Adapting Software Metrics to Analyze the Evolution of Laws

An Italian Case Study

Aaron CIAGHI [a], Andrea DALLA VALLE [a], Adolfo VILLAFIORITA [a]

[a] *Fondazione Bruno Kessler, via Sommarive 18, 38123 Trento, Italy*

Abstract. Law-makers, designers of legal information systems and citizens are often challenged by the complexity of bodies of law and the growing number of references needed to interpret a law. Quantifying this complexity is not an easy task. In this paper we present some analyses we conducted on the Italian body of laws, made available through the "Normattiva" website. Some of the metrics we applied are similar to those often used to measure the quality of software systems.

Keywords. software metrics, databases, references, Italian legal system

1. Introduction

Legal informatics, that is the application of ICT to laws, is a mature discipline [1,2] embracing various applications, such as the management of legal documents [3–5], the application of automated reasoning to laws [6,7], the definition of new tools to support the legislative process [8–10], the definition of new policies.

In this paper, we present a quantitative analysis of the Italian body of laws, starting from 1947, which we conducted by applying various metrics typically used for measuring software quality. We started by looking at the similarities between software components (specifically, classes in object oriented programming) and acts. We experimented this similarity by measuring the evolution of how laws are written in Italy since 1947. The results are a set of quantitative data and a graph representation of the Italian body of laws. Our metrics analyze both the whole set of laws (e.g. number of laws in force out of the total number of published laws at one point in time) and individual nodes in the graph by measuring the number of cross-references among different acts.

The goal of this approach is three-fold: it can assist lawmakers by providing metrics on non-trivial aspects of laws, it can assist designers and engineers of legal knowledge management systems, and it can provide citizens with a more transparent and holistic view of the laws of their country.

The rest of the paper is structured as follows. In Section 2 we present some of the current approaches to legal documents management, in Section 3 we present our approach and we discuss the similarities between laws and software. We

present a case study on the evolution of the Italian laws in Section 4. Finally, we draw our conclusions and briefly describe some future work in Section 5.

2. Background

Since the beginning, legal knowledge management has primarily dealt with the storage of legal information. As the technological progress has advanced, legal informatics has also started to support the activity of Parliaments [11, 12], as a part in the re-engineering of legislative processes, as a tool to support automated reasoning over legal documents, and to serve several other purposes [1, 13].

While legal informatics has been initially focused on the architectural question of whether legal documents should be available on centralized systems or on sparse systems [1], more recently, researchers started to agree on the fact that a standard should be adopted. A standard can guarantee several advantages: for example, it allows the exchange, the research and the retrieval of documents between several sources and across different countries [1].

Several initiatives take into account and realize the requirements described in [1]. For example, the project *"NormeInRete"* [14,15] used an open-format, compatible with MetaLEX [16,17], thus ensuring interoperability, and – consequently – accessibility of documents. The process of proposing, refining and improving standards should eventually lead to the adoption of a common standard almost worldwide.

A different line of work focuses on analyzing the correlation and interconnections among laws, with the goal of understanding the quality and complexity of a system of laws. To our knowledge, the American legal system has been the one that has undergone the widest series of studies in this direction. We mention [18–21]. Another notable example is the analysis of the French system conducted in [22–24].

3. Applying Software Metrics to Laws

From a technical perspective, the management of a legal document does not differ significantly from the management of any other document. However, the particular content that legal documents provide, together with their complex and evolving nature, highlights the need for the adoption of dedicated and context-specific solutions, as noted in [1, 25, 26].

We can identify some commonalities between laws and software lifecycles:

- Similarly to laws, software is written, modified, and deleted.
- Software requires traceability of changes, as it happens for laws.
- In order to interpret correctly a law, the references it contains should be known. Similarly, the source code cannot be compiled, until the external dependencies it declares are solved and compiled as well. In addition, in both cases the process is recursive.

- Software requires maintenance, as bugs and issues are discovered, and complexity grows. Similarly, laws should produce a legislation which is internally coherent, thus avoiding incorrect interpretations. When this does not happen, corrective actions have to be taken, in order to ensure that the internal coherency is restored.

