

法律文書の自動解析

筑波大学 情報学群 情報科学類 産学間連携推進室
情報科学類 2 年 202212117 金子尚樹 [s2212117@u.tsukuba.ac.jp]

1 はじめに

法律の世界は、六法のような基本的な法律があり、それらが埋められない隙間をより詳細な規則を定める法令が埋め、それでも埋められない隙間を判例が埋めるような構造となっている。

法的な問題の解決を行う時、問題となっている行動が法的にどう分割されるのか、そしてそれぞれがどの法令や判例に対応しどのような判定になるのかをまとめ、最終的な結論を導き出す必要がある。問題の分割や結論を出す部分は依然として人が担当しなければならないところではあるが、細かい法令や判例を読みながら分割された細かい問題に対する判断をしていく過程をソフトウェアで支援することは可能である。細かい手続きを定めた法令は実務上とても重要でありながらその条文は長く複雑であることが多く、ソフトウェアによって自動で適切な注釈を付けることができれば、法務業務の負担軽減に繋がり社会的な意義があると考えている。

本プロジェクトでは、ソフトウェアによる支援が特に効果的であるとする

- 準用・読み替え規定文
- 略称・定義規定文
- 他の法令・条項への参照
- 判例での上記項目の解析

の4つについて取り組んできた。法令を解析して構造化する研究は様々あるが、本研究では特定の項目についてのみ解析を行い、必要な情報を抽出することに主眼を置いた。これにより、多く存在する例外への対処をしやすくなり、解析の対象としない項目については無視をすることでスムーズな検証が行えるようになった。

2 読み替え規定文の解析

2.1 背景

法律文書には準用と読み替えという概念が存在する。これは、一つの条文を使いまわしつつ必要最小限の修正を加えて、設置の目的などが似ている複数の規定を定義する手法である。使いまわす手法が準用と呼ばれ、準用の際に加える必要最小限の修正が読み替えである。図1のように、まず規定Aとそれを記述する条項Aが存在する。規定Aに似た新しい規定A'ができたとき、条項Aを読み替え規定で修正しながら規定A'の記述を行う。これが準用である。

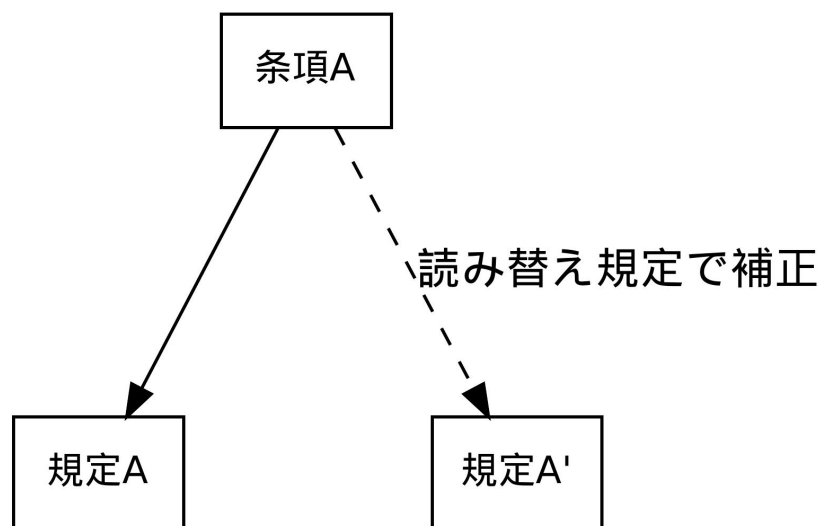


図 1：準用と読み替えの概念図

具体例を以下に挙げる。医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律（昭和三十五年法律第百四十五号）の第七条第四項では薬局の管理について以下のように定めている。

薬局の管理者（第一項の規定により薬局を実地に管理する薬局開設者を含む。次条第一項及び第三項において同じ。）は、その薬局以外の場所で業として薬局の管理その他薬事に関する実務に従事する者であつてはならない。ただし、その薬局の所在地の都道府県知事の許可を受けたときは、この限りでない。

