_2007_213_en.pdf. [Accessed 5 3 2011].

[88] R. Brighilesmo, A. Mazzei, M. Palmirani, D. Radicioni, Towards

Semantic Interpretation of Legal Modifications through Deep Syntactic Analysis, in Proceeding of the 2008 conference on Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference, Amsterdam, 2008.

[89] G. Peruginelli, F. Bargellini, Drafting Legislation with Ontology-based Support Project, [Online]. Available: http://www.dalosproject.eu/. [Accessed 26 September 2012].

[90] C.Biagioli, D.Grossi, Formal Aspects of Legislative Meta-Drafting, in eeding of the 2008 conference on Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference, Amsterdam, 2008.

[91] A.Boer, R.Winkels, T. Engers, E. Maat, A Content Management System based on an Event-based Model of Version Management Information in Legislation, in Legal Knowledge and Information Systems. Jurix 2004., Amsterdam, 2004.

[92] M. T. Pazienza, N. Scarpato, A. Stellato, STIA: Experience of Semantic Annotation in Jurisprudence Domain, in Proceeding of the 2009 conference on Legal Knowledge and Information Systems: JURIX 2009: The Twenty-Second Annual Conference, Amsterdam, 2009.

[93] M. Palmiran, L. Cervone, Legal Change Management with a Naitive XML Repository, in Frontiers in Artificial Intelligence and Applications. Legal Knowledge and Information Systems - JURIX 2009: The Twenty-Second Annual Conference, 2009.

[94] J. P. Gibbs, Norms: The Problem of Definition and Classification, vol. 70, no. 5, 1965.

[95] A. Boer, T. Engers, R. Winkels, Mixing Legal and Non-legal Norms, Bologna, Italy, 2005.

[96] A. Wagner, S. Cacciaguidi-Fahy, Obscurity and clarity in the law: prospects and challenges, Ashgate Publishing, Ltd, 2008.

[97] Clause, Merriam-Webster, Inc, [Online]. Available: http://www.merriam-webster.com/dictionary/clause. [Accessed 17 07 2010].

[98] Definition: Subject, Predicate, Webster, [Online]. Available: http://www.websters-online-dictionary.org/definitions/predicate?cx=partner-pub-

093945075352974 4%3Av0qd01-tdlq&cof=FORID%3A9&ie=UTF-8&q=predicate&sa=Search#906. [Accessed 17 07 2010].

[99] Süstemaatiline liigitus, Justiitisminsiteerium, [Online]. Available: https://www.riigiteataja.ee/jaotused.html?jaotus=S%C3%9CSTJAOT . [Accessed 9 October 2012].

[100] I. Liiv, R. Öpik, J. Übi, J. Stasko, Visual matrix explorer for collaborative seriation, Wiley Interdisciplinary Reviews: Computational Statistics, vol. 4, no. 1, p. 85–97, Issue.

[101] J. E. Mullat, Extremal subsystems of monotonic systems, Automation and Remote Control, vol. 37, p. 758–766, 1976.

[102] L. Vyhandu, Some methods to order objects and variables, Trans of Tallinn University of Technology, vol. 482, p. 43–50, 1980.

[103] W. T. McCormick, P. J. Schweitzer and T. W. White, Problem decomposition and data reorganization by a clustering technique, Operations Research , vol. 20, p. 993–1009, 1972.

[104] J. R. King, V. Narkonchaiv, Machine-component group formation in group technology: review and extension, International Journal of Production Research, vol. 20, no. 2, pp. 117-133, 1982 .

[105] M. P. Chandrasekharan, R. Rajagoplan, MODROC: an extension of rank order clustering for group technology, International Journal of Production Research, vol. 24, no. 5, pp. 1221-1233, 1986.

[106] S. Kaparthi, N. C. Suresh, Machine-component cell formation in group technology: a neural network approach, International Journal of Production Research, vol. 30, no. 6, pp. 1353-1367, 1992.

[107] A. Kusiak, Y. Chung, GT/ART: using neural network to form machine cells., Manufactoring Review, vol. 4, no. 4, pp. 293-301, 1991.

[108] S. H. S. Maine, Ancient Law: its connection with the early history of society, and its relation to modern ideas, New York: Dorset Press, 1986.

[109] Interinstitutional Agreement of 28 November 2001 on a more structured use of the recasting technique for legal acts OJ C 77, 2001. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2002:077:0001:0003:EN:PDF. [Accessed 5 3 2011].

[110] T. R. Gruber, A Translation Approach to Portable Ontology

Specifications, Knowledge Acquisition, Vols. 199-199, 1993.

[111] OECD, Information technology and new growth opportunities., Information, Computer and Communications Policy Series, ICCP, vol. 19, 1989.

[112] Federal Property and Administrative Services Act of 1949, in Public Law 89-306, 40 U.S.C. 759., 1949.

[113] Access America: Reengineering Through Information Technology, February 1997. [Online]. Available: http://govinfo.library.unt.edu/npr/library/announc/access/acessrpt.html. [Accessed 18 September 2012].

[114] Å. Grönlund, Introducing e-Gov:History, definitions and issues, Communications of the Association for Information Systems, vol. 15, 2004.

[115] M. Palmirani, R. Brighi, Electronic Government. Lecture Notes in Computer Science, Berlin: Springer Berlin / Heidelberg, 2006.

[116] G. K. Hadfield, The Quality of Law: Judicial Incentives, Legal Human Capital and the Evolution of Law, Social Science Electronic Publishing, Inc, Los Angeles, 2007.

[117] W. T. McCormick, P. J. Scweitzer. Jr., T. W. White, Problem Decomposition and Data Reorganization by a Clustering Technique, no. 20:993-1009, 1972.

[118] Tsiviilseadustiku üldosa seadus, Riigi Teataja, pp. RT I 2002, 35, 216, 2002.

[119] S. Holden, The Evolution of Information Technology Management at the Federal Level: Implications for Public Administration, pp. 53-73 in Public Information Technology: Policy and Management Issues, G. David Garson, Ed., Hershey, Pa.: Idea Group, 2003.

Automaatne meetod õigusteabe hankimiseks ja visualiseerimiseks

Lühikokkuvõte

Õiguslik informaatika on uurimisharu, mis tegeleb õigusliku informatsiooni töötlemise ja õigustoimingute toetamisega info- ja kommunikatsioonitehnoloogia vahendite abil. Käesolev doktoritöö käsitleb uudset metoodikat, mis võimaldab lihtsustada ligipääsu õigusteabele ja suurendada selle kasutatavust. Probleem pärineb tegeliku elu vajadustest ning on tihedalt seotud töö autori varasema tegevusega avalikus sektoris õigusloome ja infotehnoloogia valdkonnas.

Pakutav meetod võimaldab leida tähendusrikkamad osad seadusandlusega seotud tekstidest, esitada saadud tulemus kergemini mõistetaval kujul ja muundada see edasiseks arvutitööks sobivale kujule. Meetod on interdistsiplinaarne, ühendades endas sobivaid meetodeid keeletehnoloogia, süsteemi- ja andmeanalüüsi valdkondadest. Väljatöötatud meetod võimaldab infotehnoloogia abil seadusandlust alternatiivselt ümberstruktureerida, esitada selle ülesehitust erinevates vaadetes ja lahendada mõned olemuslikud struktuursed jms probleemid. Meetodi peamine väärtus seisneb selle võimes alandada seadusandluse ligipääsu barjääre, tõsta erinevate kasutajate õigusliku analüüsi võimekust, kuid seejuures otseselt mõjutamata seadusandlust ja selle toimimist, jäädes samas sellega otseselt seotuks. Meetodil on palju erinevaid potentsiaalseid rakendusi seadusandluses ja ka väljaspool seda.

Kõnealust meetodit testiti Eesti seadusandluse põhjal, hõlmates selle käigus pea poolt seaduste terviktekstide mahust. Kõik meetodi rakendamise põhietapid läbiti arvuti pakutavaid võimalusi kasutades. Tulemuste illustreerimiseks koostati veebipõhine prototüüprakendus, mis võimaldab huvilistel hinnata meetodi praktilist väärtust ning kasutamiskõlbulikkust. Lisaks viidi andmeanalüüsi meetoditele tuginedes läbi uurimise all oleva seadusandluse osa struktuuri süvauuring ning sellekohased tulemused on avaldatud erialases kirjanduses akadeemilisele üldsusele nii informaatika kui õigusteaduse valdkonnas. Meetod on saanud esmase heakskiidu õigusteadlastelt Eestis ja ka väljaspool seda.

Võtmesõnad: seadusandluse struktuur, seadusandluse kasutatavus, süsteemianalüüs, keeletehnoloogia, visualiseerimine, andmeanalüüs

An automated legal content capture and visualisation method

Abstract

Legal informatics is the discipline, which deals with the use of information communication technology to process legal information and support legal activities. A new approach is presented in this thesis to ease the access to legal sources and increase the usability of legislation. The problem originates from real life needs and is based partially on the author´s personal work in the public sector in close relation with litigation.

The proposed method allows extracting the most meaningful parts of legal texts, visualise the content for human readers and represent it in a way which is easily accessible by computers. The method is based on an interdisciplinary approach, combining methods from language technology, system analysis and data analysis. The method is applied upon legal texts in order to overcome some built-in weaknesses of the legislation with the help of re-engineering and restructuring. This enhances structural exploration abilities of different users without interfering or altering the existing legislation and is tightly tied with it, to enable direct access to the sources of law. The method is in essence universal, based on extraction of meaningful words; it has many potential areas of application within and very likely outside the legal domain, which need to be explored further.

The method is tested on selected part of the Estonian legislation, whereby the text processing and data analysis was done automatically. An extra compressed data layer upon the legislation was created enabling further visual and computerized analysis of the legal domain. A web based prototype was created to demonstrate the usability of the proposed method and allow the user to get personal hands-on experience. Method outcomes were analysed in cooperation with legal scientists in order to estimate the preliminary value of the visualisation and verify the method suitability for application within the legal domain. Experimental high level structural analysis of the chosen part of the legislation was performed and a new structural perspective to it is proposed for further discussion.

Keywords: legislation, system engineering, restructuring, language processing, visualisation, data mining.

Acknowledgements

I wish to thank my wife, helping me during all the process, supporting me these years and continuing this seemingly endless journey together with two adorable children born meantime.

I thank my supervisor Rein Kuusik for continuing support and belief in me. Supervisor Katrin Merike Nyman-Metcalf was of great help in many ways providing me with valuable legal remarks and suggestions how to proceed.

I also wish to thank Professor emeritus Leo Võhandu for his influential contributions developing this method. His catching enthusiasm and amazing ability to find and apply suitable mathematical method to any problem in rather unusual research fields is very motivating and inspiring.

I wish to thank my good colleagues Ahti Lohk and Rain Öpik for finding a time and sharing my passion about this research. I was lucky to have Addi Rull close by, who has provided me with a comprehensible overview about legal language issues regarding the matter.

I would also like to express my gratitude toward other numerous people helping me to find a way through a myriad of questions. Especially I want to recognize Innar Liiv and Professor Baruch Raz for their important push toward new horizons at some very critical moments of this research.

Without you, my good friends, this would not happen.

ELULOOKIRJELDUS

Isikuandmed

Ees- ja perekonnanimi: Ermo Täks
Sünniaeg ja -koht: 15.03.1971, Viljandi
Kodakondsus: Eesti

Kontaktandmed

Aadress: Akadeemia tee 15a, 12618 Tallinn
Telefon: +3725149040
E-posti aadress: ermo.taks@ttu.ee

Hariduskäik

Õppeasutus (nimetus lõpetamise ajal)	Lõpetamise aeg	Haridus (eriala/kraad)
Tallinna Tehnikaülikool	2000	M.Sc (informaatika)
Eesti Põllumajandusülikool	1993	B.Sc (mehaanika)

Keelteoskus (alg-, kesk- või kõrgtase)

Keel	Tase
Eesti keel	Kõrgtase
Inglise keel	Kõrgtase
Vene keel	Kõrgtase
Läti keel	Algtase
Saksa keel	Algtase

Täiendusõpe

2005	**Finance and contracts in fp6** Paul Drath, Singleimage Ltd, GB	**Tallinn**
2004	**The 17th annual conference on legal knowledge and information systems** Fraunhofer Institute for Open Communication Systems (FOKUS).	**Berlin**
2004	**Tarkvara ja infosüsteemide arendamise formaalsed meetodid.** TTÜ Küberneetika Instituut	**Tallinn**

2001	**Nõustav müük** Vain SMD	**Tallinn**
2000	**Meeskonnatöö** Invicta	**Pärnu**
1999	**IT juhtide strateegiline koolitus** Baltic Computer Systems	**Tallinn**
1997	**Seminar of the Danish EU decision – making procedure and the parliamentary control** Taani Välisministeerium	**Kopenhaagen**

Teenistuskäik

Töötamise aeg	Tööandja nimetus	Ametikoht
1996 – 1997	Vanglate Amet	IT peaspetsialist
1997 – 1999	Justiitsministeerium	Infosüsteemide osakonna spetsialist
1999 – 2000	Sideamet	IT peaspetsialist
2000 – 2001	Riigikantselei	Majandusosakonna informaatikatalituse juhataja
2001 – 2002	Oracle Baltics	Avaliku sektori äriarendusjuht Balti riikides
2002 –…	Tallinna Tehnikaülikool	Lektor
2005 –…	Stainless Surface Technology OÜ	Tegevjuht

Teadustegevus:

Liiv, A. Vedeshin, E. Täks, Visualization and structure analysis of legislative acts: a case study on the law of obligations. Proceedings of the 11th international conference on Artificial intelligence and law, Stanford, 2007.

E. Täks, A. Lohk, An alternative method for computerized legal text restructuring., in Frontiers in Artificial Intelligence and Applications. Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference, 2010.

E. Täks, L. Vohandu, A. Lohk, I. Liiv, An alternative method for computerized normative system, in ICAIL 2011.

E. Täks, L. Võhandu, A. Lohk, K. Nyman-Metcalf, A. Rull, A tool for exploring the hidden structure of legislation, in Proceedings of the fundamental concepts and the systematisation of law. Workshop at Jurix 2011 in Vienna, 2011.

E. Täks, K. Nyman-Metcalf, Simplifying the law - can ICT help us?, International Journal of Law and Information Technology, Oxford Press, teade vastuvõtmise kohta.