Several studies have proposed metrics to measure complexity and maintainability of software systems. We mention [27–30]. In this work, in particular, we start from complexity metrics, such as those proposed in [31], which are meant to measure the complexity of a software system by looking at the size of each class (that is, a legal document, in our analogy), the number of methods per class (that is, the number of articles per document), the size of each method within a class (that is, the size of articles in a legal document), the coupling between objects (that is, the number of references to other documents), and their evolution over time.

4. An Italian Case Study

In 2010, we applied our adaptation of software metrics to laws in order to obtain a quantitative measure of how legal production has changed in Italy in the last 60 years and to gain insights on the maintainability of laws. We are currently completing the necessary infrastructure to make the results publicly available[1]. In this section we present our analysis of the following aspects:

- Size of the whole body of laws, in terms of number of new acts per year and acts currently in force.
- Average size of acts between 1947 and 2010.
- Average number and type of references among acts (i.e., coupling).
- Stability of acts, namely the average number of modifications undergone by an act.
- Types of acts published between 1947 and 2010, in order to identify mutating trends in Italian law-making.

We used the content of the Italian database of laws "Normattiva"[2] [32], maintained by the *"Istituto Poligrafico e Zecca dello Stato"*[3] as dataset for this case study. The database contains all the laws published after 1945 (more than 70,000 documents) and it allows users to search for acts and view their consolidated text at any point in time within their efficacy period. "Normattiva" uses the same URN based identification of acts used by Normeinrete [14]. Each act is also associated to any other act that it modifies and the modified act is enriched with a new version of its text (similarly to what happens in code versioning systems).

Unfortunately, "Normattiva" also contains documents stored as PDFs or plain text without annotations. Since it is not possible to run identify references within these documents using "Normattiva"'s parser, we restricted our analysis to the

---

[1] The results will be published on `http://www.ict4g.org/legalmetrics` in late 2011.
[2] `http://www.normattiva.it`
[3] The institute responsible for the publication of laws in Italy.
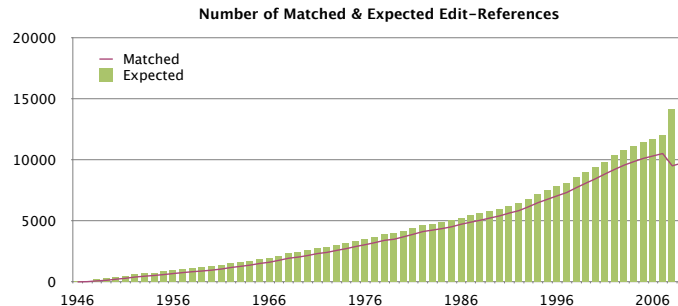
Figure 1. Absolute number of references detected in documents using "Normattiva"'s parser.
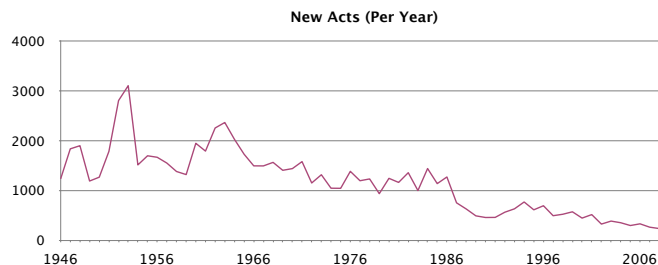

Figure 2. New acts published per year.

set of documents whose references to other acts are recognized by "Normattiva". However, Figure 1 displays that, as we look at more recent laws, detecting references to other documents becomes easier for the parser built in "Normattiva". We observe a sudden drop in effectiveness around 2008 due to a massive simplification process undergone by the Italian body of laws that affected primarily very old acts only stored as PDF scans.

4.1. New Acts and Acts in Force

We have taken a quantitative measure of the number of new acts that are published each year in Italy and of the number of acts currently in force. Figure 2 shows the number of new acts published each year since 1946. We can observe a trend since the end of the '80s of reducing the number of new laws promulgated. This can be interpreted as a change in the process of law-making. In fact, the other notable phenomenon is observable in 1953, in which more than 1500 acts dealing with "property transfer" were promulgated. In more recent decades, law-makers seem to have preferred grouping similar measures in one single act.