そして同法律の第八条第一項では薬局の管理者の義務について以下のように定めている。

薬局の管理者は、保健衛生上支障を生ずるおそれがないように、その薬局に勤務する薬剤師その他の従業者を監督し、その薬局の構造設備及び医薬品その他の物品を管理し、その他その薬局の業務につき、必要な注意をしなければならない。

そして同法律の第十七条第八項では医薬品等総括製造販売責任者について以下のように定めている。

医薬品製造管理者については、第七条第四項及び第八条第一項の規定を準用する。この場合において、第七条第四項中「その薬局の所在地の都道府県知事」とあるのは、「厚生労働大臣」と読み替えるものとする。

薬局の管理について規定している第七条第四項と薬局の管理者の義務について規定している第八条第一項を医薬品製造管理者について準用することで、語句が少し違うだけの条文を再度書くコストを減らしている。この条文を読む人は「医薬品製造管理者について準用する」という文言をもとに、第七条第四項と第八条第一項内に登場する「薬局の管理者」という語を「医薬品製造管理者」に置換し、第十七条第八項の条文が意図する文章を考えて読むこととなる。ただし、薬局に許可を出すのが都道府県知事であるのに対して、医薬品製造所に対して許可を出すのが厚生労働大臣であるという制度上の差異があり、これは非自明であるため条文中で読み替えの指示を行っている。このため、第十七条第八項が意図する条文は以下のようなものになる。

医薬品製造管理者は、その医薬品製造所以外の場所で業として医薬品製造所の管理その他薬事に関する実務に従事する者であつてはならない。ただし、その医薬品製造所の所在地の厚生労働大臣の許可を受けたときは、この限りでない。

医薬品製造管理者は、保健衛生上支障を生ずるおそれがないように、その医薬品製造所に勤務する薬剤師その他の従業者を監督し、その医薬品製造所の構造設備及び医薬品その他の物品を管理し、その他その医薬品製造所の業務につき、必要な注意をしなければならない。

元の条文が以下の物であったことを考えると、文章量の圧縮の効果があることがわかる。

医薬品製造管理者については、第七条第四項及び第八条第一項の規定を準用する。この場合において、第七条第四項中「その薬局の所在地の都道府県知事」とあるのは、「厚生労働大臣」と読み替えるものとする。

準用と読み替え規定は文章量の圧縮の他に「似ている条文をまとめ、似ている中での相違点を明確にする」という役割も存在する。これにより、過去の違う条文に関する判例をどこまで適用することができるという判断が適切にできるようになる。

このように準用と読み替え規定は便利な物であるが、読み替え規定は単語の置き換えを自然言語で指示するという特性上、

- 登場する文が不自然になる
- 置き換える単語によっては一文がとても長くなる

という難しさが存在する。保険商品の難しさについて調査を行った竹井らも保険約款などの法律文書の内容の解釈が困難になっている原因の一つに読み替え規定の多用があると指摘している [1]。

複雑な条文の具体例を挙げる。租税特別措置法（昭和三十二年法律第二十六号）の第七十条の七の四第十三項は以下のような条文である。

第七十条の七の二第二十二項から第二十六項までの規定は、認定相続承継会社について同条第二十二項に規定する評定が行われた場合における納税猶予分の相続税額の計算及び免除について準用する。この場合において、同項から同条第二十五項までの規定中「経営承継期間」とあるのは「経営相続承継期間（第七十条の七の四第一項の規定の適用を受ける経営相続承継受贈者に係る贈与者が同条第二項第五号イ又はロに掲げる日のいずれか早い日の翌日以後に死亡した場合にあつては、当該経営相続承継受贈者に係る前条第二項第六号に規定する経営贈与承継期間）」と、「第一項」とあるの