Kaitstud lõputööd
Infosüsteemide rakendusprobleeme
M.Sc. (informaatika) Tallinna Tehnikaülikool, 2000
B.Sc. (mehhaanika) Eesti Põllumajandusülikool, 1993

Teadustöö põhisuunad
Seadusandluse struktuur, seadusandluse kasutatavus, süsteemianalüüs, visualiseerimine, Keeletehnoloogia, andmeanalüüs

Muud uurimisprojektid
eVikings II
Establishment of the Virtual Centre of Excellence for IST RTD in Estonia
FP5 IST accompanying measures project IST-2001-37592

Evolutsioneeruvate infosüsteemide mudelipõhine loomine ja haldamine
SF 2010, Projekt SF0140013s10

CURRICULUM VITAE

Personal data

Name: Ermo Täks
Date and place of birth: 15.03.1971, Viljandi

Contact information

Address: Raja 15, 12618 Tallinn
Phone: +3725149040
E-mail: ermo.taks@ttu.ee

Education

Educational institution	Graduation year	Education (field of study/degree)
Tallinn University of Technology	2000	M.Sc (informatics)
Estonian Agricultural University	1993	B.Sc (mechanics)

Special Courses

2005	**Finance and contracts in fp6** Paul Drath, Singleimage Ltd, GB	**Tallinn**
2004	**The 17th annual conference on legal knowledge and information systems** Fraunhofer Institute for Open Communication Systems (FOKUS).	**Berlin**
2004	**Formal methods in software and system design** Institute of Cybernetics	**Tallinn**
2001	**Advisory selling** Vain SMD	**Tallinn**

2000	**Teamwork training** Invicta	**Pärnu**
1999	**Strategic training for it managers** Baltic Computer Systems	**Tallinn**
1997	**Seminar of the Danish EU decision – making procedure and the parliamentary control** Danish Foreign Ministry	**Kopenhaagen**

Language competence/skills (fluent; average, basic skills)

Language	Level
Estonian	High level
English	High level
Russian	High level
Latvian	Basic
German	Basic

Professional Employment

Period	Organisation	Position
1996 – 1997	Estonian Prison Department	IT Chief Specialist
1997 – 1999	Ministry of Justice	IT department specialist
1999 – 2000	Estonian National Communication Board	IT Chief Specialist
2000 – 2001	Estonian State Chancellery	IT development manager
2001 – 2002	Oracle Baltics	Business Development Manager for Public Sector in Baltic States
2002 –…	Tallinn University of Technology	Lector
2005 –…	Stainless Surface Technology OÜ	Chief executive manager

Scientific work

Liiv, A. Vedeshin, E. Täks, Visualization and structure analysis of legislative acts: a case study on the law of obligations. Proceedings of the 11th international conference on Artificial intelligence and law, Stanford, 2007 .

E. Täks, A. Lohk, An alternative method for computerized legal text restructuring., in Frontiers in Artificial Intelligence and Applications. Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference, 2010.

E. Täks, L. Vohandu, A. Lohk, I. Liiv, An alternative method for computerized normative system, in ICAIL 2011.

E. Täks, L. Võhandu, A. Lohk, K. Nyman-Metcalf, A. Rull, A tool for exploring the hidden structure of legislation, in Proceedings of the fundamental concepts and the systematisation of law. Workshop at Jurix 2011 in Vienna, 2011.

E. Täks, K. Nyman-Metcalf, Simplifying the law - can ICT help us?, International Journal of Law and Information Technology, Oxford Press, accepted for publication

Defended theses

Certain implementation problems of the information systems in Public Sector
M.Sc. (informaatika) Tallinn University of Technology, 2000
B.Sc. (mechanics) Estonian Agricultural University, 1993

Current research topics

Legislation, system analysis, restructuring, language processing, visualisation, data mining.

Other research projects

eVikings II
Establishment of the Virtual Centre of Excellence for IST RTD in Estonia
FP5 IST accompanying measures project IST-2001-37592

Model-based Creation and Management of Evolutionary Information System
SF 2010, Project SF0140013s10

PAPER I

I. Liiv, A. Vedeshin, E. Täks, Visualization and structure analysis of legislative acts: a case study on the law of obligations, In Proceedings of the 11th international conference on Artificial intelligence and law (ICAIL '07), ACM, New York, 2007, pp. 189-190

Visualization and Structure Analysis of Legislative Acts: A Case Study on the Law of Obligations

Innar Liiv, Anton Vedeshin, Ermo Täks
Department of Informatics, Tallinn University of Technology
15 Raja Street, 12618 Tallinn, Estonia
innar.liiv,anton.vedeshin,ermo.taks@ttu.ee

ABSTRACT

This paper introduces visualization methods for an initial approximation of the structure in any legislative act, procedures for relevance calculation and metrics for calculating the distance between the sections. To evaluate the methodology in real world scenarios, experimental results for the law of obligations are included and analyzed. The results of the study show that it is possible to get insightful and interpretable results when analyzing only the inner-structure of the legislative act constructed by references between the sections. Methods and metrics presented in this paper can be directly applied to current keyword-based searching systems of legal texts for relevance ranking enhancement.

1. INTRODUCTION

Individual sections in legislative acts are not of equal importance. Evaluating the influence of an individual section typically requires expert judgement and understanding of the whole system and its underlying structure. Our goal is to develop a methodology that could present approximate estimations of importance based on the structure of references in the legal text. Such approach requires no extra information and minimal or no changes from the current systems available. Possible applications include: advanced expert systems for citizens, enhanced search engines, legal drafting tools. Evaluation of influence could also help officials to estimate the impact of changes to be made in the legal text. A related research has been reported by Palmirani et al. [1] concerning the extraction of references from unstructured legal text. The initial parsing phase of the research is rather similar - as references are expressed in a natural language, the research group developed a prototype to identify references using regular expressions. However, the goal of that study was not to use the extracted information for further analysis, but to standardize and rewrite (tag) the references to allow unique and undubious identification. In this article, we take that step forward and present two kinds of approaches for such an analysis - human-interpretable visualization techniques that enable the perception of the inner structure of any legal text and computer-implementable methods for ranking and measuring the sections.

ICAIL '07 June 4-8, Palo Alto, CA USA

2. PROPOSED METHODOLOGY

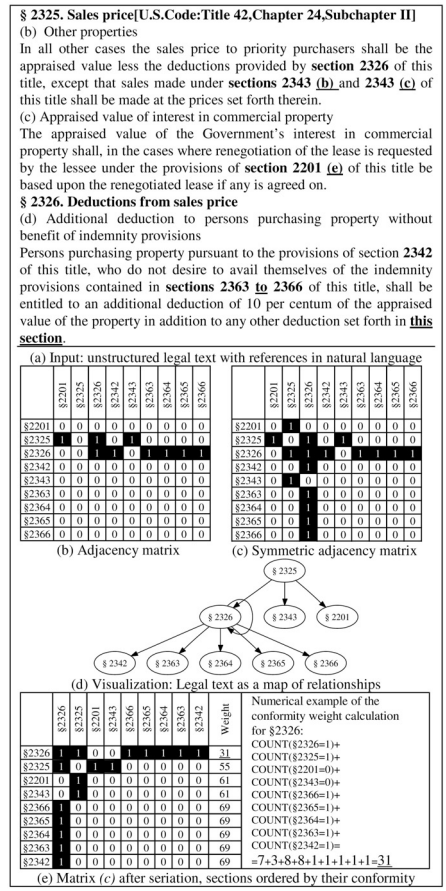


Figure 1: A walkthrough of the proposed methodology

We use legal text from U.S. code as a simple example (Fig.1) to establish common understanding of the methodology. The first step is to extract references from the initial unstructured legal text using regular expressions. This parsing and conversion process is not trivial due to variety of possible phrases for referencing, lack of textual referencing standards, references to multiple sections, ranges and comma separated nonconsecutive sections. References identified in the example (Fig.1.a) have been highlighted to illustrate the construction of adjacency matrices (Fig.1.b and Fig.1.c). Self-references were allowed, but references to subsections were merged with references to parent sections (underlined in the example). From adjacency matrix, we can draw a graph (Fig.1.d) to get the first insight of the complexity in the legislative act's reference system. Extra constraints should be set for minimizing the crossings of the edges - this allows the natural emergence of the inner structure and clusters. Such visualization method is informative and intuitive for manual inspection, but unsuitable for automatic numerical relevance calculations. A classical centrality measure, *betweenness*, from the area of social network analysis [2], should be used in this case to get a numerical importance estimations for sections comparable to manual judgements inferred from Fig.2. The final step is to apply matrix seriation method called conformity analysis [3] (see Fig.1.e for a numerical example of the calculation), which enables us to align the sections according to the typicality of their connections. For bigger datasets, in order to open the inner structure of references more efficiently, we will pursue the following additional procedure [3]. After calculation of conformity weights, we will find the least typical element, eliminate it and recalculate the weights. Eventually, after repeating the procedure until no elements are left, we will get the list of eliminations, which gives us better rank order (e.g. new order for rows and columns) than pure scale of conformity. It should be noted, that in our case, the scale is inverted, as we are dealing with a sparse structure, where the most typical section is a section with no references.

3. EXPERIMENTS AND CONCLUSION

The aim of the experiments is to investigate the applicability of proposed methods in real world scenarios. The Estonian Law of Obligations (*Võlaõigusseadus*) was selected for this research because of the large amount of sections (1075) and references between them (480). Due to the dimensions of the adjacency matrix (1075x1075), we plotted matrices as bitmap images (Fig.3), black dots represent references (positive values in the matrix). In the sense of a graph problem, we are dealing with a sparse matrix and a very specific structure. From the initial matrix, (Fig.3,left) we are able to identify a majority of connections near the diagonal, which can be explained as referencing mainly neighbouring sections and seldom to other regions. From practical point of view, such structure to be analyzed is very similar to knowledge discovery from retail transaction history, where assortment changes extremely frequently and purchases are not linked with loyalty cards. With both of those cases and structures, we are dealing with indirect relations and patterns, where classical data mining methods like clustering and association rules fail to give reasonable results. Seriation result is a chain of relations with underlying typicality scale and direction. It can also be used to reorder matrix elements (Fig.3,right), which results a great compression of the reference area, allowing the emergence of chains and clusters. Both of the visualization methods complement each other: if graph visualization (Fig.2) describes only the attraction of the elements, then seriation illustrates both -

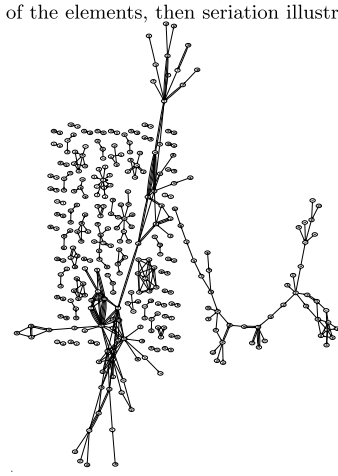


Figure 2: Chains of references and clusters

structure and the lack of structure, including the balance between elements attracting and repelling in the system. Such methods enable us to present the legislative act as a map of relationships with clusters, chains and hubs. Finally, as to the implementation status of our methodology, we have a running prototype. Implementation (PHP scripting language), along with the initial legislative act, transformed dataset and results are available electronically upon request for benchmarking and research purposes.

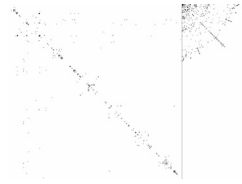


Figure 3: Initial and reordered matrix of references

4. ACKNOWLEDGMENTS

This work was supported in part by the Estonian Information Technology Foundation under Grant 06-03-99-08.

5. REFERENCES

[1] Palmirani, M., Brighi, R., and Massini, M. Automated extraction of normative references in legal texts. In *Proc. of ICAIL'03*, pp. 105-106, 2003.

[2] Freeman, L. C., Centrality in social networks: Conceptual clarification. *Social Networks*, Vol 1, pp. 215-239, 1979.

[3] Võhandu, L. Some Methods to Order Objects and Variables in Data Systems. *Trans. of Tallinn Technical University*, No. 482, pp. 43-50, 1980.

PAPER II

E. Täks, A. Lohk, An alternative method for computerized legal text restructuring. In Proceedings of the 2010 conference on Legal Knowledge and Information Systems: JURIX 2010: The Twenty-Third Annual Conference, Radboud G. F. Winkels (Ed.). IOS Press, Amsterdam, 2010, pp. 171-174

An alternative method for computerized legal text restructuring

Ermo Täks, Ahti Lohk
Department of Informatics,
Faculty of Information Technology,
Tallinn University of Technology

Abstract. Legal quality is a manifold issue; the presented research is focusing on the structural and referential aspects. Focusing on verbs and nouns, the text is turned into graphs, opening a compressed visual view of the legal act and offers several useful features:

- information-enriched graphical representation of the law
- search tool
- mind map to open visual insights
- input to ontology
- comparison basis between legal acts
- basis for restructuring the legislation

Keywords. Legislation structure, text analysis, legal norms, graph, visualization

Introduction

A legal system is supposed to reduce the level of uncertainty within society, but inevitably has its weaknesses to be dealt with. The discussion in this paper addresses the structure and referential indeterminacy of the legislation.

Many research papers offer a graph-like presentation of the law [1], [2], [3], [4], [5], [6]. Still, it is not possible, for example, to visualize the Estonian legislation as a full graph due to inadequately determined legal references. There therefore emerged a need for an additional, fully automated layer to interconnect different yet similar parts of the legislative documentation.

This work defines the substantial structural units of the legislation similarly to Biagoli, who uses norms as a natural part of the legal system [7]. The legislative system is a collection of legal texts, where the language and norms are tightly bound. Legal texts are the written formulation of norms, a thought content expressed through language [8] and norms can be therefore found or targeted with the help of linguistic or grammatical constructions [9]. It is argued that similar norms should be constructed with the help of the same words [10].

A short analysis about the use of different types of words in Estonian legislation identified the appearance of nouns on 47% and verbs on 13% of the cases. The informational value of these word groups is much higher compared to others, such as numerals or conjunctions. A clause is a minimal grammatical construction able to deliver the thought content of the norm, containing at least a subject (noun) and a

predicate (verb). The clause is also acting as a natural language container, binding together specific nouns and verbs. Such interpretation allows us to form a visual backbone of the specific norm, containing by rough estimations 60…80% of the norm content (Figure 1).

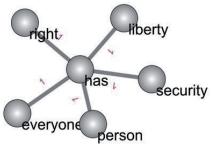


Figure 1. Nouns and a verb sub graph of the norm "Everyone has the right to liberty and security of person."

1. Example of use

A specific part of Estonian legislation was chosen for the analysis in order to test the current approach: European Parliament Election Act, Local Government Election Act, Referendum Act, and Riigikogu Election Act. Altogether, it covers a text with 2739 sentences and 32 187 words.

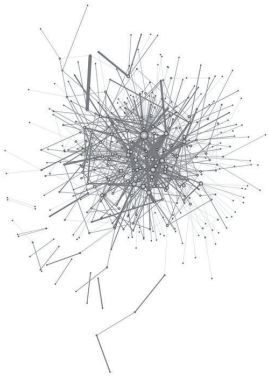


Figure 2. Complete graph of connected keywords

The consolidated text was then analyzed based on language structure. The borders of clauses were defined and the concordance of the verbs and nouns within the clauses was calculated. As a result, a table was created, showing the frequency with which certain verbs appeared with certain nouns in the same clause. In total, 396 unique nouns and 197 unique verbs were found. Compared to the original document the amount of words was reduced more than 54 times.

This table was turned into an undirected graph (Figure 2). The result is a dense and almost fully connected network of keywords, a "skeleton" of the legal acts. Graph edges represent the words and the size of the edges shows the degree of it - to how many different words it connects. Arcs show the connections between the words and the line thickness shows how many times the pair of words appears together in one clause.

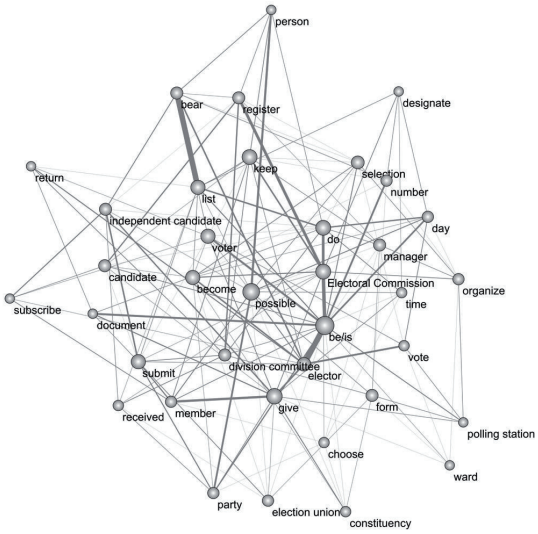


Figure 3. Graph based on verb- and noun-concordance in the clause

In order to reduce the complexity, the graph was reduced for more substantial analysis (Figure 3). This view helps to obtain a quick overview of the content of the legal act, showing the most important objects and activities.

