

Towards an Automatic Consolidation of French Law

GEORGES-ANDRÉ SILBER, CRC, Mines Paris, PSL University, France

CCS Concepts: • **Information systems** → **Information extraction**; • **Software and its engineering** → **Domain specific languages**; • **Applied computing** → **Law**.

1 INTRODUCTION

The texts that are part of French law are modified by amending texts published in the Official Journal of the French Republic (JORF). The life cycle of a legislative or regulatory text begins with the publication of its complete version in the JORF and continues with the possible publication of texts amending it. The full amended text, called its *consolidated* version, is never published in the JORF and has no legal value: only the initial version and the suite of the ordered modifications of the text are authentic [8].

Since 2008, the French Légifrance [4] website presents most of the legal texts in their original versions as well as in their successive versions, consequences of the modifications brought to these texts over time. The operator of the Légifrance website, the Direction of Legal and Administrative Information (DILA), manually reports the modifications described in natural language in the texts in order to obtain, at each modification date, the complete consolidated version of the text.

This convenience of access to the texts in an easier-to-read and easier-to-use version has de facto changed the status of these consolidated versions: they are seen by most users, including legal professionals, as the reflection of the applicable law [7]. Moreover, the drafters of new texts, in the French parliament or in the ministries, start from this consolidated version to conceive the modifying texts. It is therefore of the utmost importance that this consolidation work be free of errors and available as soon as possible.

We present here preliminary work integrated in the Legistix tool that we are developing, whose objective is to create an automated and reliable consolidation system for French legal texts. This work is based both on regular expressions used in several compound grammars, similar to the successive passes of a compiler, and on a new specialized language of functional type, allowing to describe the changes applied to the texts.

For each modifier text, our tool generates fully automatically a computer program in this new language which, when executed, performs the changes induced by the modifier text on the target texts. In previous work on this topic, presented for example in [6], only the problem of classifying types of modification is addressed. To the best of our knowledge, our work is the first to present a complete approach to identify target texts and to transform the natural language instructions of the modifier text into a computer program formalizing the actual transformation rules.

Unlike other topics of study related to legal texts, such as extracting automatically reasonings, rules of law or trying to understand the semantics of the text, we place ourselves in a purely *legistic* [8] perspective, i.e. we do not try to understand the meaning of the text, but only its structure and its relations to other texts. Since our problem is clearly defined, it lends itself well to automation, since the consolidation must be careful not to interpret the changes but only to apply them to the letter [7].

2 EXAMPLE OF CONSOLIDATION

In order to illustrate the complexity of a tool like Legistix, we will use in this document an example of consolidation based on the law 2022-1348 of 24 October 2022 published in the JORF of 25 October 2022 and presented fig. 1. Like all texts published in the JORF, unless otherwise indicated, this law came into force the day after its publication, i.e. on October 26, 2022. We can also note in

paragraph III of the law a *delayed entry into force* of some induced changes on January 1, 2023. The text targeted by the changes is article L723-4 of the Commercial Code is presented in fig. 2.¹

Article 1 of law 2022-1348 of october 24, 2022 (partial)

- I. - Article L723-4 of the Commercial Code is amended as follows:
- 1° At the beginning of the first paragraph, the following is added: "I. -";
 - 2° In 1°, the second occurrence of the word: "and" is replaced by the word: "or";
 - 3° In 3° and 4°, after the word: "judicial", if inserted the word: "rescue";
 - 4° In 4° bis, the second occurrence of the word: "were" is suppressed;
 - 5° In 5°, after the word: "qualities", are inserted the words: "and duties";
 - 6° The last paragraph is replaced by a II redacted as follows:
 "II. - Also eligible are, [...]
 "2° [...] establishments registered in the trade directory or within the jurisdiction of the courts. [...]"
- II. - In the first sentence of 2° of II of article L723-4 of the French Commercial Code, as amended by I of this article, the words: "trade directory" are replaced by the words: "national register of company or establishment in the trades and crafts sector".
- III. - II of this article shall apply as of January 1, 2023. This act shall be executed as a law of the of the state.
-

Fig. 1. Partial reproduction of the unique article of the law 2022-1348 of october 24, 2022 published in the official journal of october 25, 2022 [13], with the identifier fr/loi/2022-1348/1/20221025. Non official translation from the French. Missing text is between [...].

