arXiv:2301.06469v1 [cs.CL] 16 Jan 2023

迈向法国法律的自动合并

GEORGES-ANDRÉ SILBER，CRC，巴黎矿业学院，PSL大学，法国

CCS 概念：• **信息系统** → **信息抽取**；• **软件及其工程** →**领域特定语言**；• **应用计算** → **法律**。

1 引言

构成法国法律的文本会通过在《法兰西共和国官方公报》（JORF）上发布的修订文本进行修改。立法或规章文本的生命周期始于其完整版本在 JORF 上发布，之后可能会继续发布对其进行修订的文本。修订后的完整文本称为其整合版本，但从未在 JORF 上发布，也不具有法律效力：只有初始版本以及按顺序对文本所作的一系列修改才具有真实性 [8]。

自2008年以来，法国的 Légifrance [4] 网站提供大多数法律文本的原始版本以及历次版本，这些版本反映了文本随着时间推移所经历的修正。Légifrance 网站的运营方——法国法律与行政信息局（DILA）——会手动记录文本中以自然语言描述的修正，以便在每个修正日期生成该文本的完整合并版本。

这种便捷的访问方式使文本以更易阅读、更易使用的版本呈现，事实上改变了这些合并版本的地位：包括法律专业人士在内的大多数用户都将其视为现行法律的体现 [7]。此外，法国议会或各部委起草新文本时，也会以该合并版本为基础来拟定修正文本。因此，这项整合工作至关重要，必须确保没有错误，并尽可能及时完成。

本文介绍了我们正在开发的 Legistix 工具中的一项初步工作，其目标是为法国法律文本创建一个自动化且可靠的整合系统。这项工作一方面采用了若干复合语法中使用的正则表达式，其处理方式类似于编译器的连续多轮处理；另一方面采用了一种新型专用函数式类型语言，用于描述应用于文本的变更。

对于每段修改文本，我们的工具都会在这种新语言中全自动生成一个计算机程序；该程序执行时，会对目标文本实施由修改文本引起的变更。此前关于这一主题的研究，例如 [6], , 仅讨论了修改类型分类问题。据我们所知，我们的工作首次提出了一种完整的方法，用于识别目标文本，并将修改文本中的自然语言指令转换为能够形式化实际转换规则的计算机程序。

与法律文本相关的其他研究主题不同，例如自动提取推理、法律规则或尝试理解文本的语义，我们采取的是纯粹的立法 [8]视角，即我们不试图理解文本的含义，而只关注其结构及其与其他文本的关系。由于我们的问题定义明确，因此非常适合自动化，因为整合必须谨慎行事，不对变更作出解释，而只是严格按照字面应用这些变更 [7]。

2 整合示例

为说明 Legistix 这类工具的复杂性，本文将以整合为例，依据2022年10月24日颁布的2022-1348号法律。该法律发表于2022年10月25日的 JORF，见图1。与所有发表于 JORF 的文本一样，除非另有说明，该法律自公布次日起生效，即2022年10月26日。我们还可以注意到

法律第III段规定，部分由此引发的变更将延迟生效，生效日期为2023年1月1日。变更所针对的文本是《商法典》L723-4条，见图2。[1]。

2022年10月24日2022-1348号法律第1条（部分内容）

I. - 《商法典》L723-4条修改如下：
1° 在第一段开头添加以下内容："*I.* -";
2° 在1°中，将第二次出现的词语："和"替换为词语："或";
3° 在3°和4°中，在词语："司法"之后，如已插入词语："救助，";
4° 在4° bis中，删除"曾经"一词首次出现之处；
5° 在5°中，"资质"一词之后，插入"和职责"一语；
6° 最后一段替换为以下拟定的II：
"II. - 以下主体也符合资格：*[...]*
"2° [...] 在商业名录中登记或属于法院管辖范围内的机构。*[...]"*
II. - 在经I修订的《法国商法典》L723-4条II之2°的第一句中
本条中的"商业名录"一词替换为"公司或
手工业领域的营业场所"。
III. - 本条第II款自2023年1月1日起适用。本法应作为国家法律予以执行。