2. Discussion

2.1. Information-enriched graphical representation of the law

The graph is representing a content-rich environment, opening new ways of treating legal documents. Built up on the existing texts, the graph can be directly related to it and does not conflict with any part of it. A visual layout of the graph is formed from the keywords, which are considered to be important within the text. The cost of such condensation is a certain loss of information, but the generality and visibility of it is greatly enhanced and one could use this graph together with legal text.

2.2. A search tool

The data and graphical representation of important nouns/objects and verbs/activities in the legal act supports a specific search within the legal text. Such document-specific listings support the text search, but can also be used for broader searching tasks.

2.3. A mind map: giving valuable insight into the most important issues

The networked presentation of the matter, especially in a reduced graph, forms a good visual picture of the main regulated issues under the regulation. This way the graph layout is a kind of a roadmap or a mind map for human users, guiding them quickly through the legislation, giving them a general understanding of it and a group of keywords reflecting the main issues.

2.4. Entry to ontology

Presented graph layout is independent from ontology, but can be used as an entry to ontology, offering a set of statistically important keywords and their measured relationships. This approach is based in combination of the word use frequency and its influence range; the main advantage is quick, objective and fully automated process.

2.5. Comparison basis

The graph layout is suitable for the comparison of legal documents or its parts. The keywords most frequently used and the relations between them form a sufficient basis for quick and rough comparison of different legal documents.

2.6. A basis for structuring legislation

The ability to compare graphs is a key to the restructuring of the legal domain, taking into consideration the graph-based similarities of documents and the existing hierarchies in the legal domain. This opens the way to the restructuring of the legislation and visual presentation of its structural relations based on the probability of the similarity.

References

[1] I. Liiv, A. Vedeshin, E. Täks, *Visualization and structure analysis of legislative acts: a case study on the law of obligations*, Proceedings of the 11th International Conference on Artificial Intelligence and Law, ACM, Stanford, California, 2007.

[2] D. Bourcier, P. Mazzega, *Codification, Law Article and Graphs,* IOS Press Amsterdam, Leiden, Netherlands, 2007.

[3] M. J. Bommarito II, D. M. Katz, *Mathematical Approach to the Study of the United States Code,* Physica A, Vol. 389, March 25, 2010.

[4] M. J. Bommarito II, D. M. Katz, Properties of the United States Code Citation Network, *Working Paper Series,* SSRN eLibrary, 2009.

[5] F. B. Cross, T. A. Smith, *The Reagan Revolution in the Network of Law*, SSRN, 2006.

[6] P. Zhang, L. Kappaka, *Semantics-Based Legal Citation Network*, Proceedings of the 11th International Conference on Artificial Intelligence and Law, ACM, Stanford, California, 2007.

[7] C. Biagioli, *Towards a legal rules functional micro-ontology*, In Proc. of workshop LEGONT '97, 1997.

[8] A. Aarnio, The *rational as reasonable: a treatise on legal justification,* Dordrecht, Reidel, 1987.

[9] E. de Maat, R. Winkels, *Automated classification of norms in sources of law*, In Semantic Processing of Legal Texts, Springer, 2010.

[10] A. Wagner, S. Cacciaguidi-Fahy, *Obscurity and clarity in the law: prospects and challenges*, Ashgate Publishing Ltd, 2008.

PAPER III

K. Nyman-Metcalf, E.Täks, Simplifying the law – can ICT help us? International Journal of Law and Information Technology 2013; doi: 10.1093/ijlit/eat003

Simplifying the law- can ICT help us?

Katrin Merike Nyman-Metcalf, Ermo Täks
Tallinn University of Technology
katrin.nyman-metcalf@ttu.ee, ermo.taks@ttu.ee
Ehitajate tee 5, 19086 Tallinn

1. Introduction

Usability is a well-known term in information science. According to the International Usability Standard, usability is related to: "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments".[1] In other words: information is useful if it is delivered to the right person in the right location at the right time in the right amount and the right form (and so on). The same principle can add value also in other areas of our social life meaning in our existence as members of a society, in legislation for example.

There is no denying that legislation plays a big role in any society. From a sociological viewpoint it can be said to be a tool of government to organize society and protect citizens. Versions of the kind of legal systems that exist today have been used for thousands of years and have thus in essence proven their value, but inevitably the systems have organic weaknesses that have to be considered and overcome. It is interesting to see if modern technology in the form of information communication technology (ICT) can assist in counteracting such weaknesses. Legislation is usually a huge collection of different normative documents, obligatory to know and follow for everyone yet effectively handled by a few. This huge collection furthermore tends to keep growing and changing with time. It is no exaggeration to say that finding a relevant norm may be a challenging task even for experts and all the more so for laymen. However, the information age is offering new ways of organizing and presenting the legal content and there are various on-going research activities to study such possibilities. Despite this, most legal work is still performed in a very traditional manner.[2]

[1] ISO 9241-11:1998 *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability* (International Organisation on Standardisation) 1998

[2] T. Agnoloni, L. Bacci, E. Francesconi, P. Spinosa, D. Tiscornia, S. Montemagni and others, *Building an ontological support for multilingual legislative drafting*. *Proceedings of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference* (IOS Press, Amsterdam 2007) 9-18

The various possibilities that ICT offers appear not to be fully appreciated and used for legislative or law implementing purposes.

Key elements of the normative system, such as its construction and through that its accessibility are constantly under observation by those involved in legislative or law implementing activities. The law must be understood in order to be properly implemented. The law must also be sufficiently understood by the general public. This is essential in order to be able to assess if the society is one governed by the rule of law, with legal certainty and equal opportunities under the law. The constant growth of both the national and the international legal corpus and the rapidity of changes to it, mean that the legal system gradually and constantly becomes more complex to manage. This complexity affects legal hierarchies as well as the substance of laws. The unwanted consequence and by-effect of the complexity is its negative influence on the intelligibility of laws for citizens.[3] Not knowing the law is not an excuse for not following it, so navigating at least to a certain extent in the ever more complex legal system is a necessity not just for specialists. Any help to do this is worthwhile and modern technology should be able to give such help to an even greater extent than what is already the case.

A well accepted and widely recognized ICT solution in the legal field is representing legislation and related documentation in the web, making the legislation physically easily accessible for everyone and thus solving many distribution and production related problems. Still this can be considered as a relatively small advancement compared with traditional presentation of laws. Usually the legal documentation is "reflected" on the web, thus following the logic and build-up of the traditional "paper based" legal system. Such an approach greatly underestimates the possibilities and services ICT systems are capable of offering. To understand the different possibilities that exist, to present the legal information better with help of computers and choosing the best option from these many possibilities, is a big and extremely challenging task. In order to succeed in the field of systematisation and simplification of the law, a combination of well-established traditional methods and promising new ones has to be explored.

[3] A. Boer, R. Hoekstra and R. Winkels, *METALex: Legislation in XML. Legal Knowledge and Information Systems. JURIX 2002: The Fifteenth Annual Conference* (IOS Press, Amsterdam 2002) 1-10 and D. Bourcier and P. Mazzega, *Codification, Law Article and Graphs. Proceeding of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference (*IOS Press, Amsterdam 2007)

This article offers a brief overview of the research that has been undertaken recently specifically regarding use of ICT systems for legislative simplification and as a key element, the article presents some novel suggestions developed by the authors. This article primarily focuses on the research carried on within the European Union (EU) to support the usage of EU legislation at different levels. The first part of the article analyses the subject of systematisation and simplification of the law from the EU legislator´s point of view, while the second part gives an overview over research activities carried out in this field. The idea behind the overview of the EU initiatives is to see how well these can be supported by ICT, which through matching identified problems with proposed solutions also helps to determine how valuable in practice the various ICT solutions may be. The idea behind the overview of research activities is to verify whether on-going academic research is adequate and addresses the same issues as the EU legislator has identified. The scope of the article does not permit a detailed evaluation of all such research activities, but the aim of the information is to provide a descriptive framework for determining what possibilities ICT provides and how the suggestions of the authors fit in with other on-going research. The illustration in a table format of the correlation between issues and research undertaken shows the links very clearly. Finally, a new system is presented to assist in the endeavours to use ICT for simplification of legislation and its overview.

2. A need for simplification

Practical effects resulting from the complexity of legislation and the difficulties in getting an overview over the normative system are apparent, not least in the EU with its large mass of normative instruments applicable in 27 (soon 28) different countries. When the EU is criticised for being distant from the citizens, the complex legal system is often mentioned as one aspect. There are several initiatives of the European Commission that aim at improving quality in European (and consequently also national) legislative texts. In particular, harmonizing legal terminology is considered a precondition for improving the quality of legislation and for facilitating access to legislation by legal experts as well as citizens. In a multilingual environment, only genuine awareness of the subtleties of legal terminology in the different languages enables drafters to maintain coherence among the different linguistic versions of the same text. This is essential for the EU Member State legal orders, especially in relation to regulations that are directly applicable in all Member States. It is essential to know that the meaning of the legal text is the same in all Member States. Coherent legal terminology is however also important in order to be able to implement EU

directives properly into national law. Over the years, the European Court of Justice (ECJ) has had the occasion to point out the importance of terminological correctness and the equal validity of language versions in a number of cases, in different contexts (Case 29/69 *Stauder*, Case 283/81 *CILFIT* and Case C-257/00 *Nancy Givane* for example).

Well planned, easily searchable and user-friendly presentation of legal documents contributes greatly the accessibility and understandability of the legislation and through this may improve the quality of implementation. This is understood well by EU officials, declaring that "Developing a user-friendly and easily accessible European law is a major concern for the Commission and the other EU institutions."[4]

Against this background, the EU has at the highest level undertaken a lot of efforts in legislation simplification. Officials have found that improving the quality of legislation is in fact a public good in itself, enhancing the credibility of the governance process and contributing, it is said, to the welfare of citizens, business and other stakeholders.[5] Better regulation is a driver to improvement of the policy making process through the integrated use of effective tools - not an attempt to impose further bureaucratic burdens. High quality regulation instead prevents unnecessary burdens on businesses, citizens and public administration. If the regulatory system and legislation are clear and effective it helps to avoid damage to competitiveness caused by increased costs and market distortions. Indeed, studies from various sources have estimated the burden of regulation to fall in the range 2-5% of GDP in Europe.[6] Although such a burden can only be estimated, nonetheless the figure does indicate the importance of this issue to European economies. Another aspect, also hard to measure exactly but still very essential, is that high quality regulation assists in the restoration of confidence in government and consequently is better able to accomplish its desired purpose as it is more likely to be followed. Implementation of such regulation is less problematic for public

[4] European Commission, A*ccessibility and Presentation of EU Law,* available at http://ec.europa.eu/governance/better_regulation/access_eu_law_en.htm (accessed 1 September 2012)

[5] Dieudonné Mandelkern, Ernesto Abati Garcia-Manso, Kees Burger, Marc Cabane, Luigi Carbone, Bruno Chavanat, Colin Church, Dermot Curran, Erwin De Pue, Michael Fruhmann, Pierre Gehlen, Roberto Hayder, Lars Hjortnæs, Panagiotis Karkatsoulis,Erich Milleker, *Mandelkern Group on Better Regulation, Final report* (European Commission 2001) i.

[6] Kai Wegrich, *The Administrative Burden Reduction Policy Boom in Europe: Comparing mechanisms of policy diffusion* (Centre for Analysis of Risk and Regulation, London School of Economics and Political Science, London 2009), e.g. 2 and 8.

administrations and compliance is easier for citizens.[7] Thus the benefit of simplification are easy to see in any legal system and especially against the background of the EU legal system.

3. Legislation simplification approaches in the European Union

The cited Mandelkern report proposes an Action Plan with deadlines, suggesting that the implementation of the Plan would contribute significantly to achieving the required improvements of the EU regulatory process. It describes a comprehensive overall approach with a set of seven core principles: necessity, proportionality, subsidiarity, transparency, accountability, accessibility and simplicity.[8] The suggestions of the plan have been implemented to some degree through various initiatives (discussed below), although work still remains and the core content remains relevant, even after more than ten years.

The core content of the Action Plan in the Mandelkern report can be summarized as follows, using the bullet points of the report.[9] The issues described in each of these bullet points is subject to analysis later on in this article, primarily in light of how ICT solutions can support the suggestions made:

Policy implementation options: This, in the words of the report, means that EU and national policymakers should consider the full range of possible policy options and choose among them that which is most appropriate in any specific case. The practical meaning of it is that although adoption of a legal act may frequently be the most appropriate option it should not always and automatically be the only choice.

Impact assessment (IA): The Mandelkern report states that regulatory impact assessment (RIA) is an effective tool for modern, evidence-based policy making, providing a structured framework for handling policy problems and as such should be an integral part of the policy making process at EU and national levels and not a bureaucratic add-on. The importance is described as ensuring that decisions are taken based on sufficient knowledge.

Consultation: Early and effective consultation of interested parties is described as means of and an important tool for open governance. This should not take over the role of civil servants, Ministers or

[7] Mandelkern and others (fn 5) 7.
[8] *Ibid.* 9-10
[9] *Ibid.* i-ii

Parliamentarians in the policymaking process (such as it is under national or EU law) but aims to supplement the information these experts have. Even if consultations take some time and effort, if they are correctly done they can help avoid delays in policy development due to controversies that may arise later. Consultations can thus speed up and facilitate rather than delay or hinder progress.

Simplification: In the Mandelkern report simplification is highlighted in a special bullet point. Simplification does not mean deregulation. In this context it is aimed at systematic and targeted efforts that while preserving rules, makes them more effective, less burdensome, and easier to understand and to comply with. The report states that programmes need to be established at both EU and national levels.

Access to regulation: The report points out that those affected by European or national regulation have the right to be able to access it and understand it. Access here has two sides. One is that the coherence and clarity of regulations must be enhanced through consolidation (including codification and recasting) and the other that access should be improved by better practical arrangements (especially using ICT). The suggestions of the report are that the coherence and clarity should be achieved through consolidation programmes at national and EU level, while the practical access is best served by having within Member States as well as at EU level a public access service (either free or for a small fee). Since the report was written, it is more common that access if free.

Structures: Under this bullet point, the Mandelkern report reaches the conclusion that better regulation needs the appropriate supporting structures for its promotion if it is to be successful. The best arrangement at EU or national level will depend on various relevant circumstances. To give this task to a single unit at or near the centre could be considered, but the essential is that an effective solution must be found at each level – central structures alone may not be sufficient.

Implementation of European regulation: The final bullet point in the summary of the Mandelkern report states that high quality regulation forms a chain from the earliest stages of its preparation through to its implementation. The consequences drawn from this is that more attention should be paid at the European level to implementation concerns in order to ensure that the full consequences are understood and taken into consideration. Member States should also give higher priority to the implementation of EU legal acts.

3.1 EU initiative "Better regulation"

Based on the Mandelkern report a series of initiatives have been launched in the EU aimed at making EU legislation "better". The many initiatives include the following Commission instruments: "European governance - A white paper"[10]; "Interinstitutional Agreement on Better Law-Making"[11]; "Communication from the Commission - European Governance: Better lawmaking"[12]; "Communication from the Commission - Action Plan: Simplifying and improving the regulatory environment"[13]; "Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - Updating and simplifying the Community acquis"[14]; "Communication from the Commission to the Council and the European Parliament - Better Regulation for Growth and Jobs in the European Union"[15]; "A strategy for the simplification of the regulatory environment"[16] and "Smart Regulation in the European Union"[17].