Figure 3 shows a comparison of the total number of acts published and the total number of acts currently in force since 2000. We can easily observe that most of the published acts remain in force. Only recently – after 2008 – we see a change in this trend, due to the already mentioned simplification initiative that repealed a large number of old acts.
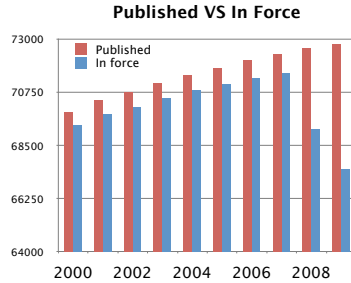
Figure 3. Number of acts in the dataset vs. number of acts currently in force after 2000.
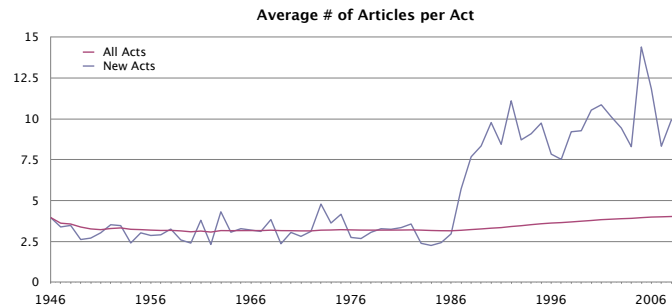

Figure 4. Average number of articles per act. The blue line indicates the size of new acts, while the red line represents the average length of all published acts.
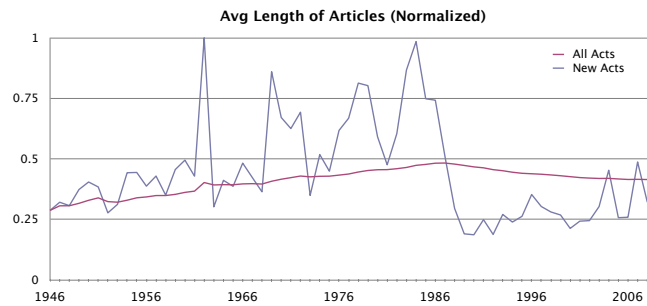

Figure 5. Average length of articles (normalized).

4.2. Size of Acts

We measured the size of acts by looking at how the number of articles per act and the length (number of words) of an article have changed over the years. The graphs in figures 4 and 5 show that after 1988 the average number of new articles explodes, more than doubling the overall number of articles in the whole legislation. On the other hand, the average length of articles similarly decreases.
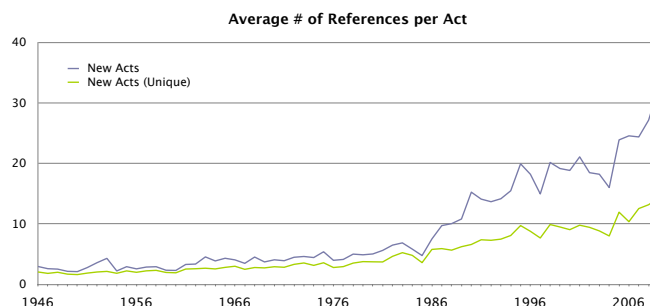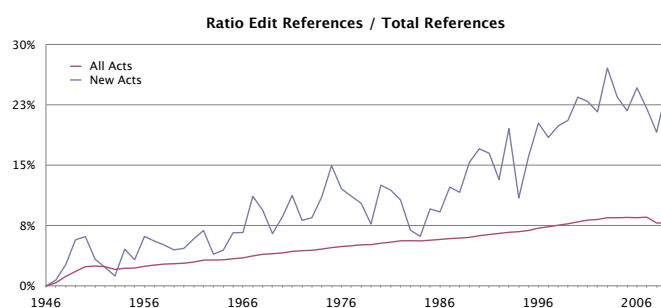
Figure 6. Average number of references per act.

Figure 7. Average number of *Update* references per act over total number of references.
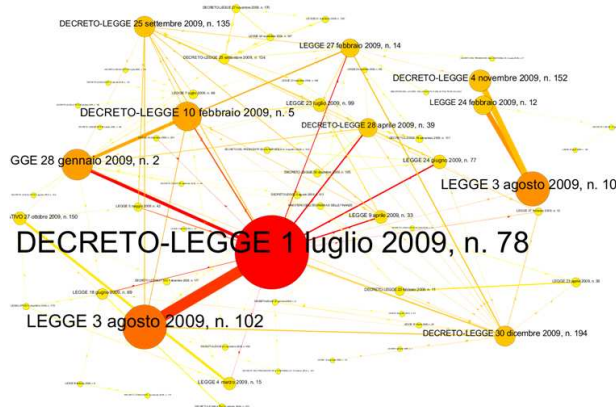

Figure 8. Visualization of the graph of references of a set of acts published in 2009.