图1. 2022年10月24日第2022-1348号法律唯一条文的部分摘录，该法律刊载于 25, 2022 [13], 官方公报，标识符为 fr/loi/2022-1348/1/20221025。法语非官方译文。缺失文本以 [...]标示。

A《法国商法典》第L723-4条，自2021年10月13日起施行

年满三十周岁者，符合下列条件的，有资格当选商事法院法官：1° 登记在工商会和行业与手工业协会的选民名册上 [...]ȝ [ ] ° 涉及司法重整或清算程序.....° [...][4] 在投票日仍处于司法重整或清算程序中；4°之二 未曾受到制裁 [ ] ... ° 并且能够证明 [...] 所列品质 [...]以下人员也有资格当选商事法院成员， [...]。[5]

图2. 《法国商法典》L723-4条的部分内容 [12], , 自2021年10月13日起生效，标识符为 fr/code/commerce/L723-4/20211013。法语的非官方译文。省略的文本位于 [...]之间

修正文本对目标法典条文进行修改，其中包含供人类修改原文并生成新文本的指示。这些指示包括插入、删除、替换等。我们还可以在该条文的第4° bis段落中发现一处错误，这是先前一部由议会表决通过的修正法的立法者所犯的："原文为....."。我们工具的目标是将这些面向人类的指示转换为计算机程序，从而能够自动应用这些指示。

[1]这些由 DILA 手动实施的更改可在 Légifrance 网站上在线查看：https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000005634379/LEGISCTA000006161381/2022-10-26/#LEGISCTA00000616138

3 数据、标识符与版本

整合到 Legistix 中的数据由 DILA 提供并每日更新的 JORF [1] 和 LEGI [2] 数据库的全部内容组成。这些数据库包含自 1990 年以来在 JORF 上发布的所有文本，以及 DILA 整合的所有文本。借助一套能够改进数据的整合流程，数据会持续整合到 Legistix 数据库中；该流程尤其会重建原始数据中未提供的段落以及目录版本。例如，《商法典》自 2000年起已有 600 多个版本。

每份文件都会获得一个统一资源标识符形式的唯一标识符。该统一资源标识符的形式延续了与 DILA 在 Légimobile [9] 项目中开展的工作。这些标识符的特点是永久有效，并且能够根据文件本身的信息系统地推导出来。例如，图1所示的2022年10月24日第2022-1348号法律第1条，其唯一统一资源标识符为 fr/loi/2022-1348/1，代表该条款随时间推移的所有版本。

为指定某个特定版本，这些标识符还支持法律生命周期中涉及的多个日期：签署日期、在 JORF 上公布的日期、导致版本变更的修改日期，以及废止日期。因此，前述条款于2022年10月25日在 JORF 上公布的版本，其统一资源标识符为 fr/loi/2022-1348/1/20221025。按照同样的逻辑，图2所示《商法典》第L723-4条自2021年10月13日起生效的版本，其统一资源标识符为 fr/code/commerce/L723-4/20211013。

4 检测修改后的文本并创建新版本 S

以2022-1348号法律为例，Legistix 检测到《商法典》第 L723-4 条这一引用出现了两次："*L723-4 of the Commercial Code*"。借助上文所述的标识符机制，该引用将被解析为标识符 fr/code/commerce/L723-4，无需查询外部解析机制：只需引用中的字符串即可。

现在，我们可以开始介绍 Legistix 自动生成的程序：为便于理解，这里用一种类似 Python 语法的函数式语言来呈现，以此说明我们正在实现的一些机制。首先，可以将源文本 s 定义为 JORF 于 2022年10月25日 发布的 2022-1348号法律 第1条当日版本，其中 db 是一个表示文本数据库的对象，而 get_version 是一个返回现有版本引用的方法：

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

Legistix 现在必须找出将用于所述修订的目标文本的不同版本。我们需要的第一个版本是作为起点的版本，也就是将要施加这些修订的版本。对于第 I 段中找到的第一个引用，其日期是源文本在 JORF 上的公布日期，即2022年10月25日。该引用：

```
t = db.get_version("fr/code/commerce/L723-4", Date(2022, 10, 25)) # t
是 fr/code/commerce/L723-4/20211013(!) # 即2022年10月25日生效的版本
```