Article L723-4 of the French Commercial Code, in force since october 13, 2021

Persons who are at least thirty years of age shall be eligible for election as a judge of a commercial court:

- 1° Registered on the electoral lists of the chambers of commerce and industry and the chambers of trade and crafts [...]
- 3° In respect of which a judicial recovery or liquidation procedure [...]
- 4° [...] a judicial recovery or liquidation procedure is in progress on the day of the vote;
- 4° bis Who were not were subject to the sanctions [...]
- 5° And that justify [...] of the qualities listed [...]

The following are also eligible for election as members of the commercial courts, [...].

Fig. 2. Partial reproduction of the article L723-4 of the French Commercial Code [12], in force since october 13, 2021, with the identifier fr/code/commerce/L723-4/20211013. Non official translation from the French. Missing text is between [...]

The amending text that modifies the target code article contains instructions for humans to modify the original text and to obtain the new text. These instructions are insertions, deletions, replacements, etc. We can also note in the 4° bis paragraph of the article a mistake made by the legislator of a previous amending law voted in the Parliament: "Who were not were...". The objective of our tool is to transform these instructions intended for humans into a computer program allowing to apply them automatically.

¹These changes, applied manually by the DILA, can be viewed online on the Légifrance website: https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000005634379/LEGISCTA000006161381/2022-10-26/#LEGISCTA000006161381

3 DATA, IDENTIFIERS AND VERSIONS

The data integrated into Legistix are made up of the whole of the JORF [1] and LEGI [2] databases provided and updated daily by the DILA. These databases contain all the texts published in the JORF since 1990 as well as all the texts that have been consolidated by the DILA. The data are integrated into the Legistix database on an ongoing basis, thanks to an integration process that improves the data, reconstructing in particular the paragraphs and the versions of the tables of contents not provided in the original data. For example, the Commercial Code has over 600 versions since 2000.

Each document receives a unique identifier in the form of a URI. The form of this URI is a continuation of the work carried out with the DILA in the context of the Légimobile [9] project. These identifiers have the particularity of being perennial and of being able to be systematically derived from the information in the document itself. Thus, for example, article 1 of the law 2022-1348 of October 24, 2022 presented in fig. 1 has the unique URI `fr/loi/2022-1348/1`, representing all the versions of this article over time.

To specify a particular version, the identifiers also support the many dates governing the life cycle of a law: date of signature, date of publication in the JORF, dates of modification leading to versions, date of abrogation. Thus, the version published in the JORF on October 25, 2022 of the previous article has as the URI `fr/loi/2022-1348/1/20221025`. Following the same logic, article L723-4 of the Commercial Code in its version in force since October 13, 2021 presented fig. 2 has the URI `fr/code/commerce/L723-4/20211013`.

4 DETECTION OF MODIFIED TEXTS AND CREATION OF NEW VERSIONS

In the example of law 2022-1348, Legistix detects the reference "*L723-4 of the Commercial Code*" present twice. Thanks to the identifier mechanism described above, it will be resolved into the identifier `fr/code/commerce/L723-4`, without the need to query an external resolution mechanism: the character string of the reference is enough.

Here we can start describing the program obtained automatically by Legistix: presented here for more clarity in a functional language with a Python-like syntax, it will allow us to illustrate some of the mechanisms we are implementing. First of all, the source text `s` can be defined as the article 1 of the law 2022-1348 in its version as of October 25, 2022 as published in the JORF, with `db` an object representing the database of the texts and `get_version` a method returning a reference to an existing version:

```
s = db.get_version("fr/loi/2022-1348/1", Date(2022, 10, 25))
# s is fr/loi/2022-1348/1/20221025
```

Legistix must now find the different versions of the target text that will be used in the described changes. The first version we need is the one from which we start, i.e. the version on which the changes are to be applied. For the first reference found in paragraph I, the date is the date of publication in the JORF of the source text, i.e. October 25, 2022. The reference:

```
t = db.get_version("fr/code/commerce/L723-4", Date(2022, 10, 25))
# t is fr/code/commerce/L723-4/20211013 (!)
# that is, the version in force on October 25, 2022
```

is positioned in article L723-4 in its version as of October 13, 2021. Indeed, the version in force of this text on October 25, 2022 is that of October 13, 2021 [12], as modified by Article 1 of Law 2021-1317 which came into force on the same day [11].