は「、第七十条の七の四第一項」と、「対象非上場株式等に」とあるのは「対象相続非上場株式等に」と、「認定承継会社」とあるのは「認定相続承継会社」と、「経営承継相続人等」とあるのは「経営相続承継受贈者」と、「対象非上場株式等（）」とあるのは「対象相続非上場株式等（）」と、「相続により取得をした対象非上場株式等の当該相続の時ににおける」とあるのは「対象相続非上場株式等の」と、同条第二十三項中「を第一項」とあるのは「を同条第一項」と、「第二項第五号」とあるのは「同条第二項第四号」と読み替えるものとする。

これを読み替えられる語と読み替える語で対応が取れるように色分けをすると次のようになる。

第七十条の七の二第二十二項から第二十六項までの規定は、認定相続承継会社について同条第二十二項に規定する評定が行われた場合における納税猶予分の相続税額の計算及び免除について準用する。この場合において、同項から同条第二十五項までの規定中「経営承継期間」とあるのは「経営相続承継期間（第七十条の七の四第一項の規定の適用を受ける経営相続承継受贈者に係る贈与者が同条第二項第五号イ又はロに掲げる日のいずれか早い日の翌日以後に死亡した場合にあつては、当該経営相続承継受贈者に係る前条第二項第六号に規定する経営贈与承継期間）」と、「、第一項」とあるのは「、第七十条の七の四第一項」と、「対象非上場株式等に」とあるのは「対象相続非上場株式等に」と、「認定承継会社」とあるのは「認定相続承継会社」と、「経営承継相続人等」とあるのは「経営相続承継受贈者」と、「対象非上場株式等（）」とあるのは「対象相続非上場株式等（）」と、「相続により取得をした対象非上場株式等の当該相続の時ににおける」とあるのは「対象相続非上場株式等の」と、同条第二十三項中「を第一項」とあるのは「を同条第一項」と、「第二項第五号」とあるのは「同条第二項第四号」と読み替えるものとする。

これからわかるように、この長い条文では

- ・「経営承継期間」→「経営相続承継期間（第七十条の七の四第一項の規定の適用を受ける経営相続承継受贈者に係る贈与者が同条第二項第五号イ又はロに掲げる日のいずれか早い日の翌日以後に死亡した場合にあつては、当該経営相続承継受贈者に係る前条第二項第六号に規定する経営贈与承継期間）」
- ・「、第一項」→「、第七十条の七の四第一項」
- ・「対象非上場株式等に」→「対象相続非上場株式等に」
- ・「認定承継会社」→「認定相続承継会社」
- ・「経営承継相続人等」→「経営相続承継受贈者」
- ・「対象非上場株式等（）」→「対象相続非上場株式等（）」
- ・「相続により取得をした対象非上場株式等の当該相続の時ににおける」→「対象相続非上場株式等の」
- ・「を第一項」→「を同条第一項」
- ・「第二項第五号」→「同条第二項第四号」

の9個の読み替えが規定されており、とても長い文や読点から始まる読み替えが存在することがわかる。これを誤読せずに読むことはとても難しく、機械的な抽出が可能にな

ると人への負担を大きく軽減することができる効果が望める。

法律文書の構造化を行っている先行研究は複数あるが、信岡らや木村らの提案手法の中では読み替え規定文の解析を行っておらず、今後の課題とされている [2, 3]。

2.2 手法

読み替え規定文の書き方について磯崎は「この場合において、「A」とあるのは「B」と読み替えるものとする。」となることが基本形であると指摘している [4]。しかし、より詳細に読み替え規定文の構造を観察したところ、

1. まず (a)・(b) のパターンが 0 回以上繰り返される

(a) “「～～」とあり” という表記が 0 回以上繰り返される

(b) その後 “「～～」とあるのは「～～」と、” という表記が出現する

2. 次に “「～～」とあり” という表記が 0 回以上繰り返される

3. 最後に “「～～」とあるのは「～～」と読み替える” という表記が出現するという規則で記述されていることが分かった。このうち、

- 「～～」とあり

- 「～～」とあるのは

と書かれている場合の鍵括弧の中の用語は読み替えられる語であり、

- 「～～」と、

- 「～～」と読み替える

と書かれている場合の鍵括弧の中の用語は読み替える語である。

「～～」とあり」と書く条文の具体例を挙げる。学校教育法（昭和二十二年法律第二十六号）第十三条第二項には以下のようになっている。

前項の規定は、市町村の設置する幼稚園に準用する。この場合において、同項中「それぞれ同項各号に定める者」とあり、及び同項第二号中「その者」とあるのは、「都道府県教育委員会」と読み替えるものとする。