位于 L723-4 条中，该条的版本截至2021年10月13日。实际上，这段文本在2022年10月25日生效的版本，是经 13, 2021 [12]，同日生效的2021-1317号法律第1条修订后的版本 [11]。

由于 JORF 中的文本默认于公布次日生效，也就是此处的 2022年10月26日，因此必须在该日期使用 new_version 函数，根据 t 所表示的版本创建 2022-1348号法律第 I 部分所述的 L723-4条新版本：

v1 = t.new_version(Date(2022, 10, 26)) # v1 是 fr/code/commerce/L723-4/20221026，由目标 t 创建

在 II 中，修改内容是参照经本条 I 修订的《商法典》L723-4条进行说明的，也就是说，在应用 I 中的修改之后。结合 III 中的信息（该部分指出，II 中的修改自 2023年1月1日起适用），必须在该日期创建一个新版本：

v2 = v1.new_版本(Date(2023, 1, 1)) # v2 是 fr/code/commerce/L723-4/20230101，由版本 v1 创建。

名称 s、t、v1 和 v2 表示对文档的抽象引用。正如我们将在下一节中看到的，这些引用可用于抽象地引用文本片段。例如，par 和 sen 方法可以生成函数；这些函数在求值时将分别返回对某个段落或句子的引用。因此，s.par("I") 会生成一个函数，该函数在求值时将返回对 2022-1348号法律第1条第 I 段的引用。这些生成器可以如下组合：

v1.par("II").par("2°").sen(1)
该表达式会生成一个函数，用于获取《商法典》L723-4条第二部分第2°项的第一句在2022年10月26日版本中的内容。

5 修正文本中的变更检测

识别修正文本中提及的目标文本版本后，Legistix 会检测变更，并将其转换为一系列函数以供应用。根据图1中的文本，Legistix 会生成图3所示的完整程序。

它会创建《商法典》L723-4条的两个新版本：v1自2022年10月26日起生效，纳入2022-1348号法律第1条第I段所述的变更（s.par("I")），并应用于截至2021年10月13日的《商法典》版本；v2自2023年1月1日起生效，纳入第II段所述的变更（s.par("II")）。请注意，第二个版本是在第一个版本的基础上创建的。prepend、replace、insert、suppress、replace_par这些方法是函数生成器，只要尚未调用apply_changes()方法，就不会由程序直接执行。

程序中使用的每种方法都将一种自然语言操作转换为程序操作。描述变更的每个函数均采用 action(source, target, what...) 的形式，其中 source 和 target 分别表示变更的起始文本片段和目标文本片段。参数 what 用于描述变更，例如 ("and", 2) 表示单词 "and" 的第二次出现，而 "rescue" 表示替换词。

6 结论与展望

我们从一个涵盖数十年内容的参考数据库着手，其中的整合工作此前已由人工完成。这种历史整合可作为衡量我们方法可靠性的参照。Legistix 工具能够自动完成此前由人工执行的 93% 的整合操作。我们将继续努力，争取达到 100%，并分析那些未能检测出的案例。其中一些案例即使对人类而言也存在歧义，因此与立法者 [7] 所期望的可机械化特性相矛盾。