To some extent the Commission has in these many proposals accepted the proposals made in the Mandelkern report. Some suggestions were however modified or not included.

In this paper we give a short overview over the adopted initiatives, based on the various instruments listed. The list includes a combination of methods and tools adopted and statements of fact, regarding elements that affect the process and that need attention if real improvements are to be made. The steps are listed on the European Commission web-site under the general headline of Better Regulation, with some links to the proposals or normative documents. This outline of what steps to take toward better regulation is based on the mentioned Commission overview.[18] Later in the paper, the potential of ICT to contribute to the goals set out for better regulation will be examined and presented.

3.1.1 Impact assessment

[10] COM(2001) 428 final
[11] 2003/C 321/01
[12] COM(2002) 275 final
[13] COM(2002) 278 final
[14] COM(2003) 71 final
[15] COM(2005)97 final
[16] COM(2005)535 final
[17] COM(2010)543 final
[18] See fn 4

As suggested in the Mandelkern report and brought up in many of the EU documents since (and highlighted at the European Commission web-site), impact assessment is designed to help in structuring and developing policies. If it is correctly carried out it helps to identify the main options for achieving the objectives and analyses their likely impacts in the economic, environmental and social fields. Through such a system the Commission aims to improve the quality and transparency of its proposals as well as identify balanced solutions to reach policy objectives through:

A. "a coherent analysis of potential impacts,
B. consideration of various policy choices (e.g. to use alternative instruments to 'control and command' regulation or non-intervention),
C. consultation of stakeholders; and
D. enhanced transparency (IA roadmaps and IA reports published on the Impact Assessment website).
E. Executive summaries of impact assessments are translated into all EU languages."[19]

3.1.2 Consultation

The next mentioned element of better regulation consists of consultations, meaning that the Commission consults and is in constant touch with external parties when elaborating its policies and before making proposals and taking policy initiatives. Among other questions, the Commission must consider whether EU legislation is needed in the specific situation. Consultation can be made in different ways, using various methods. Listed methods include consultation papers (Green and White Papers), communications, advisory committees, expert groups, workshops and forums. In practice, any combination of such methods or more informal ones can fulfil the goal of consulting properly. Examples of consultations include public consultations on regulatory issues for a possible future EU – US trade agreement or on the future for EU and US trade and economic relations[20] or on the implementation of a data and transaction reporting framework for wholesale energy markets[21]. All policy home pages of the EU include a heading for consultations where open and recent public consultations can be seen and there is a possibility to participate. The exact manner and content of a consultation will be decided based on what is suitable for the issue

3.1.3 Expertise

[19] *Ibid.*
[20] http://trade.ec.europa.eu/consultations/index.cfm (accessed 1 December 2012)
[21] http://ec.europa.eu/energy/index_en.htm#t_0_2 (accessed 1 December 2012)

Just like consultations, expertise can take various forms. If consultation is needed specifically in order to get external input, the expertise is to a high degree available in-house but the Commission also (and increasingly, given increased technical or other complexity of many policy areas) calls upon external specialists in different fields. There may be expert groups or workshops, hearings, conferences or seminars to get access to the expertise.

3.1.4 Administrative costs

To have a sub-heading called "Administrative costs" under the main heading of better regulation may not be immediately obvious, but the idea is to show awareness of the fact that implementing regulations and laws entails costs and there is consequently a need to take this into consideration in regulatory reform work. To deal with this issue has become an important driver for reform, as some legal obligations like on providing information or access to information have become expensive as well as time-consuming, complex and perhaps even useless. The Commission finds that there can be saving of time and money if certain reporting requirements are terminated. To support this, an EU Standard Cost Model has been developed as part of the European Commission Impact Assessment Guidelines[22] () to help making it easier to make comparisons cross-country or cross-policy.

3.1.5 Choice of regulatory instruments

Public authorities intervene in the markets and lives of the subjects through setting standards, by levying taxes and charges, by financing specific actions and groups, or by providing information and advice. The most direct means of intervention is regulation through law and other regulatory instruments. For the EU, in addition to the same kind of considerations that states take before deciding on actions, the question of division of competence between the EU and its Member States arises. The instruments of regulation at the disposal of the EU are clearly set out in the Lisbon Treaty (Article 288 Treaty on the Functioning of the European Union, TFEU) that provides the current legal basis for the EU. Forms of legislation in the EU are:

A. Regulations and directives
    a. A large proportion of EU law is found in directives which set out a result to be achieved but leave national authorities the choice of methods: "A directive shall be binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and

[22] ec.europa.eu/governance/impact/commission_guidelines/commission_guidelines_en.htm

methods.“ (Article 288 TFEU). Directives may go beyond what is required by EU legislation, adding obligations and procedures, which is called 'gold-plating'. This can affect the implementation of the legal requirement as well as the coherence of the entire regulatory system. The effect of gold-plating and how to deal with it is subject to current debate. Another aspect of the debate is whether directives are becoming too detailed and in fact are not in accordance with the original idea of this type of legislation.

b. The other main legal instrument for the EU is the regulation. Regulations are directly applicable in all Member States and require no transposition. Consequently, there is no problem of gold-plating. On the Commission Better Regulation web-site it is thus suggested that replacing directives with regulations may under certain circumstances be conducive to simplification. However, although some provisions of the EU Treaties permit a choice between different types of legal instruments, in many instances the choice has been made in the Treaties and cannot be changed without a Treaty change, which is a slow and cumbersome process.

B. Review/sunset clauses: Another points raised on the Better Regulation web-site is that of obsolete legal acts that no longer have real effect, but which remain in force because they have not been expressly repealed. To deal with obsolescence of legislation, specifically in rapidly developing areas such as high-technology, the Commission often introduces review, revision or sunset clauses in its legislative proposals. Yet, in some areas alternative mechanisms to legislation, such as self- and co-regulation, may be more suitable to handle rapid developments.

C. Alternative instruments: To avoid obsolescence but also for other reasons, alternatives to traditional regulation should be considered, like self- and co-regulation.

a. Self-regulation by private parties, usually members of a profession or producers in a sector or between operators and their clients/consumers. This can include codes of conduct in a specific sector, with obligations to submit to tests by approved laboratories and labels to this fact. Media also self-regulates to a large extent as do liberal professions (lawyers etc.).

b. Co-regulation is used for occasions when authorities give recognized private parties the duty to ensure that certain objectives are reached. These may be economic operators, the social partners, non-governmental organizations or associations.

c. Self- and co-regulation require monitoring and evaluation. This is essential to their credibility, the Commission states. For this purpose, to share information but also encourage the setting up of self- and co-regulatory systems, the Commission and the European Economic and Social Committee (EESC) have established an internet EU self- and co-regulation database.[23]

3.1.6 Transposition and application of EU law

Continuing down in the Commission´s list of measure for better regulation[24] the timely and correct implementation of EU law by the Member States is the next issue listed. This ensures that the results intended by EU policy are attained and conversely late or incorrect implementation can deprive EU subjects of their rights. The transposition into national law is done by national governments and parliaments, sometimes involving regional and local authorities. At this stage, laws risk being 'gold plated' i.e. requirements or procedures which are not required by the initial directive are added. This may be unproblematic, provided the original intent of the directive is not lost. However, often the effect may be that added matters distort the intent of the directive. Member States may even have legislation that similarly affects the substance of a regulation (that itself may not be transposed into national law, as it is directly applicable). If the way in which EU law becomes part of national law is thus distorted, this affects both transposition and implementation of EU law as well as the quality of national and regional regulation.

Concerning directives, the monitoring of the transposition process relies on the correlation tables provided by the Member States. These tables show the link between the provisions in directives and national rules – either new legal acts adopted to implement the directive or existing legislation that already contains the relevant objective. This highlights the importance of such correlation tables and that the content in them really shows actual correlation and not just surface correlation. The complexities of EU legal terminology and making it fit with traditional national legal terminology highlight potential problems. This is a concrete example of where the use of systematisation with the help of ICT can have a great practical impact.

3.1.7 Simplification

[23] *Self- and Co-Regulation* from the European Economic and Social Committee Website: http://www.eesc.europa.eu/?i=portal.en.self-and-co-regulation (accessed 10 August 2012)
[24] Fn 4

On this matter, the Commission says that "It is essential in a rapidly changing world to review laws, streamline and remove overlaps to ensure that EU legislation is clear and poses as few burdens as possible for operators and citizens. The EU has progressively developed a broad strategy to improve the regulatory environment and thus provide a more effective, efficient and transparent regulatory system for the benefit of citizens and reinforce competitiveness, growth and sustainable development [--]."[25]In its strategy to simplify the regulatory environment[26], the Commission uses the following methods:

1. repeal: removes from the statute-book those legal acts which are unnecessary, irrelevant or obsolete;
2. codification: contributes to the reduction in volume of EU legislation, and at the same time, provides more readable and legally secure texts, thus facilitating transparency and enforcement;
3. recasting: is a simplification method as it simultaneously amends and codifies the legal acts in question;
4. co-regulation: can be a more cost efficient and flexible method for addressing certain policy objectives than classical legislative tools. Standardization by independent bodies is an example of a well-recognized 'co-regulation' instrument;
5. use of regulations: replacing directives with regulations can under certain circumstances be conducive to simplification as regulations are directly applicable (i.e. no need for transposition into national legislation) and guarantee that all actors are subject to the same rules at the same time. This presupposes, as mentioned above, that EU law allows such a choice of legal instrument.[27]

For such simplification a system that provides a simple, graphic overview of legislation can be very useful. It will be a starting point from which the most appropriate measures can be decided. Here is thus another area where ICT – to help make the graphic overview – is of immediate use.

[25] ec.europa.eu/governance/better_regulation/simplification_en.htm
[26] Commission, Communication of the Commission to the European Parliament, the Council - *Implementing the Community Lisbon programme: A strategy for the simplification of the regulatory environment,* COM(2005)535 final
[27] *Ibid.*

3.1.8 Accessibility and presentation of EU law

The EU makes a lot of use of ICT to give access to EU law. This is mainly in the form of reflecting legislation, adopted and structured for a traditional, non-ICT environment. The EU operates two websites for free access to EU law, directed at the general public as well as specialists:

1. EUR-Lex website: free access to the full range of EU law and treaties, including consolidated legislation, international agreements, parliamentary questions, case law, new legislative proposals and much of the EU Official Journal in all EU languages. EUR-Lex also contains a register of documents of the EU institutions.
2. Pre-Lex: a possibility to follow the major stages of the decision-making process between the Commission and the other EU institutions starting from the Commission proposals. Commission communications are also accessible through various search possibilities.

These web-sites are easily accessible for specialists, while their design may be somewhat complicated for infrequent users from the general public. This may have more to do with the amount of information and the rather complex legislative procedure of the EU than with any serious weakness in the design of the web-pages. The EU presents a lot of information about its activities as well as for example summaries of legislation and collections of main instruments under clear subject-matter headings that are more easily accessible for the general public than the legal databases as such. The need to think about how accessible internet information is in practice, for the intended target groups, is an issue given increased prominence as the web-based information amount grows.[28]

3.1.9 Evaluation

Evaluation can be made ex-ante to determine prospective results of regulation or ex-post to retrospectively see the real results and impact. There is often a combination, to better determine results and impacts in relation to the stated needs and objectives compared with resources used. The outcome of evaluation should be used for planning, designing and implementing future EU policies.

3.1.10 Inter-institutional coordination

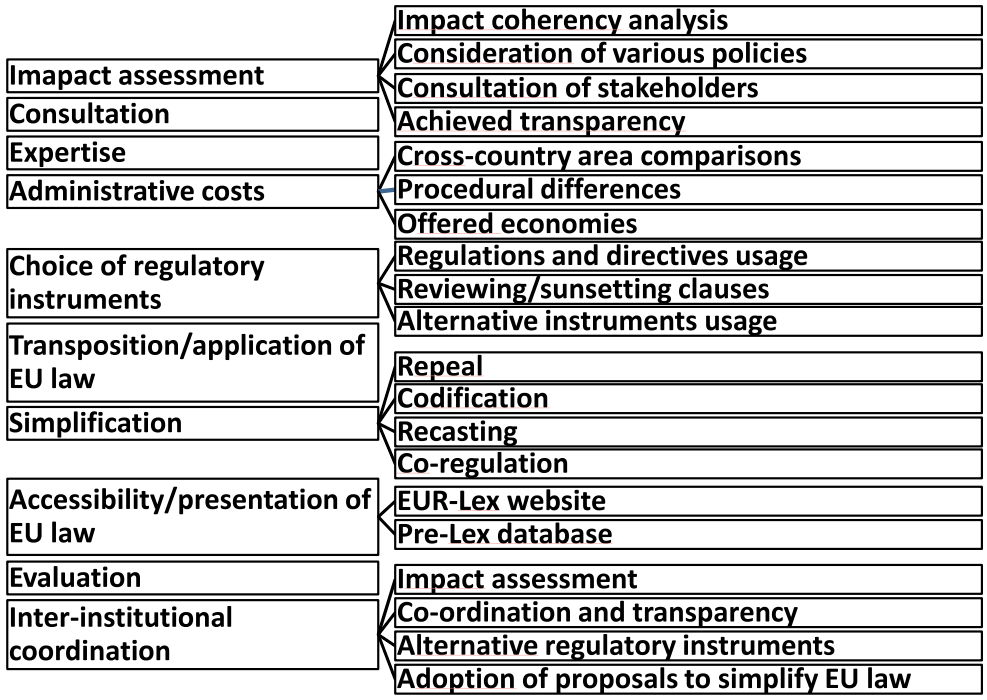
Given the distribution of tasks and competences in the EU, it is clear that better regulation is a shared responsibility (between the EU and the Member States as well as between different EU

[28] See for example http://www.accessforall.eu/2011/10/level-of-accessibility-in-public-websites-in-comparison-with-legislative-provisions-and-the-deadline-for-implementation-of-wcag-2-0 (accessed 1 December 2012)

institutions), so although the Commission lists the different initiatives and issues on its web-page, the real impact will only be felt if there is proper coordination. As is known, the process very briefly described is that the Commission submits proposals for adoption to the European Parliament and the Council. EU laws are transposed into national law by national governments and parliaments (depending on type of legal act) and often applied at regional and local levels. The responsibility for regulating well is hence a shared one. The Inter-institutional Agreement on Better Law-Making[29] concentrates on:

A. improving inter-institutional coordination and transparency,
B. providing a framework for alternative regulatory instruments (self- and co-regulation),
C. increasing the use of impact assessment in Community decision-making, and
D. working methods for the adoption of proposals to simplify EU law.

Below the EU better regulation components are graphically presented.



[29] 2003/C 321/01

Figure 1. Different EU activities within the initiative "Better regulation"

4. How can ICT help?

The past decade has shown an impressive increase of legal informatics-related works in science as well as in the public sector. Many problems can at the present time be solved more efficiently with help of computers and many promising research fields indicate the potential to offer even more solutions in the near future.[30] The question is not any more whether ICT can be usefully exploited in the legal field but more where and how ICT can support or even replace existing approaches.

In this part of the article we are going to introduce shortly some of the most promising research aimed to support legal specialists in their work for better regulation. One purpose of such an approach is to tie up both ends of the same issue, traditional and well proven legislative activities versus novel and innovative ICT centric aspects. Due to the complexity of the issue, these two aspects tend to drift apart so that it is hard for outsiders to keep track of the main issues. There may be too much emphasis on the technological side of various solutions, whereas ideally the solutions should primarily relate to the issues to be solved in a comprehensive manner: including ICT solutions and any other means to achieve better regulation. The area selected to be described in somewhat more detail is that which constitutes the main research topic of the current authors, with other research alluded to briefly in order to keep the article in reasonable length.