As a text in the JORF comes into force by default on the day after its publication, i.e. here on 26 October 2022, the new version of article L723-4 described in I of law 2022-1348 must be created on this date with the function `new_version`, from the version represented by `t`:

```
v1 = t.new_version(Date(2022, 10, 26))
# v1 is fr/code/commerce/L723-4/20221026, created from target t
```

In II, changes are described with reference to Article L723-4 of the Commercial Code as revised by I of this article, i.e. *after* the changes in I have been applied. Combining this information with the information in III, which indicates that the changes in II are to apply on January 1, 2023, a new version must be created on that date:

```
v2 = v1.new_version(Date(2023, 1, 1))
# v2 is fr/code/commerce/L723-4/20230101, created from version v1
```

The names `s`, `t`, `v1` and `v2` represent abstract references to documents. As we will see in the next section, these references can be used to abstractly reference portions of text. For example, the methods `par` and `sen` will allow to generate functions which will respectively return references to a paragraph or a sentence when they are evaluated. Thus, `s.par("I")` generates a function which, when evaluated, will return a reference to paragraph I of article 1 of law 2022-1348. These generators can be combined as in:

```
v1.par("II").par("2°").sen(1)
```

which generates a function allowing to obtain the first sentence of the 2° of the II of the article L723-4 of the commercial code in its version of october 26, 2022.

5 DETECTION OF CHANGES IN THE MODIFYING TEXTS

After identifying the versions of the target texts mentioned in the modifying texts, Legistix detects the changes and transforms them into a sequence of functions to apply them. From the text in fig. 1, Legistix generates the complete program shown in fig. 3.

It creates two new versions of article L723-4 of the Commercial Code, `v1` effective October 26, 2022, with the changes described in paragraph I of article 1 of law 2022-1348 (`s.par("I")`), applying to the version of the Commercial Code as of October 13, 2021, and `v2` effective January 1, 2023 with the changes described in paragraph II (`s.par("II")`). Note that the second version is created from the first. The methods `prepend`, `replace`, `insert`, `suppress`, `replace_par` are function generators which are not applied directly by the program, as long as the method `apply_changes()` is not called.

Each method used in the program translates an operation from natural language. Each function describing a change is of the form `action(source, target, what...)`, where `source` and `target` indicate respectively the text fragment at the origin of the change and the target. The parameters `what` describe the changes, with for example (`"and", 2`) which represents the second occurrence of the word `"and"` and `"rescue"` a replacement word.

6 CONCLUSION AND PERSPECTIVES

We started with a reference database of several decades where the consolidation was already carried out manually. This historical consolidation serves as a reference to measure the reliability of our approach. The Legistix tool is able to automate 93% of the consolidation operations that were previously performed manually. Our efforts will continue in order to reach a rate of 100%, by analyzing the undetected cases, some of which are ambiguous even for a human, thus contradicting the mechanizable aspect desired by the legislator [7].

The next step will be to extend our work to the EUR-Lex [3] database containing the law of the European Union and where regulations and directives are published on a publication model similar to French law.

Third, we will propose an expert system [5] to assist legislators for the drafting of modifying texts where the formal rules of modification (the program) could be generated directly during the

```
s = db.get_version("fr/loi/2022-1348/1", Date(2022, 10, 25))
t = db.get_version("fr/code/commerce/L723-4", Date(2022, 10, 25))

v1 = t.new_version(Date(2022, 10, 26))
pI = s.par("I")
v1.schedule_changes(s, [
    v1.prepend(pI.par("1"), v1.par(1), "I. -"),
    v1.replace(pI.par("2"), v1.par("1°"), ("and", 2), "or"),
    v1.insert(pI.par("3"), v1.par("3°"), "judicial", "rescue,"),
    v1.insert(pI.par("3"), v1.par("4°"), "judicial", "rescue,"),
    v1.suppress(pI.par("4"), v1.par("4° bis"), ("were", 2)),
    v1.insert(pI.par("5"), v1.par("5°"), "qualities", "and duties"),
    v1.replace_par(pI.par("6"), v1.lastpar(),
        "II. - Also eligible[...]")
])
v1 = db.add_version(v1.apply_changes())

v2 = v1.new_version(Date(2023, 1, 1))
pII = s.par("II")
v2.schedule_changes(s, [
    v2.replace(pII, v2.par("II").par("2°").sen(1),
        "trade directory",
        "national register of company[...]")
])
v2 = db.add_version(v2.apply_changes())
```