4.3. References

In order to measure how tightly acts are coupled, we analyzed how the average number of references in acts change by year of publication. It is reasonable to expect that newer acts tend to contain more references. Figure 6 displays the evolution of the number of references to acts. We can again observe a significant
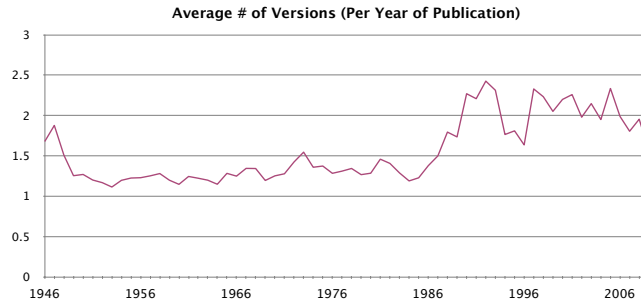
Figure 9. Average number of versions (i.e., edits) per act per year of publication.

change after 1988. This can be interpreted as an increase in complexity of legal documents if we consider the number of references as the number of times in which a reader will be forced to read another document in order to understand a given act. This also implies that modifying an act published after 1988 might indirectly impact more acts.

In our analysis, we distinguish two types of references: *Read* (i.e., direct references to other parts of other acts) and *Update* (i.e., references made in the context of a change or repeal of part of a law). We are interested in understanding whether a growth in the size of the body of laws corresponds to a higher tendency to modify the existing documents. Figure 7 shows the ratio between the number of edits and the total number of references. The chart shows that our hypothesis is valid as we see an increment in *Update* references among laws.

Moreover, as the natural representation of the dependencies among laws is a directed graph, we can visualize an act and its references as shown in figure 8. Acts are represented as nodes in the graph and cross-references are represented as arcs. Nodes appear larger according to the number of times they are referenced and the color of the arcs is the same as the color of their destination node.

4.4. Stability of Acts

Finally, in our study we wanted to measure how "stable" acts are by measuring the average number of versions of the same act stored in "Normattiva" 's database (it should be noted that "Normattiva" stores different versions for each article. We therefore counted the number of modifications undergone by the single act in order to obtain the number of versions).

Figure 9 shows the average number of versions per act per year of publication. The chart highlights the fact that after 1988 the number of modifications that an act undergoes on average almost double with respect to acts published prior to 1988. We should consider the fact that the lifecycle of older acts spans over multiple decades, while the lifecycle of newer acts spans over (at most) two decades. This means that, if the average number of versions per year were considered, we would have that the most recently published acts are the ones that are more frequently modified. One possible interpretation is that an act might be more likely to be modified during the beginning of its lifecycle, while less changes happen as the act gets older.
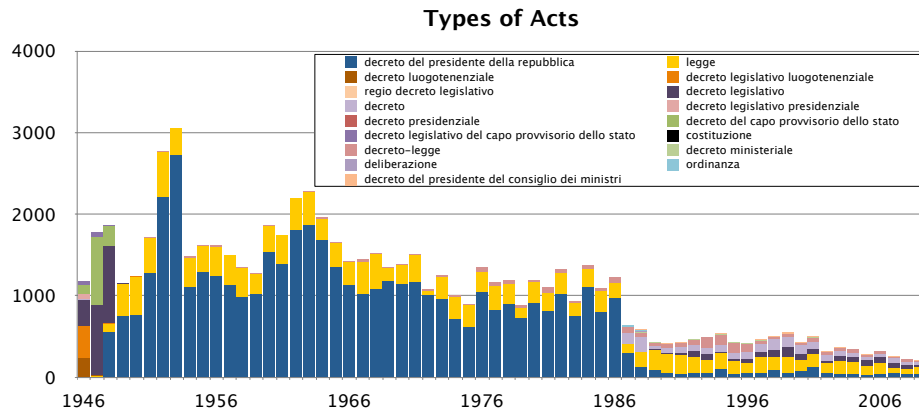
Figure 10. Types of acts published since 1946.