Many efforts have been made in the last few years to investigate two problems in the analysis of legal texts: how to automatically identify structural portions of legal documents through their mutual references and how to grasp semantic information of the legal text. Legislation usability for human beings as well as for computers is under consideration as these two elements of usability are quite different. Human accessibility can be increased with help of additional structural analysis and a visualisation technique, whereas computer usability can be enhanced through new standards and mark-up technologies

Most of these approaches are based on XML and RDF, so it might be useful to first give some background information about these acronyms before describing the visualisation system proposed by the authors.

4.1 XML, RDF and OWL

[30] Such solutions are presented below in this chapter, including that proposed by the authors of this article.

During the last couple of thousand years the usual, paper based document has evolved to the most convenient form considering human readers. Now, the digital revolution is introducing new challenges and one of them is to teach computers to understand open text. XML stands for eXtensible Markup Language. XML is designed to transport and store data and has a limited ability to add some meaning to text. It is done through text structuring and adding markings (tags) to these structural units. This forms a basis for a next generation World Wide Web- semantic web or WEB 2.0.

For example, a heading is easily recognizable for human readers within a text. It is usually separated from the rest of the text and is visually different. However, it is very complicated to create a good rule with the help of which computers can independently recognize headings in the same ways as humans do. Even if it were possible to create such a function, it is impossible for computers to perform a reverse operation: decide which parts of text are suitable to use for headings. So the best thing we can do is to divide text manually into smaller but meaningful parts - like heading, signature, date, addressee, etc.- and surround them with tags: *<heading> Chapter 1</heading>*.

RDF is used to define the structure of the data and is used to give additional semantic meaning to the structured text. If XML is a method to separate and mark textual parts, RDF provides a way to link these parts together in a meaningful way. The way how it is done is estimated like this: *Subject - Predicate – Object*. This is called triple and allows us to describe needed relationships between two entities: *Chapter 1 is written by John More*. For more complex relationships, different ontology related methods are used like a Web Ontology Language (OWL) etc. This kind of manually added information enriching information is called metadata.

The abovementioned method increases the usability of the legal text dramatically and therefore related technologies are among the main topics of research in Europe and abroad. There is a reasonable hope that such novel technology could increase the usability of legal information a lot, not least in the EU with its multilingual structurally complex, constantly changing and sometimes internally conflicting legal information.

4.2 Visualization of the law (short name: Visualization)

A normative system is a collection of static legal resources. The hierarchical structure of the normative system is vaguely estimated - it has a well-developed referencing system, but does not form a systematic and well-structured connected graph.

The normative system is also a collection of norms. Legal and linguistic aspects of a norm are tightly linked: the norm can be understood as "thought (i.e. meaning) content expressed through language".[31] The norms can be found or targeted with the help of linguistic or grammatical constructions. The clause is a minimal grammatical construction able to deliver the thought content of the norm, containing at least a subject (noun) and a predicate (verb). The clause is also acting as a natural language container, binding together specific nouns and verbs. Such interpretation allows forming a visual backbone of the specific norm, containing by rough estimations 60…80% of the norm content. [32]

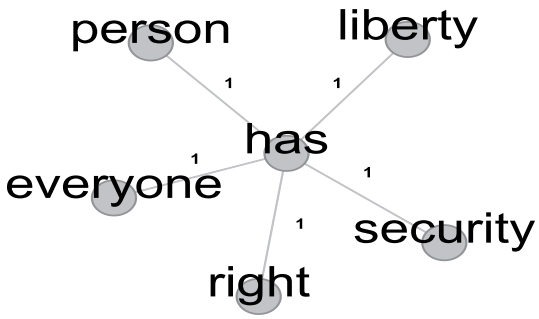


Figure 2. Sub-graph of the Estonian Constitution norm "Everyone has the right to liberty and security of person."

Such a representation has many useful features, but most importantly this provides a good formal comparison basis between legal acts. A graph-view allows us to check graph-structural similarity. All legal acts can be compared with the help of graph research methods and their similarity can be measured. This forms a basis also for possible restructuring of legislation:

[31] Aulis Aarnio, *The Rational as Reasonable. A Treatise on Legal Justification.* s.l. (Blackwell Publishing Ltd, 1988)

[32] Ermo Täks, Leo Vohandu, Ahti Lohk, and Innar Liiv, *An Experiment to Find a Deep Structure of Estonian Legislation. Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference* (IOS Press, Amsterdam 2011) 93 - 102

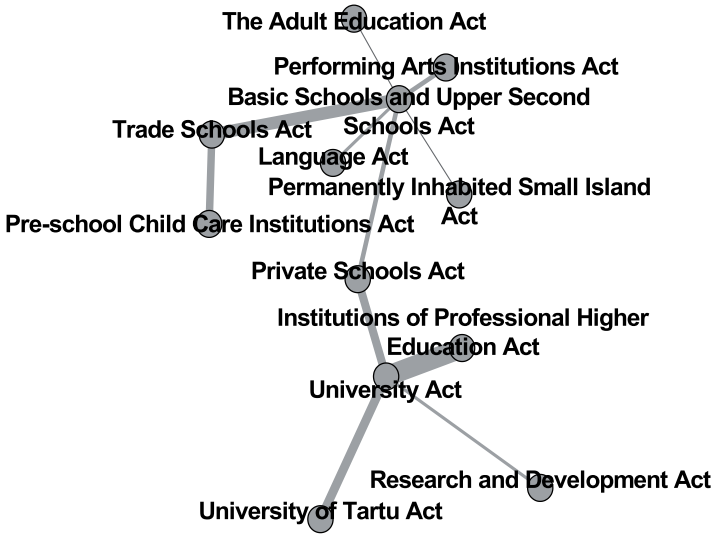


Figure 3. Extracted part example of similarity structure of the Estonian legislation, where each legal act is connected to the most similar one and the thickness of lines is referring to the similarity strength

The presented approach offers a method to create a systematic alternative structure to any naturally evolved normative system. The proposed method allows creating an easily readable and compressed visual picture (a graph) of the legal act, specifying concordance of verbs and nouns within clauses. By using these graphs to measure the similarities of different legal acts it is possible to analyse the whole normative system (the maximal connection graph). Qualitatively different new applications can be derived from presented graphs to perform a normative system analysis. A systematic visualized picture of legal documentation at different levels is capable of guiding the user through the sources of law without lengthy training. It provides a quick overview and a basis for more detailed study. Such visualization of the legal language presents many benefits and areas of use.[33]

This leads us to describe the possible positive aspects of this research little bit deeper, linked to the simplification procedures outlined above:

- **Impact assessment:** A visualized layout of the legal act presents a suitable tool for problem estimation, description, decomposition and negotiations between parties. This helps in the process of consultation with stakeholders, creating a good basis for comparison with which to measure the achieved transparency.

[33] *Ibid.*

- **Consultation:** A visualized layout of the legal act presents a suitable tool for problem spotting, description, decomposition and negotiations between parties. Visual layout is good support for creating a common understanding about the issue and the ways to analyse the influence of decisions.
- **Expertise:** Visual layout gives a bird-view of the legal domain, offering an overview of relations and locations affecting the legal question.
- **Administrative costs:** The structure of legal acts and connections of different legal acts form a good starting point for a cross-nation area comparisons and making savings in data handling processes.
- **Choice of regulatory instruments:** The location and role of the proposed legal act can be tested in visual layouts to estimate the potential effect of different legal acts and for the legislation as whole.
- **Simplification:** Different viewing and representing options based on different connections between legal acts allows broadening or narrowing the scope of comparison, to represent a collection of connected documents and parts of it, to merge or decompose parts of existing legislation and so on, to provide an overview of the legislation based on which simplification can be decided.
- **Accessibility/presentation of EU law**: Visualized and compressed layout of the legal text helps to access and present the legal content more efficiently and it can be read quicker.
- **Inter-institutional coordination:** Visualized and compressed layout of the legal text helps to get a quick overview of the content and applied changes.

4.3 Legislation related complexity and ways to measure it (short name: Complexity)

Moving on to other on-going research activities, we start with the research of Bourcier and Mazzega from 2007.[34] This is based on the assumption that there is a constant and accelerating growth of the national and international legal corpus and together with the rapidity of changes the result is that it becomes more and more complex for anyone applying law to manage this process. The complexity has an impact on the intricacy of both legal hierarchies and legal contents and may have other unwanted by-effects on the intelligibility of laws for citizens, as mentioned above.

[34] Bourcier and Mazzega, fn 3

In their work Bourcier and Massega have analysed the French environmental code to understand how the drafters have organized the previous laws scattered in various fields of law into the structure of a hierarchical table of contents. Relying on graph representation, they have observed this architecture through the various levels of its organization and connections with other legal corpuses. They were also able to find some invariant distributions that shape various statistical distributions of the vertices and edges of the legal graph.[35]

4.4 Electronic presentation of law in METALex (short name: METALex)

XML is a next generation documentation standard created mainly for better computer processing. Information on a network that connects many different types of computers has to be usable on all of them. Public information in particular cannot be restricted to one make or model or to give the control over its data format to one private body. Furthermore, public information must be possible to reuse in many different ways, which must be supported by the encoding system. This rules out proprietary data formats and this is what has led to XML use. The efficiency of managing and processing information in legal documents can be dramatically improved by applying XML techniques. As a part of the more general idea of an integrated semantic web, documents are enriched with metadata to enable smart applications such as (intelligent) retrieval and reasoning. Various national initiatives have established XML standards for describing legal sources and documents, which have grown into projects aiming at integration and interoperability across all legal domains.

A necessary precondition for effective legal document management is the electronic availability of legal sources in a structured and standard format. Boer, Hoekstra and Winkels explain how the standard intends to provide a generic and easily extensible framework for the XML encoding of the structure and contents of legal and paralegal documents.[36]

METALex is a generic open standard for legislative documents specifically designed to facilitate the maintenance of decision support software used by public bodies. In addition, it offers provisions for more or less traditional functionalities offered by publishers and search engines. The METALex XML schema aims to be a standard interchange format for legal documents for the purposes of

[35] *Ibid.*
[36] Boer, Hoekstra and Winkels fn 3 (at 2)

presentation, description of the relations between legislative documents, search and filtering on meaningful levels of detail, and version management and file exchange.

As Boer, Hoekstra and Winkels say,[37] the classification level presupposes that the user of the classification system can read the document to find out why the classification was attached. They find that such domain classification schemas, mainly predating the use of computers for storing legal information, are not necessarily adequate for electronic use. In such classification the attributes used tend to be fairly traditional like author, creation, modification and promulgation dates, jurisdiction, legal status and language. Boer, Hoekstra and Winkels find such attributes rather crude in meaning, without a lot of relevant detail and the usefulness for automated reasoning thus is questionable. Some issues they point to are that identification of documents by jurisdiction assumes that the user of a search service knows what jurisdictions he is in and the specific XML standard of that country. This standard comes from the legislative style and language of the country, although in the EU there should be a common standard. It can be explained so that the EU provides the frame that can be filled by each country.[38]

Boer, Hoekstra and Winkels as an alternative to rigid domain classification, suggest that statements can be directly identified in the contents of a document. Documents can make statements about other documents and (fragments) of the document itself: the metadata on one document is distributed over different locations. For this, there is the Resource Description Framework5 (RDF).[39] RDF is used to store a description of the events that lead to the resource with the attached metadata. In this way RDF and related technologies helps the computer to make generalisations: it can understand that if the object is a girl it is also a woman, a human being and so on. Thus a wider range of norms can be applied.

4.5 Syntactic analysis

[37] *Ibid:*
[38] *Ibid.*
[39] *Ibid.* (at 3)

Brighi, Lesmo, Mazzei, Palmirani and Radicioni stress[40] that hand-made annotations in law or other texts are time-consuming and error-prone, alternative tools for modelling and representing the structure and content of norms are needed and in the view of these authors such tools could greatly benefit from automatic approaches to extract both structural and semantic data from legal texts, conceivably generating XML output.

They concentrate on the annotation of modificatory provisions. In any legal system one norm refers to other norms, either for completing its own content or to change these other norms in some way through a modificatory provision. Even if legal language is stricter, or as Brighi and others put it, more controlled than ordinary language, tracking modifications requires considering the deep syntactic structure of sentences and encoding prior knowledge on possible modifications (content and how they practically may occur).[41] For the automatic approach before any semantic annotations are added, the text is marked up with structural data as well as with normative references and "quotation mark" elements, i.e. pieces of text referring to a passive norm (such as any additional or replacement text, along with a string indicating where this text belongs in the passive norm). Semantic annotation enriches a text with the <mod> element, which delimits a modificatory clause and with the metadata that fully qualifies the modification and its attributes according to one of the following classes identified by the authors referred to:

"Type 1. A change made to the actual text or form of the norm (an integration, replacement, deletion, relocation) or to the meaning of the norm (an interpretation or variation of meaning or a modification of clauses);

Type 2. A change made to the range of a norm (an extension of its subject matter or range of application or a provision stating a derogation to it);

Type 3. A change made to the temporal parameters of the norm (the time of its entry into force, and the time when it becomes applicable or effective);

Type 4. A change made to the status of the norm within the legal system (a decree-law that is made into law, an international treaty that is transposed into domestic law)

[40] R. Brighil, L. Lesmo, A. Mazzei, M. Palmirani, and D. P. Radicioni, *Towards Semantic Interpretation of Legal Modifications through Deep Syntactic Analysis. Proceedings of the 2008 conference on Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference* (IOS Press, Amsterdam 2008) 202-206 at 203-204

[41] *Ibid.*

Type 5. A change made to the powers conferred under a norm within the legal system (e.g. a EU directive transposed into domestic law"[42]

This system thus identifies what the content of a modification is, if it is a substantive modification of a specific provision or a modification in relation to the system of norms. To facilitate the classification of modifications, the language and its deep meaning is used as modifications use certain language.

4.5 Multilingual drafting

The aim of multilingual drafting initiatives is to improve the quality of the legislative production, to enhance accessibility of legislation at European level and to promote awareness and democratic participation of citizens to the legislative process through providing special tools for such drafting. This is essential in a legal system such as that of the EU with a large number of different languages but where still the law has to apply in the same way in all Member States.

This is the background to the project "Drafting Legislation with Ontology-based Support (DALOS)"[43], which aims at ensuring that legal drafters and decision-makers have control over the legal language at national and European level. What the project does is to provide law-makers with linguistic and knowledge management tools to be used in the legislative processes, more specifically in legislative drafting. The tools make sure the terminology is properly understood to its deep legal meaning. DALOS uses an ontological characterisation of legal language, giving conceptual meaning to the lexical units and providing connection with other terms. The combinations provided make the legal language easier to work with.

In legal language every collection of terms emanating from a specific language and a specific legal system is an autonomous vocabulary resource that can be mapped through relationships of equivalence with other systems. Words are the main tools for lawyers and the use of the correct terms as well as the correct combinations of them are of paramount importance. The best approach to map terms and term collections consists of developing parallel alignment with the same methodology and the same conceptual model.

[42] *Ibid.*

[43] G. Peruginelli and F. Bargellini, *Drafting Legislation with Ontology-based Support Project* from Drafting Legislation with Ontology-based Support: http://www.dalosproject.eu/ (accessed 10 August 2012)

DALOS shows how different methods may be applied, depending on the characteristic of the domain, the data structure and on the results sought. The project highlights how among structured data different degrees of formalization can be distinguished: controlled vocabularies (such as thesauri, classification trees, directories, keyword lists), semantic lexicons as well as foundational, core, and domain ontologies. Agnoloni and others[44] show that the integration of lexical resources (heterogeneous because belonging to different law systems, or expressed in different languages, or pertaining to different domains) can lead to different solutions depending on the desired results:

- generate a single resource covering both (merging);
- compare and define correspondences and differences (mapping);
- combining different levels of knowledge representation, basically interfacing lexical resources and ontologies.