下一步将把我们的工作扩展到 EUR-Lex [3] 数据库，该数据库收录欧盟法律，并以类似于法国法律的发布模式发布法规和指令。

第三，我们将提出一种专家系统 [5]，以协助立法者起草修正文本，并可在起草过程中直接生成修正规则的形式化表达（程序）

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

v1 = t.new_version(Date(2022年, 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°"), ("和", 2), "或"), v1.insert(pI.par("3"), v1.par("3°"), "司法", "救济， "),
v1.insert(pI.par("3"), v1.par("4°"), " 司法 ", " 救济， "), v1.suppress(pI.par("4"),
v1.par("4° bis"), ("是", 1)), v1.insert(pI.par("5"), v1.par("5°"), "资格", "和职责"),
v1.replace_par(pI.par("6"), v1.lastpar(), "II. - 也符合条件[...]")]) 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), "商业名录", "
national register of company[...]")]) v2 = db.add_version(v2.apply_
changes())
```

图3. Legistix 基于2022年10月24日颁布的2022-1348号法律第1条所编写的程序，刊载于官方公报 2022年10月25日。

文本的起草。这将建立一个良性循环，使修改的影响能够立即得到可视化，从而提升法律制定工作的效率和可靠性。

当达到 100% 的比例时，修正文本发布与合并版本可供使用之间将不再有延迟。合并文本将仅仅是从原始文本开始执行一系列程序所产生的副产品。这些程序可以与修正文本同时付诸表决，从而使“代码即法律 [10] ”这一格言在法律上成为现实。

致谢

感谢巴黎矿业学院的学生 Salomé Ouaknine，在她为期两个月、以自动合并为主题的研究实习期间提供帮助，尤其是在梳理未来结合正则表达式与机器学习的混合方法方面。

参考文献

[1]法国法律与行政信息局（DILA）（编）。2022年。以 *XML* 格式提供的数据，包含自 *1990* 年以来发表于《法兰西共和国官方公报》（*JORF*）的所有法律和法规。https://www.data.gouv.fr/fr/datasets/jorf-les-donnees-de-l-edition-lois-et-decrets-du-journal-officiel/ [2]法国法律与行政信息局（DILA）（编）。2022年。以 *XML* 格式提供的数据，包含经法国法律与行政信息局（*DILA*）整合的法国法律。https://www.data.gouv.fr/fr/datasets/legi-codes-lois-et-reglements-consolides/ [3]欧盟出版局（编）。2022年。*EUR-Lex*，获取欧盟法律。https://eur-lex.europa.eu

[4] 法国法律与行政信息局（DILA）（编）。2022年。*Légifrance*。https://www.legifrance.gouv.fr [5] Alain Delaët、Denis Merigoux 和 Aymeric Fromherz。2022年。将 Catala 转化为法律证明平台。载于*POPL 2022* 03447072 - 编程语言与法律。美国费城。https://hal.inria.fr/hal- [6] Samuel Fabrizi、Maria Iacono、Tesei Andrea 和 Lorenzo De Mattei。2022年。首次尝试实现法律文件自动整合：文本修改的可靠分类。载于第八届意大利计算语言学会议论文集。意大利米兰。http://ceur-ws.org/Vol-3033/paper26.pdf [7] Thierry-Xavier Girardot。2014年。获取法律：整合的重要性与挑战。《文献工作者——信息科学》51, 4（2014年），30–32。https://www.cairn.info/revue-documentaliste-sciences-de-l-information-2014-4-page-30.htm [8] 法国政府秘书处和法国国务委员会。2017年。立法起草指南。法国文献出版社。[9] Pierre Larrède 和 Georges-André Silber。2011年。法国法律数据分类体系：Légimobile 的实践。载于*iExpo2011*。https://www.gfii.fr/uploads/docs/un-plan-de-classement-des-donnees-juridiques-francaises-l-experience

[10] Lawrence Lessig。2000年。《代码即法律》。《哈佛杂志》（2000年）。http://harvardmagazine.com/2000/01/code-is-law.html [11] 法兰西共和国。2021年。*2021-1317*号法律第*1*条，*2021*年*10*月*11*日颁布，刊载于法国官方公报*10*月 *12, 2021* _ 000044183798。https://www.legifrance.gouv.fr/jorf/article [12] 2021 jo/JORFARTI 法兰西共和国。。《法国商法典》第*L*条——于*2021*年*10*月*13*日生效，依据*2021*年*10*月*11*日*2021-1317*号法律第*1*条所作的修订。*723 4* https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044191424/2021-10-13/ [13] 法兰西共和国。2022年。*2022-1348*号法律第*1*条，*2022*年*10*月*24*日颁布，刊载于*2022*年*10*月*25*日的法国官方公报。https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000046480974