This analysis confirmed what we observed in Section 4.3. As the body of laws becomes older, it is more likely to undergo modifications rather than increase in size.

4.5. Types of acts

Analyzing the types of acts produced since 1947 can help us determining the event that caused the changes in all trends after 1988. Figure 10 displays a bar chart that describes legal production in Italy, divided by type of document. We can identify three different phases:

1. 1946-1948, which corresponds to a period in which the Italian Constitution was being written and the most frequent type of act is the *decreto del capo provvisorio dello stato* (decree of the provisional president).
2. 1949-1986, in which only three types of act are found: *legge* (law), *decreto legge* (decree law), *decreto del Presidente della Repubblica* (decree of the President of the Republic).
3. 1987-present, in which two new types of act are introduced and the *decreto del Presidente della Repubblica* becomes less used. This is due to the publication of a law[4] that redefined the circumstances in which a *decreto del Presidente della Repubblica* can be used, introducing *"decreto legislativo"* (legislative decree) and *"decreto ministeriale"* (ministerial decree). Our hypothesis is that this might have standardized emerging law-making practices.

5. Conclusion

In this work, we have presented a new way of approaching the problem of managing legal documents. Exploiting a possible parallelism with software, we have

---

[4]Legge 400/1988, published on August 23rd, 1988.

tried to define, apply, and examine a narrow set of metrics, in order to understand whether their usage can lead to a better comprehension of the way a legislation is structured, changes and evolves.

Our analysis of the Italian set of laws allowed us to experiment the application of metrics on a real and significant dataset. Our case study demonstrated that metrics are not only applicable, but they also produce meaningful results that can provide insights to different stakeholders.

The results produced by a metrics based analysis are clearly not obtainable through traditional methods. Furthermore, the visualization of our results as line or bar charts – as shown throughout this paper – can clearly convey the results of the analysis and highlight properties that would not be exposed otherwise.

While this work is obviously not complete, we claim that it provides enough justification to pursue this line of research and to complement it with a multi-disciplinary approach that involves also legal experts. Among the advantages we expect by drawing the analogy between the two domains is the possibility of re-using best-practices in both domains.

The identification of more meaningful metrics, a more complete exploitation of the graph representation and the refinement of the interpretation of our results are the natural next steps of this work.

References

[1] M. Biasiotti, E. Francesconi, M. Palmirani, G. Sartor, and F. Vitali. Legal Informatics and Management of Legislative Documents. *Global Center for ICT in Parliament Working Paper*, 2, 2008.

[2] Josep Aguiló-Regla. Introduction: Legal Informatics and the Conceptions of the Law. In *Law and the Semantic Web*, pages 18–24, 2003.

[3] John McClure. The Legal-RDF Ontology. A Generic Model for Legal Documents. In *LOAIT*, pages 25–42, 2007.

[4] Gioele Barabucci and Fabio Vitali. XDTD as a Simple Validation Language for XML-based Legal Documents. In *JURIX*, pages 1–10, 2009.

[5] Caterina Lupo. NormeinRete: a Federative Approach to On-line Legislation Access. In *International Conference Parliaments' Information Management in Africa: Challenges and Opportunities of ICTs to Strengthen Democracy and Parliamentary Governance, Nairobi, Kenya*, 2005.

[6] Emile de Maat and Radboud Winkels. Automated Classification of Norms in Sources of Law. In *Semantic Processing of Legal Texts*, pages 170–191, 2010.

[7] Emile de Maat, Radboud Winkels, and Tom van Engers. Automated Detection of Reference Structures in Law. In *Proceeding of the 2006 conference on Legal Knowledge and Information Systems*, pages 41–50, Amsterdam, The Netherlands, The Netherlands, 2006. IOS Press.

[8] A. Xenakis and A. Macintosh. Using business process re-engineering (BPR) for the effective administration of electronic voting. *The Electronic Journal of e-Government*, 3(2):91–98, 2005.