It is shown that the methodological approach chosen in the DALOS project is the third one: it requires the definition of mapping procedures between semantic lexicons. This is driven by the reference to an ontological level where the basic entities, which populate the legal domain, are described. Such an approach has been followed to obtain correspondence between terms of different languages as well to align corresponding terms towards a common conceptualization at a higher knowledge level.[45]

4.6 Legislative Meta-Drafting (short name: Meta Drafting)

The semantic mark-up of legal texts calls, first of all, for the development of suitable sets of meta-data, supposed to capture the formal structure of the legal text, as opposed to its content, as stressed by Biagioli and Grossi.[46] Such meta-data need then to be systematically interconnected, to reveal the semantic structure underlying the mark-up.

Biagioli and Grossi start from the presumption that legal orders are perceived as accumulated sets of laws, created through a dynamic process. Legislative archives reflect the historical organisation of the legal order and the law is the documentary unit on which the archive is created. The lack of an

[44] Agnoloni and others, fn 2
[45] *Ibid.*
[46] C. Biagioli and D. Grossi, *Formal Aspects of Legislative Meta-Drafting. Proceedings of the 2008 conference on Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference* (IOS Press, Amsterdam 2008) 192-201 at 200.

analytical/systematic vision of the whole is an obstacle to obtaining information about and exercising control over the contents as a whole.[47]

For Biagioli and Grossi the meta-data needed for legislative meta-drafting can be obtained from suitable theories of provisions. Such theories according to these authors make explicit the functional structure of the legal text. They divide provisions in the model into two main families: rules and rules on rules. They call the rules on rules "a peculiar category that includes the provisions related to the dynamics of a legal system". Further, the main family of rules is divided into a further two major classes, which are the subject of normative theories - constitutive rules and regulatory rules. Biagioli has developed this approach, which has become popular. The text is not divided into paragraphs and sections but into so called provisions: small pieces of text, without a strict structure.[48]

4.6 Content management.

Boer, Winkels, van Engers and de Maat have developed a content management system. Their starting premise is that governments as well as legal publishers usually make legislation available in some special purpose XML format or XHTML annotated with metadata describing what version of which legislation it is. According to the authors, the electronic documents containing the legislation are taken from largely autonomous and heterogeneously organized repositories. Versioning metadata is closely tied to the organization of the repository it came from.[49]

To understand the relationship among multiple metadata descriptions it is first necessary to understand the resources they purport to describe. Understanding the resource in turn requires a comprehension of its lifecycle including events and corresponding transformations of the resource that constitute this lifecycle. This leads to the conclusion that an electronic document and its metadata description represent a certain time point in this lifecycle. It also means that values of metadata attributes may change over the lifetime of a legal document, even if the document itself

[47] *Ibid.*

[48] C. Biagioli, *Towards a legal rules functional microontology* Istituto di Teoria e Tecniche dell'Informazione Giuridica, www.ittig.cnr.it/Ricerca/Testi/Biagioli97Legont.pdf (accessed 10 August 2012)

[49] A. Boer, R. Winkels, T. Engers, and E. Maat, *A Content Management System based on an Event-based Model of Version Management Information in Legislation. Legal Knowledge and Information Systems. Jurix 2004: The Seventeenth Annual Conference* (IOS Press, Amsterdam 2004) 19-28 at 19

does not change. Metadata most often concerns information included in the document itself, or in another document that refers to it, so it is not added extra information.[50]

One main interest in proper content management systems is that commentaries from different sources should be available in internal applications for civil servants or others who need them in their work, but this is complicated because they are all organized in a different way and frequently updated, and sometimes even reorganized, by the publishers. Versioning, structuring, and naming practices are different between different publishers, but not because the content requires this. This makes it difficult to establish the exact identity of the legal sources contained in a product without continuous human intervention.

Among shared content elements of legislation in different jurisdictions in regard to which improved handling of metadata can be useful is that what Boer et. al. call 'timepoints'. This is a simple date, but one can distinguish three kinds of timestamps used in legislation as outlined by Boer and others[51]:

Version management timestamps define the validity of the document for reference as well as the validity for application - the document can be applied by a competent decision-maker in the time-interval in which it is active.

Legislative Drafting timestamps relate to the procedures that have to be followed by the legislator including timestamps for certain events (e.g. signing) or minimum time intervals that must elapse before some consequence follows.

Application to cases: These timestamps define objects in the outside world that the legislation refers to, which can be immediate events that take negligible time (like traffic events); persistent 'objects' (mortgages, pension arrangements); or delayed payoffs of choices (financial products exempt from certain kinds of taxation). Transitory regimes are needed to minimize any damage caused by changes in legislation and legal principles like limits on retroactive application belong in this category.

[50] *Ibid.*
[51] *Ibid.* 20-21

Another important component of the structures developed by Boer et. al. is the lifecycle of legislation, which they divide into four phases:[52]

> **Fixed:** This indicates the point in time when a design of legislation becomes an official proposal that cannot any longer be modified by the drafters in the regular way (by opening the document in an editor, changing it, and saving it). In the legislative process this will be an event linked to formal signing or confirmation, after which a new formal decision would be needed to change the text of the proposal. The authors (Boer and others) quote the example of the Netherlands, where this point is the date of signing of legislation by the monarch and where this point is furthermore characterised by auxiliary provisions like date of publication, date of enactment, official name of the legislation, official acronym of the legislation, and delegate legislative competence, becoming valid law.
>
> **Knowable:** The date of publication is also the date from which the legislation is (presumed to be) publicly known. This date is usually prescribed by law either so that the legislation sets a date (specific or a general one – how many days after some event) or so that publication is relative to some other event (like the publication of another closely related law. This is an alternative date at which any such auxiliary provisions as those mentioned above may become valid law. From the date of publication, legislation may be applied and only to events happening after this point in time (with some few exceptions for retroactive or delayed application).
>
> **Repealed:** An end-point in the lifecycle of legislation is when legislation is repealed. This date is usually announced by law as a specific date or relative to some other event or after some fixed time interval. If legislation is repealed, for document management purposes it does not 'disappear': it can still be referenced by its official name and acronym.

Not all jurisdictions distinguish between the 'Fixed' and 'Knowable' dates, and this distinction is therefore not part of METALex XML. Events have input and/or output, and if they are actions they have an actor in a certain role (e.g. legislator) and optionally instruments (e.g. a legislative competence). Boer and others divide into five types of events that can cause a transition, which leads to a new version of (a part of) a legislative text :[53]

[52] *Ibid.* 22-22
[53] *Ibid.*

Fix: Initiated by the legislator, based on legislation attributing legislative competence, having as its output fixed legislation.

Publish: Initiated by the same legislator, using a publication channel as an instrument and possibly legislation requiring the publication. Here one can see input (fixed legislation) as well as output (published legislation).

Enact: Initiated by the legislator, using as instrument legislation (or in all cases, a part if legislation) requiring enactment. Input - published legislation; output - enacted legislation.

Repeal: Initiated by the legislator, the instrument is legislation requiring cancellation of the legislation. Input is enacted legislation and output is repealed legislation.

Modify: Initiated by the legislator, the instrument is legislation requiring modification by replacing its text with text quoted in modifying (or possibly external, legislation). It is most common that parts of a legal act are modified, but it may be done also by inserting or removing an article. Input is legislation and output is amended legislation.

Boer, Winkels, Engers and Maat propose a novel use of such an event-based framework with legislation in an instrumental role. All relevant dates are attached to the event (if there is a corresponding source that contains the specific date, it is in external legislation in the instrumental role and not the legislation to which the date is usually attached as metadata). [54]

4.6 STIA

In research linked to semantic annotation in jurisprudence, the starting point is the spread of norms and laws containing explicit cross references or overlapping concerning the same or similar topics. This has entailed various actions for legislative simplification, not so much to reduce the amount of sources, but rather as a necessity in order to achieve the correct application of existing normative principles. These should be pronounced, discussed and dealt with as monolithic utterances instead of being spread across several distinct codes as stated by Pazienza, Scarpato and Stellato.[55]

[54] *Ibid.*

[55] M. T. Pazienza, N. Scarpato and A. Stellato, *STIA: Experience of Semantic Annotation in Jurisprudence Domain. Proceedings of the 2009 conference on Legal Knowledge and Information Systems: JURIX 2009: The Twenty-Second Annual Conference* (IOS Press, Amsterdam 2009) 156-161

These authors take as their example Italian legislation that is composed of more than 100 000 different acts. Such amount of legal acts is not unique in any way for Italy but in any document collections in jurisprudence the cross referencing assumes huge proportions. The biggest problem in this case is to retrieve useful information in such enormous collections in relatively short time.

What Pazienza, Scarpato and Stellato suggest as a first step on the way to legal simplification is to identify relations of pertinence between distinct laws. Such identification allows for these laws to be unified and reproduced in new synthetic codes. Information Retrieval is typically used to retrieve relevant information from document collections. Matching queries and documents is generally term-based: words within documents are used to describe the documents and to determine their relevance for a given query. This is the simple way of doing it, but this does however not show up other links than the simple correspondence between words. For a legal practitioner the real correlation between acts may not be shown in this manner.[56]

To introduce more information about the meaning of a document, semantic annotations can be added, containing additional information about the text or part of it, that are important to improve retrieval processes. In recent years, collaborative tagging systems have become very popular among users as a means for organizing their resources. These systems use semantic annotations taken by users to improve retrieval by using the information held into them.

The knowledge model of the framework extended by STIA offers two concept layers, consisting of the application layer, containing ontologies from the project called Semantic Turkey and its extensions, which are necessary to drive the application, and the user layer, containing specific domain ontologies and allowing the user to add instance data.[57]

4.7 Legal Change Management

The final area of research to be described is what can be called legal change management. Palmirani and Cervone have pointed out that not only are there many national and international XML standards for modelling and representing legal resources but furthermore for modelling norms by

[56] *Ibid.*
[57] *Ibid.*

way of rules, there are very many different standards. They find a need to collectively manage heterogeneous legal resources that use different XML formats.[58]

There are good ways to answer this need (Metalex etc). Unfortunately they are not sufficient for managing a common query layer between heterogeneous collections of XML legal resources and providing the semantics with which to manage change over time. Therefore there is a need for developing a repository architecture, which is capable of managing heterogeneous documents through common document ontology and metadata mapping, while managing document change over time. A native XML document database of mixed resources has to be developed, which is able to maintain the flexibility and the expressiveness of the original local standard while also providing and interchanging a solution between several XML standards – all the while also building a solid common basis for making meta-queries between mixed documents.[59]

5. Conclusion

There are many initiatives to achieve "better regulation" in different contexts, not least in the EU. The above review of the various EU measures designed to reach better regulation shows how this question is one of great importance, as the volume and complexity of EU law keeps growing and as its correct implementation is essential if the objectives set at EU level are not to be lost. In many of the contexts mentioned in the better regulation review, ICT could be of great use but until now, only rather limited use of the potential of ICT has been made. The paper suggests ways to improve such use. These suggestions are partly built on the efforts made in recent years on ICT support for the analysis of legal texts, most specifically on how to automatically identify structural portions of legal documents through their mutual references and how to grasp semantic information of the legal text. In addition, increasing accessibility of legislation through structural analysis and new visualisation techniques show how ICT can play an important role. The use of correlation tables for the implementation of directives is just one concrete example of where the use of systematisation with the help of ICT can have a great practical impact. As it is important to identify actual correlation, the use of systematisation based on the legal language using the help of ICT can be

[58] M. Palmirani and L. Cervone, "Legal Change Management with a Naitive XML Repository" in G. Governatori (Ed.), *Frontiers in Artificial Intelligence and Applications. Legal Knowledge and Information Systems - JURIX 2009: The Twenty-Second Annual Conference* (IOS Press, Amsterdam 2009) 146-155
[59] *Ibid.*

instrumental. Simplification of legislation is another example and accessibility and presentation could be supported by more extensive and imaginative ICT use.

The approach designed by the authors and presented in this paper offers a method to create a systematic alternative structure to any naturally evolved normative system. It does this through an easily readable and compressed visual picture (a graph) of the legal act and words used in it. The graphs can provide a background for comparisons between selected laws as well as to analyse the entire normative system (the maximal connection graph). This is why such a systematic visualized picture of legal documentation can be very useful also for non-specialists as it gives a quick overview and a basis for more detailed study. The ways to use the visualization inked to the simplification procedures identified for the EU have been outlined above.

Table 1 gives a general picture about specific research activities and how these can contribute to increasing the quality of EU regulation in the broad sense. The table represents the indicative potential of chosen researches to add value to the problem solution. The ability is described by three different values, where meanings of the digits are:

0 –no or very little value

1 –ability to support problem solution process in some way

2 – ability to deal directly with problem through the use of novel solutions and ICT support.

This table sums up the different research activities described in that it places these in the context of the needs identified to improve EU law making. It thus provides an assessment of these system including the system suggested by the authors in a table format – extracting the usefulness of the various theories and projecting these onto the issues to be addressed.

A column called “Computerizable?” is summarizing the potential of all the researches going on and uses the same indications to describe the extent and ability potential of ICT support for problem solution in the close future. As it appears, all the above mentioned issues are under development and have a potential to pay back the expected results to the society during next 10 years. Some are highly “computerizable” and it would be possible to achieve gains from ICT use immediately, provided the technical solutions are at hand and properly adjusted to the issues. A very large number of matters can in some way benefit from ICT.

Table 1. A cross table reflecting the ongoing research in the field of subject and

EC Better Regulation activity		Legal informatics research field									Computerizable?
		Metalex	Complexity	Visualisation	Syntac-tic analysis	Multi-lingual draf-ting	Meta draf-ting	Cont-ent mana-geme-nt	STIA	Chan-ge mana-geme-nt	
Impact assessment	Impact coherency analysis	1	0	1	0	0	0	0	0	1	1
	Consideration of various policies	0	0	1	0	0	0	0	0	0	1
	Consultation of stakeholders	0	1	1	1	0	0	0	0	1	1
	Achieved transparency	1	1	1	1	1	0	0	0	1	1
Consultation		0	1	1	0	0	1	0	1	1	1
Expertise		0	1	1	1	1	0	1	0	0	1
Administrative costs	Cross-country area comparisons	1	2	1	0	1	1	1	0	1	2
	Procedural differences	1	0	0	0	1	0	1	0	1	2
	Offered economies	2	1	2	0	1	1	1	0	1	2
Choice of regulatory instruments	Regulations and directives usage	1	1	1	1	0	0	0	0	0	1
	Reviewing/sun-setting clauses	1	1	1	1	1	1	0	0	0	1
	Alternative instruments usage	1	1	1	1	1	1	0	0	0	1
Transposition/application of EU law		2	0	1	0	2	1	2	0	2	2
Simplification	Repeal	1	1	1	1	1	1	0	1	1	1
	Codification	2	1	2	1	2	1	0	1	1	2
	Recasting	2	1	2	1	1	1	0	0	1	2
	Co-regulation	1	1	2	1	1	1	0	0	1	1
Accessibility/presentation of EU law	EUR-Lex website	2	1	2	2	2	1	2	2	1	2
	Pre-Lex daatabase	2	1	2	2	2	1	2	2	1	2
Evaluation		1	0	1	0	0	1	1	0	0	1
Inter-institutional coordination	Impact assessment	0	0	1	1	0	0	0	0	0	1
	Co-ordination and transparency	2	1	1	0	1	1	2	1	1	2
	Alternative regulatory instruments	1	1	1	1	1	0	1	0	1	1
	Adoption of proposals to simplify EU law	1	1	1	0	0	0	0	0	0	1

0 No support
1 Work process support
2 Content handling support

PAPER IV

E. Täks, L. Vohandu, A. Lohk, and I. Liiv, An experiment to find the deep structure of Estonian legislation. ;In Proceedings of JURIX. 2011, pp 93-102

An experiment to find a deep structure of Estonian legislation

Ermo TÄKS[1] , Leo VOHANDU Ahti LOHK and Innar LIIV.
Department of Informatics,Tallinn University of Technology, Raja 15, 12618, Tallinn, Estonia

Abstract. A normative system evolves over time and, due to several reasons, forms a huge and complex, yet not very systematic nor consistent collection of legal norms. So far qualitative, thus subjective methods are used to create a structure in legislation. A new, experimental and quantitative approach is presented, that opens a rough structure of the legislation with the help of graph theory and visualization. As a test case, a deep structural analysis, based on a part of Estonian legislation is introduced.