Fig. 3. Legistix program derived from article 1 of the law 2022-1348 of October 24, 2022, published in the official journal of October 25 October 25, 2022 [13].

drafting of the text. This would set up a virtuous circle where the impacts of the modifications could be immediately visualized, thus improving and making the production of law more reliable.

When the 100% rate will be reached, there will be no delay anymore between the publication of the modification texts and the availability of the consolidated versions. The consolidated text will only be a by-product of the application of a series of programs starting from the original text. These programs could be voted on at the same time as the amending text, making the adage *code is law* [10] legally real.

ACKNOWLEDGEMENTS

Thanks to Salomé Ouaknine, a student at Mines Paris, for her help during her two-month research internship on the subject of automatic consolidation, especially on the clearing of the future hybrid approach mixing regular expressions and machine learning.

REFERENCES

- [1] Direction of Legal and Administrative Information (DILA) (Eds.). 2022. *Data in XML format containing all laws and regulations published in the Official Journal of the French Republic (JORF) since 1990*. <https://www.data.gouv.fr/fr/datasets/jorf-les-donnees-de-l-edition-lois-et-decrets-du-journal-officiel/>
- [2] Direction of Legal and Administrative Information (DILA) (Eds.). 2022. *Data in XML format containing French law consolidated by the Direction of Legal and Administrative Information (DILA)*. <https://www.data.gouv.fr/fr/datasets/legi->

codes-lois-et-reglements-consolides/

- [3] Publications Office of the European Union (Ed.). 2022. *EUR-Lex, access to European Union Law*. <https://eur-lex.europa.eu>
- [4] Direction of Legal and Administrative Information (DILA) (Eds.). 2022. *Légifrance*. <https://www.legifrance.gouv.fr>
- [5] Alain Delaët, Denis Merigoux, and Aymeric Fromherz. 2022. Turning Catala into a Proof Platform for the Law. In *POPL 2022 - Programming Languages and the Law*. Philadelphia, United States. <https://hal.inria.fr/hal-03447072>
- [6] Samuel Fabrizi, Maria Iacono, Tesei Andrea, and Lorenzo De Mattei. 2022. A First Step Towards Automatic Consolidation of Legal Acts: Reliable Classification of Textual Modifications. In *Proceedings of the Eighth Italian Conference on Computational Linguistics*. Milan, Italy. <http://ceur-ws.org/Vol-3033/paper26.pdf>
- [7] Thierry-Xavier Girardot. 2014. Accéder au droit: importance et défis de la consolidation. *Documentaliste – Sciences de l'Information* 51, 4 (2014), 30–32. <https://www.cairn.info/revue-documentaliste-sciences-de-l-information-2014-4-page-30.htm>
- [8] Secrétariat général du gouvernement and Conseil d'État. 2017. *Guide de légistique*. La documentation française.
- [9] Pierre Larrède and Georges-André Silber. 2011. Un plan de classement des données juridiques françaises: l'expérience Légimobile. In *iExpo2011*. <https://www.gfii.fr/uploads/docs/un-plan-de-classement-des-donnees-juridiques-francaises-l-experience-legimobile-laureat-au-prix-de-la-meilleure-contribution-scientifique-i-expo-2011.pdf>
- [10] Lawrence Lessig. 2000. Code Is Law. *Harvard Magazine* (2000). <http://harvardmagazine.com/2000/01/code-is-law.html>
- [11] French Republic. 2021. *Article 1 of law 2021-1317 of october 11, 2021, published in the french official journal of october 12, 2021*. https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000044183798
- [12] French Republic. 2021. *Article L723-4 of the French Commercial Code, in force on October 13, 2021, following the modifications made by art. 1 of the Law 2021-1317 of October 11, 2021*. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044191424/2021-10-13/
- [13] French Republic. 2022. *Article 1 of law 2022-1348 of october 24, 2022, published in the french official journal of october 25, 2022*. https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000046480974