[9] Sebastian Olbrich, Carlo Simon. Process Modelling Towards e-Government – Visualisation and Semantic Modelling of Legal Regulations as Executable Process Sets. *Electronic Journal of e-Government*, 6, 2008.

[10] A. Ciaghi, A. Mattioli, and A. Villafiorita. A Tool Supported Methodology for BPR in Public Administrations. *International Journal of Electronic Governance*, 2010.

[11] Rinke Hoekstra, Joost Breuker, Marcello Di Bello, and Alexander Boer. LKIF Core: Principled Ontology Development for the Legal Domain. In *Proceeding of the 2009 conference*

on Law, Ontologies and the Semantic Web, pages 21–52, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press.

[12] A. Villafiorita, K. Weldemariam, A. Susi, and A. Siena. Modeling and Analysis of Laws using BPR and Goal-oriented Framework. 2009.

[13] H. Prakken, T. Gordon, D. Walton, T. Bench-Capon, FJ Bex, S.W. den Braak, H. Oostendorp, H. Prakken, HB Verheij, and GAW Vreeswijk. Logical tools for modelling legal argument: a study of defeasible reasoning in law. *Dordrecht, Boston*, 1997.

[14] C. Lupo and C. Batini. A federative approach to laws access by citizens: The "Normein-Rete" system. *Electronic Government*, pages 1061–1061.

[15] C. Lupo and L. De Santis. CNIPA and legislative XML: an update on projects and new initiatives.

[16] A. Boer, R. Winkels, and F. Vitali. Metalex XML and the legal knowledge interchange format. *Computable Models of the Law*, pages 21–41, 2008.

[17] A. Boer, R. Hoekstra, E. de Maat, E. Hupkes, F. Vitali, M. Palmirani, and B. Ratai. CEN MetaLex Workshop Proposal. Technical report, Technical report, European Committee for Standardization (CEN), 2009.

[18] T.A. Smith. The web of law. *University of San Diego Legal Working Paper Series*, page 8, 2005.

[19] J.H. Fowler and S. Jeon. The authority of supreme court precedent: a network analysis. *Preprint as of June*, 29:2005, 2005.

[20] S.J. Chandler. The network structure of supreme court jurisprudence. *Preprint as of June*, 10:2005, 2005.

[21] II Bommarito, J. Michael, D. Katz, and J. Zelner. Law as a seamless web?: comparison of various network representations of the united states supreme court corpus (1791-2005). In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 234–235. ACM, 2009.

[22] Pierre Mazzega, Danièle Bourcier, and Romain Boulet. The network of french legal codes. In *ICAIL* [24], pages 236–237.

[23] Arno R. Lodder and Laurens Mommers, editors. *Legal Knowledge and Information Systems - JURIX 2007: The Twentieth Annual Conference on Legal Knowledge and Information Systems, Leiden, The Netherlands, 12-15 December 2007*, volume 165 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2007.

[24] *The 12th International Conference on Artificial Intelligence and Law, Proceedings of the Conference, June 8-12, 2009, Barcelona, Spain*. ACM, 2009.

[25] M.F. Moens and M. Logghe. Digital legislation: Reflections on the agora-lex project. In *Proceedings of Business Information Systems*, pages 228–233. Citeseer, 2002.

[26] M.F. Moens, M. Logghe, and J. Dumortier. Legislative databases: Current problems and possible solutions. *International Journal of Law and Information Technology*, 10(1):1, 2002.

[27] D. Kafura and G.R. Reddy. The use of software complexity metrics in software maintenance. *Software Engineering, IEEE Transactions on*, (3):335–343, 1987.

[28] D. Coleman, D. Ash, B. Lowther, and P. Oman. Using metrics to evaluate software system maintainability. *IEEE Computer*, pages 44–49, 1994.

[29] N.F. Schneidewind. The state of software maintenance. *Software Engineering, IEEE Transactions on*, (3):303–310, 1987.

[30] S.S. Yau and J.S. Collofello. Design stability measures for software maintenance. *IEEE Transactions on Software Engineering*, pages 849–856, 1985.

[31] S. R. Chidamber and C. F. Kemerer. A metrics suite for object oriented design. *IEEE Trans. Softw. Eng.*, 20:476–493, June 1994.

[32] F. Venturini. La banca dati normattiva. *Le Carte e la Storia*, (1):37–39, 2010.