Keywords. Legislation structure, quantitative text analysis, legal norms, graph theory, visualization.

Introduction

A legal system is supposed to reduce the level of behavioral uncertainty within a society, but it inevitably has weaknesses that have to be dealt with. The legislation is created by many participants (lawyers, officials, politicians etc.) over a long time period to respond to the important changes within the society, therefore inevitably carrying an element of ambiguity and uncertainty in itself. The discussion in this paper addresses the structure and the referential indeterminacy of the legislation. The main constraint here is the active nature of the role of the legislation - it is used every day by many members of the society for sometimes very important purposes and therefore cannot be changed or altered willingly.

A hierarchical legal system with constitution on top should guarantee the consistency of legal norms in such a way that each and every legal norm within the legal system is not opposing the norms above it - otherwise the norm with a higher hierarchical position, will overrule that with the lower hierarchical position. Thus, there must exist a number of dependencies between different legal norms within the chosen legislation. Visualization of such dependencies could reveal the role and the purpose of a chosen legal norm better.

Many research papers offer a graph-like presentation of the law, helping to visualize the connected parts of the legal documentation based on textual references. [1] [2] [3] [4] [5] [6] [7] However, it is not possible, for example, to visualize neither the Estonian legislation nor a part of it as a connected graph due to inadequately determined legal references. [5] The practical use of such references is often not targeted toward the creation of a logical and consistent structure of the legislation, which is partly causing the above mentioned legal ambiguity and uncertainty.

[1] Corresponding Author.

Therefore a search began to find additional links for better structure and clarity of the legislation.

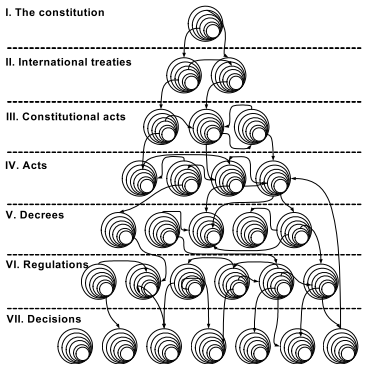


Figure 1. A normative system hierarchy

1. Problem statement

A normative system is a collection of static legal resources. The normative system has a hierarchical structure. A legal system has a built-in demand for internal consistency - legal norms depend on each other and should not be in conflict, at least as far as different hierarchical levels go. These dependencies are visualized in Figure 1.

Hierarchical structure of the normative system is vaguely estimated - it has a well developed referencing system, but does not form a systematic and well structured built-up to form a connected graph. [5] Chosen example, presented in Figure 2, reveals a tight interconnectivity between some parts of it, but hardly provides a systematic overview of the subject.

2. An in-depth structural analysis approach

In order to find a hidden structure of the legislation, a suitable method for legal act similarity measurement had to be found. Due to quite specific needs (relatively small amount of available texts, different impact of legal acts depending on its position in the hierarchical structure, desire to visualize the content and discover the in-depth structure of the legislation etc.) and to open new perspectives, existing text analysis methods were put aside. The main purpose of the search was to find additional dependencies within the normative system, to complement the existing textual references that inadequately visualize the legislation structure.

A suitable way of performing a quantitative comparative analysis of legal texts was found, encompassing system analysis, natural language processing and graph theory methods. System analysis was implemented in two aspects: reverse engineering and business process modeling. Natural language was used as a basis for text segmentation, text compression and content analysis. Graph theory was used for compressed text visualization and similarity calculations.

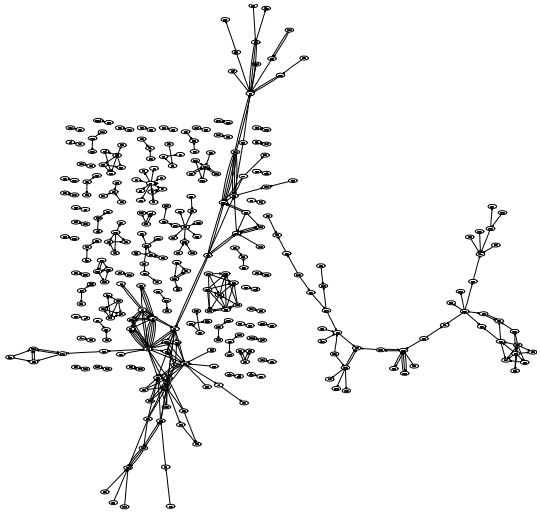


Figure 2. The visual representation of textual references between subdivisions within the Law of Obligations Act

2.1. Reverse engineering

In order to create a new structure for the normative system a smallest, atomic unit had to be found. These legal particles can be used for restructuring the existing system without demolishing it. Alternating views, created in such a way can complement each other, open a number of hidden perspectives and interact successfully.

Reverse engineering is a process of analyzing a subject system to:

- Identify the system components and their interrelationships;
- Create representations of the system in another form or at a higher level of abstraction. [8]

Normative system is a collection of norms. A legislative text may be seen as a vehicle that contains and transports rules; and the legal order as an organism of rules rather than that of laws - as a result enabling us to observe its contents better. Rules are, therefore, seen as the true foundations of normative systems, whilst laws are merely temporal. [9] Relevantly the norm can be used as an atomic unit, within the legal system, and also for its structural decomposition.

2.2. Natural language processing

Legal and linguistic aspects of a norm are tightly bounded: the norm can be understood as “thought (i.e. meaning) content expressed through language”. The norm as a rule receives expression in a norm sentence (norm formulation) and vice versa: the norm is a meaning content of the norm formulation. [11] The norms can be found or targeted with the help of linguistic or grammatical constructions. [10] [12]

The clause is a minimal grammatical construction able to deliver the thought content of the norm, containing at least a subject (noun) and a predicate (verb).

According to Fillmore's Case grammar ideology there exists limited amount of abstract semantic relations between verb and its nouns. [13] A short analysis of the usage of different types of words in Estonian legislation identified nouns on 47% of the cases and verbs on 13% of the cases, combining for 60% of the words in legal texts. It is important to notice that informational value of verbs and nouns is significantly higher compared to adjectives, conjunctions etc., though it is hard to measure. Therefore the rest of the word types were left out from the scope of the current work despite some very appealing features (negation, specification, role assignment etc.).

The clause is also acting as a natural language container, binding together norm specific nouns and verbs. Such interpretation allows us to form a visual backbone of the specific norm in Figure 3. [14]

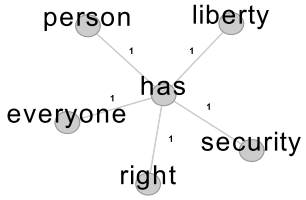


Figure 3. Subgraph of the Constitution norm "Everyone has the right to liberty and security of person."

2.3. Business process modeling (BPM)

Modeling is a common task in BPM that aims to visualize the structure of complex systems. The method allows modeling any system using main activities, events and objects. We applied this method, declaring that verbs within legal text refer to events and activities, nouns to objects. Measuring the concordance - coexistence frequency of verbs (activities and events) and nouns (objects) within sub clauses -, it is possible to create a very rough BPM model of legal text, represented as a graph. This also allows us to compress the analyzed legal text up to 53 times, still embedding, by rough estimations, 60…80% of the legal content. [14] Graph edges represent the words and the size of the edges show the degree of it - to how many different words it connects. In Figure 4, arcs show the connections between the words and the line thickness shows how many times a pair of words appears together in one clause.

2.4. Graph theory

The visual representation of legal text as the graph makes a good formal comparison basis between legal acts and is a key to perform the quantitative analyses of the legal text, using different graph theory aspects and graph mining methods. We applied a data mining technique to estimate the similarity of different graphs.

The process of evaluating the similarity of two graphs is commonly referred to as graph matching. The overall aim of graph matching is to find a correspondence between the nodes and edges of two graphs that satisfies some, more or less, stringent constraints. That is, by means of the graph matching process similar substructures in one graph are mapped to similar structures in the other graph. [15]. Graph isomorphism calculation can be NP-complete task, but as the graph edges have names (a basic form of filtered words due to Estonian language peculiarities) the similarity measurement is

relatively simple. Therefore it is possible to perform graph structural similarity calculations and this notion forms a basis for further restructuring of legislation.

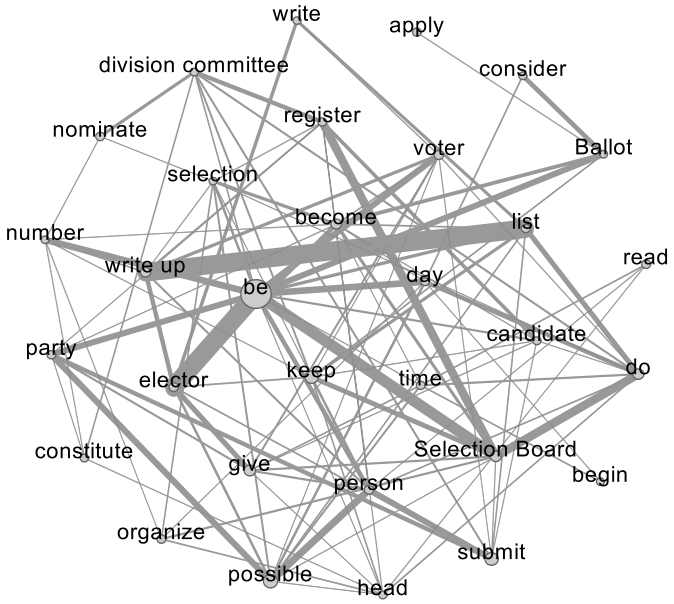


Figure 4. Estonian Parliament Election Act graph, derived from verbs and nouns concordance in the sub clauses.

2.5. Method summary

A norm is a smallest unit of the legislation. Norm is expressed through the linguistic construction, which is the sub clause, containing at least the subject (noun) and the predicate (verb). Nouns, verbs and their relations (frequency of co-use) can be represented in graph form to visualize the norm content and can be used for structural analysis. Two different legal acts are similar if they share same verbs (events or activities), tied with same nouns (objects) and have similar concordance.

3. A test case based on acts layer of Estonian legislation.

To test this theory and to demonstrate the broad applicability of such approach an experimental fitness function was generated. Purpose of this function was to measure similarity of two different legal acts. Function estimated the result in three stages- how many shared verbs were found, how many shared nouns connected to specific verbs and frequency of the use of each pair. The general weight is a sum of:

- Shared verbs account for 45%;
- Shared nouns connected to specific verbs 45%;
- Concordance frequency 10%.

Particular calculation weights were experimental, chosen willingly and are subject for further evaluation.

For the test case of normative system structural analysis:

- 386 Estonian legal acts were randomly chosen;
- Each legal act compared to each other legal act (148996 comparisons).

3.1. A similarity measurement results

As a result a similarity table was created, consisting of 386 rows and 386 columns (see small piece of it in Table 1).

Table 1. Similarity table of legal acts, %

Act ID	13360508	13360557	13360674	13360776
1014238	27	27	10	11
12732218	38	23	22	15
12755417	64	36	37	19
12833542	39	28	29	14
12857270	64	22	32	14
12861944	64	36	31	21
12862402	15	30	0	0
12911615	64	20	21	18
12936755	0	15	15	15
12969907	42	20	27	15

As a result, some interesting general characteristics appeared. In medium, all legal acts seem to share the content roughly by 1/3 (presented in Table 2), but this is not a general rule. In 167 cases (0,22%) two compared legal acts did not share any content. According to results presented in Figure 5, it can be said that there is more likely overlapping between legal acts.

Table 2. General characteristics of similarity table, %

Charateristic	Value
Lowest minimal similarity per act	0
Highest minimal similarity per act	33,1
Lowest maximal similarity per act	39,7
Highest maximal similarity per act	91
Medium similarity	29,9
Lowest medium similarity per act	12,6
Highest medium similarity per act	42,9

In 20 cases the similarity was measured as 80% and higher. Quick control of two most similar legal acts (Estonian Parliament Election Act and European Parliament Election Act) showed a remarkable similarity of texts up to some parts exactly copied from each other.

3.2. A deep structural analysis

However to get an readable overview of such a huge amount of interconnections is impossible for humans. So we have to get a better condensed overview of those similarities. We used several different visualization methods for that purpose.

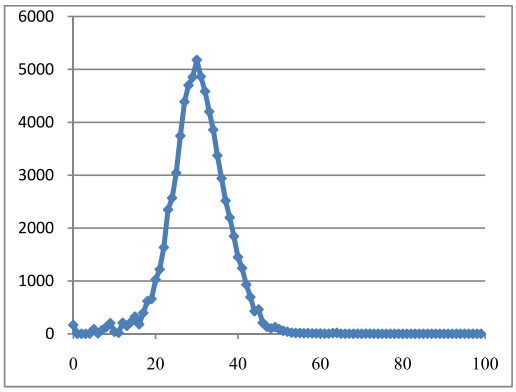


Figure 5. Frequency of similarity results

In order to present complex connected system it has been useful to use extreme representation principles. In our case for Estonian legislation we did use the principle of maximal similarity spanning tree and we computed for the similarity matrix S of legal acts a maximal connected path [16]. In Figure 6 every legal act is connected to its most similar legal act. A zoomed upper part of this graph is shown in Figure 7.

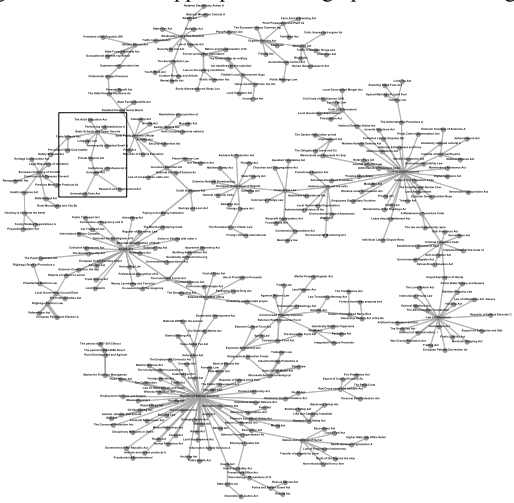


Figure 6. Similarity structure of the Estonian legislation, where each legal act is connected to the most similar one.

All 386 legal acts are represented in one connected tree, which represents the maximal similarities between acts.

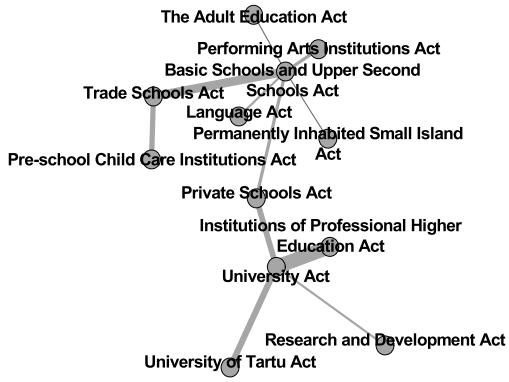


Figure 7. Extracted part of similarity structure

A symmetrical matrix of similarities was transformed into 0/1 matrix based on cutting level of lowest maximum similarity (37%) to get a single connected drawing of all acts. After using the standard visualization method we present the densest part of the acquired graph in Figure 8.

Figure 8. The most dense part of the graph constructed from the strongest connections

As the last method of data reorganization we used a BEA method. [17] In Figure 9 the dense black regions describe strongly similar legal acts subsets. The further deeper legal analysis of such dense regions belongs to the realm of legal expertise.

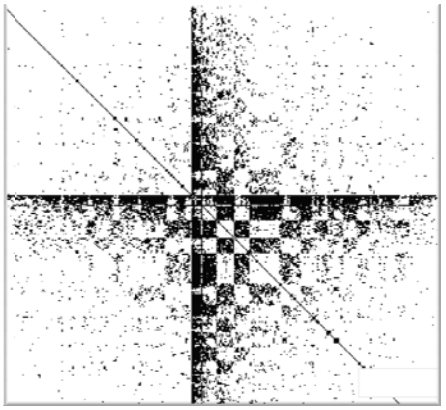


Figure 9. Reordered table to group similar legal acts.

4. Contribution

The presented approach offers a way to open hidden structure of any naturally evolved normative system. The proposed method allows to create easily readable compressed visual pictures (a graph) of the legal act specifying concordance of verbs and nouns within clauses. Using these graphs to measure the similarities of different legal acts it enables to analyze the whole normative system, visualize its internal structure, locate dense interconnectivity between legal acts, group and regroup acts accordingly. This method is mathematically well justified and is step toward objective, quantitative measurement, possibly altering the subjective legal act classification used so far.

5. A future research

The existing textual references within legal texts form a natural part of legislation. A combination of textual and similarity references could give valuable insights and greatly enrich the visual presentation.

The presented graphs and tables express the hidden structure of only one layer - legal acts layer. Similar structure can be found on each level of a normative system and it is possible to connect these levels to one another.

The proposed mathematical approach has it weaknesses. It is not quite suitable to take into account the very short or very specific legal texts. So far we found three types of legal acts, which present problems in this case- very short ones, very standard ones and specific ones (state rewarding proposals, including one or two verbs only). This is a problem because these acts form a natural part of a normative system and should not be excluded. A suitable method to create readable visual layouts including very small graphs is under development.

References

[1] M. J. Bommarito II, D. M. Katz, "A Mathematical Approach to the Study of the United States Code," *Physica A*, vol. Vol. 389, March 25, 2010.

[2] M. J. Bommarito II, D. M. Katz, "Properties of the United States Code Citation Network," *Working Paper Series*, no. SSRN eLibrary, November 2009.

[3] D. Bourcier, P. Mazzega, "Codification, Law Article and Graphs," in *Proceeding of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference*, Amsterdam, 2007.

[4] F. B. Cross, T. A. Smith, "The Reagan Revolution in the Network of Law," no. http://ssrn.com/paper=909217, 2006.

[5] I. Liiv, A. Vedeshin, E. Täks, "Visualization and structure analysis of legislative acts: a case study on the law of obligations," in *ICAIL '07: Proceedings of the 11th international conference on Artificial intelligence and law*, Stanford, California, 2007, pp. 189-190.

[6] P. Mazzega, D. Bourcier, R. Boulet, "The network of French legal codes," in *Proceedings of the 12th International Conference on Artificial Intelligence and Law*. Barcelona, Spain: ACM, 2009, pp. 236--237, http://doi.acm.org/10.1145/1568234.1568271.

[7] P. Zhang, L. Koppaka, "Semantics-Based Legal Citation Network," in *ICAIL '07: Proceedings of the 11th international conference on Artificial intelligence and law*, Stanford, California, 2007.

[8] J. E. Chikofsky, H. J. Cross, "Reverse engineering and design recovery: a taxonomy," *Software*, vol. 7, no. IEEE, pp. 13-17, January 1990.

[9] C. Biagioli, "Towards a legal rules functional micro-ontology," , vol. In *Proc. of workshop LEGONT '97, 1997.*

[10] M. Nakamura, S. Nobuokau, A. Shimazu, "Towards translation of legal sentences into logical forms," in *Proceedings of the 2007 conference on New frontiers in artificial intelligence, JSAI'07*, Miyazaki, Japan, 2008.

[11] Aulis Aarnio, *The Rational as Reasonable. A Treatise on Legal Justification*.: Blackwell Publishing Ltd, 1988.

[12] E. de Maat, R. Winkels, "Automated Classification of Norms in Sources of Law," in *Semantic Processing of Legal Texts*, Enrico Francesconi et al., Eds.: Springer Berlin / Heidelberg, 2010, pp. 170-191.

[13] J. C. Fillmore, *The Case for Case*. New York: Holt, Rinehart, Winston, 1968, vol. Universals in Linguistic Theory.

[14] E. Täks, A. Lohk, "An Alternative Method for Computerized Legal Text Restructuring," in *Proc. of conf. JURIX 2010*, 2010, pp. 171-174.

[15] H. C. Aggarwal, H. Wang, *Managing and mining graph data.*: Springer, 2010, vol. Volume 40 of Advances in Database Systems.

[16] L. K Vohandu, "Analysis of Biological Systems with many Variables," in *III Conference of Mathematical Methods in Biology.*, 1961, pp. 19-22.

[17] W. T. McCormick, Paul J. Schweitzer Jr., Thomas W. White, "Problem Decomposition and Data Reorganization by a Clustering Technique," no. 20:993-1009, 1972.

DISSERTATIONS DEFENDED AT TALLINN UNIVERSITY OF TECHNOLOGY ON *INFORMATICS AND SYSTEM ENGINEERING*

1. **Lea Elmik**. Informational Modelling of a Communication Office. 1992.
2. **Kalle Tammemäe**. Control Intensive Digital System Synthesis. 1997.
3. **Eerik Lossmann**. Complex Signal Classification Algorithms, Based on the Third-Order Statistical Models. 1999.
4. **Kaido Kikkas**. Using the Internet in Rehabilitation of People with Mobility Impairments – Case Studies and Views from Estonia. 1999.
5. **Nazmun Nahar**. Global Electronic Commerce Process: Business-to-Business. 1999.
6. **Jevgeni Riipulk**. Microwave Radiometry for Medical Applications. 2000.
7. **Alar Kuusik**. Compact Smart Home Systems: Design and Verification of Cost Effective Hardware Solutions. 2001.
8. **Jaan Raik**. Hierarchical Test Generation for Digital Circuits Represented by Decision Diagrams. 2001.
9. **Andri Riid**. Transparent Fuzzy Systems: Model and Control. 2002.
10. **Marina Brik**. Investigation and Development of Test Generation Methods for Control Part of Digital Systems. 2002.
11. **Raul Land**. Synchronous Approximation and Processing of Sampled Data Signals. 2002.
12. **Ants Ronk**. An Extended Block-Adaptive Fourier Analyser for Analysis and Reproduction of Periodic Components of Band-Limited Discrete-Time Signals. 2002.
13. **Toivo Paavle**. System Level Modeling of the Phase Locked Loops: Behavioral Analysis and Parameterization. 2003.
14. **Irina Astrova**. On Integration of Object-Oriented Applications with Relational Databases. 2003.
15. **Kuldar Taveter**. A Multi-Perspective Methodology for Agent-Oriented Business Modelling and Simulation. 2004.
16. **Taivo Kangilaski**. Eesti Energia käiduhaldussüsteem. 2004.
17. **Artur Jutman**. Selected Issues of Modeling, Verification and Testing of Digital Systems. 2004.

18. **Ander Tenno**. Simulation and Estimation of Electro-Chemical Processes in Maintenance-Free Batteries with Fixed Electrolyte. 2004.
19. **Oleg Korolkov**. Formation of Diffusion Welded Al Contacts to Semiconductor Silicon. 2004.
20. **Risto Vaarandi**. Tools and Techniques for Event Log Analysis. 2005.
21. **Marko Koort**. Transmitter Power Control in Wireless Communication Systems. 2005.
22. **Raul Savimaa**. Modelling Emergent Behaviour of Organizations. Time-Aware, UML and Agent Based Approach. 2005.
23. **Raido Kurel**. Investigation of Electrical Characteristics of SiC Based Complementary JBS Structures. 2005.
24. **Rainer Taniloo**. Ökonoomsete negatiivse diferentsiaaltakistusega astmete ja elementide disainimine ja optimeerimine. 2005.
25. **Pauli Lallo.** Adaptive Secure Data Transmission Method for OSI Level I. 2005.
26. **Deniss Kumlander**. Some Practical Algorithms to Solve the Maximum Clique Problem. 2005.
27. **Tarmo Veskioja**. Stable Marriage Problem and College Admission. 2005.
28. **Elena Fomina**. Low Power Finite State Machine Synthesis. 2005.
29. **Eero Ivask**. Digital Test in WEB-Based Environment 2006.
30. **Виктор Войтович**. Разработка технологий выращивания из жидкой фазы эпитаксиальных структур арсенида галлия с высоковольтным p-n переходом и изготовления диодов на их основе. 2006.
31. **Tanel Alumäe**. Methods for Estonian Large Vocabulary Speech Recognition. 2006.
32. **Erki Eessaar**. Relational and Object-Relational Database Management Systems as Platforms for Managing Softwareengineering Artefacts. 2006.
33. **Rauno Gordon**. Modelling of Cardiac Dynamics and Intracardiac Bio-impedance. 2007.
34. **Madis Listak**. A Task-Oriented Design of a Biologically Inspired Underwater Robot. 2007.
35. **Elmet Orasson**. Hybrid Built-in Self-Test. Methods and Tools for Analysis and Optimization of BIST. 2007.

36. **Eduard Petlenkov**. Neural Networks Based Identification and Control of Nonlinear Systems: ANARX Model Based Approach. 2007.
37. **Toomas Kirt**. Concept Formation in Exploratory Data Analysis: Case Studies of Linguistic and Banking Data. 2007.
38. **Juhan-Peep Ernits**. Two State Space Reduction Techniques for Explicit State Model Checking. 2007.
39. **Innar Liiv**. Pattern Discovery Using Seriation and Matrix Reordering: A Unified View, Extensions and an Application to Inventory Management. 2008.
40. **Andrei Pokatilov**. Development of National Standard for Voltage Unit Based on Solid-State References. 2008.
41. **Karin Lindroos**. Mapping Social Structures by Formal Non-Linear Information Processing Methods: Case Studies of Estonian Islands Environments. 2008.
42. **Maksim Jenihhin**. Simulation-Based Hardware Verification with High-Level Decision Diagrams. 2008.
43. **Ando Saabas**. Logics for Low-Level Code and Proof-Preserving Program Transformations. 2008.
44. **Ilja Tšahhirov**. Security Protocols Analysis in the Computational Model – Dependency Flow Graphs-Based Approach. 2008.
45. **Toomas Ruuben**. Wideband Digital Beamforming in Sonar Systems. 2009.
46. **Sergei Devadze**. Fault Simulation of Digital Systems. 2009.
47. **Andrei Krivošei**. Model Based Method for Adaptive Decomposition of the Thoracic Bio-Impedance Variations into Cardiac and Respiratory Components. 2009.
48. **Vineeth Govind**. DfT-Based External Test and Diagnosis of Mesh-like Networks on Chips. 2009.
49. **Andres Kull**. Model-Based Testing of Reactive Systems. 2009.
50. **Ants Torim**. Formal Concepts in the Theory of Monotone Systems. 2009.
51. **Erika Matsak**. Discovering Logical Constructs from Estonian Children Language. 2009.
52. **Paul Annus**. Multichannel Bioimpedance Spectroscopy: Instrumentation Methods and Design Principles. 2009.

53. **Maris Tõnso**. Computer Algebra Tools for Modelling, Analysis and Synthesis for Nonlinear Control Systems. 2010.
54. **Aivo Jürgenson**. Efficient Semantics of Parallel and Serial Models of Attack Trees. 2010.
55. **Erkki Joasoon**. The Tactile Feedback Device for Multi-Touch User Interfaces. 2010.
56. **Jürgo-Sören Preden**. Enhancing Situation – Awareness Cognition and Reasoning of Ad-Hoc Network Agents. 2010.
57. **Pavel Grigorenko**. Higher-Order Attribute Semantics of Flat Languages. 2010.
58. **Anna Rannaste**. Hierarcical Test Pattern Generation and Untestability Identification Techniques for Synchronous Sequential Circuits. 2010.
59. **Sergei Strik**. Battery Charging and Full-Featured Battery Charger Integrated Circuit for Portable Applications. 2011.
60. **Rain Ottis**. A Systematic Approach to Offensive Volunteer Cyber Militia. 2011.
61. **Natalja Sleptšuk**. Investigation of the Intermediate Layer in the Metal-Silicon Carbide Contact Obtained by Diffusion Welding. 2011.
62. **Martin Jaanus**. The Interactive Learning Environment for Mobile Laboratories. 2011.
63. **Argo Kasemaa**. Analog Front End Components for Bio-Impedance Measurement: Current Source Design and Implementation. 2011.
64. **Kenneth Geers**. Strategic Cyber Security: Evaluating Nation-State Cyber Attack Mitigation Strategies. 2011.
65. **Riina Maigre**. Composition of Web Services on Large Service Models. 2011.
66. **Helena Kruus**. Optimization of Built-in Self-Test in Digital Systems. 2011.
67. **Gunnar Piho**. Archetypes Based Techniques for Development of Domains, Requirements and Sofware. 2011.
68. **Juri Gavšin**. Intrinsic Robot Safety Through Reversibility of Actions. 2011.
69. **Dmitri Mihhailov**. Hardware Implementation of Recursive Sorting Algorithms Using Tree-like Structures and HFSM Models. 2012.

70. **Anton Tšertov**. System Modeling for Processor-Centric Test Automation. 2012.
71. **Sergei Kostin**. Self-Diagnosis in Digital Systems. 2012.
72. **Mihkel Tagel**. System-Level Design of Timing-Sensitive Network-on-Chip Based Dependable Systems. 2012.
73. **Juri Belikov**. Polynomial Methods for Nonlinear Control Systems. 2012.
74. **Kristina Vassiljeva**. Restricted Connectivity Neural Networks based Identification for Control. 2012.
75. **Tarmo Robal**. Towards Adaptive Web – Analysing and Recommending Web Users` Behaviour. 2012.
76. **Anton Karputkin**. Formal Verification and Error Correction on High-Level Decision Diagrams. 2012.
77. **Vadim Kimlaychuk**. Simulations in Multi-Agent Communication System. 2012.
78. **Taavi Viilukas**. Constraints Solving Based Hierarchical Test Generation for Synchronous Sequential Circuits. 2012.
79. **Marko Kääramees**. A Symbolic Approach to Model-based Online Testing. 2012.
80. **Enar Reilent**. Whiteboard Architecture for the Multi-agent Sensor Systems. 2012.
81. **Jaan Ojarand**. Wideband Excitation Signals for Fast Impedance Spectroscopy of Biological Objects. 2012.
82. **Igor Aleksejev**. FPGA-based Embedded Virtual Instrumentation. 2013.