このとき、「それぞれ同項各号に定める者」と「その者」の 2 つの語を共に「都道府県教育委員会」という語に置き換えるという指示を行っている。

このルールをもとにすると、鍵括弧の中の文字列を保持しつつ鍵括弧の直後に出現する文字が

- とあり

- とあるのは

- と、

- と読み替える

のどれであるかによってその鍵括弧の中の文字列が読み替えられる語か読み替える語かを判定することができ、読み替え規定文からそれらの情報を自動で抽出することができることがわかる。

2.3 ソフトウェアの実装

手慣れていることやすでに自身が作成した法令解析用のライブラリを利用するために、ソフトウェアの実装は Rust 言語で行った。作成したソフトウェアは MIT ライセンスの元 GitHub で公開している^{*1}。

入力は e-gov 法令検索 [5] で公開されている法令 XML データとし、出力は

- 出現箇所
- 読み替えられる語
- 読み替える語

の情報が記述された JSON 形式となる。

2.4 結果

2022 年 12 月 4 日時点で公開されていた法令データをもとに解析した。読み替え規定文 11672 個に対して提案する手法を用いて解析したところ、11428 個は抽出に成功し、残りの 244 個は失敗した。

解析結果については CC0 ライセンスの元 GitHub で公開している^{*2}。

2.5 まとめ

解析に失敗した条文を調べたところ、大きく分けて

- 文中の括弧の対応が壊れているもの
- 表形式で与えているもの
- 読み替える語等の指定に鍵括弧をしていないもの

の 3 つに分けられることがわかった。

文中の括弧の対応が壊れているものの例として、農業委員会等に関する法律（昭和二十六年法律第八十八号）の平成二七年九月四日法律第六三号附則の第三十九条を挙げる。

組織変更については、附則第十三条第二項及び第八項、第十六条並びに第十七条の規定を準用する。この場合において、附則第十三条第二項中「前項」とあるのは「附則第三十七条第一項」と、「旧農協法第七十三条の四十三第二項」とあるのは「第二条の規定による改正前の農業委員会等に関する法律第七十六条」と、同条第八項中「第四十九条並びに」とあるのは「第四十九条第一項及び第二項（第二号を除く。）並びに」と、「内容」とあるのは」とあるのは「内容」とあるのは、と、「。次項において「改正法」という。）附則第十三条第一項」とあるのは「附則第三十七条第一項」と、「組織変更をする旨」と、同項第二号中「計算書類」とあるのは「財産目録」と、同条第三項中「第九十七条の四第二項」とあるのは「改正法附則第十条の規定によりなおその効力を有するものとされた改正法第一条の規定による改正前の農業協同組合法第九十二条第二項」とあるのは「組織変更をする旨」と、附則第十七条中「附則第十二条から前条まで」と

^{*1} https://github.com/japanese-law-analysis/analysis_yomikae

^{*2} https://github.com/japanese-law-analysis/data_set/blob/master/law/yomikae.json

あるのは「附則第三十六条から第三十八条まで並びに附則第三十九条において読み替えて準用する附則第十三条第二項及び第八項並びに前条」と読み替えるものとする。

下線を引いたところのように、読み替えられる語と読み替える語に閉じ鍵括弧だけが含まれていることがわかる。

表形式で与えているものの例としては信託業法施行規則（平成十六年内閣府令第百七号）第五十一条の九の一部を挙げる。

法第五十条の二第一項の登録を受けた者については信託会社（第二十三条第二項及び第三項並びに第二十五条にあっては、管理型信託会社）とみなして、第八条、第十七条から第二十三条まで、第二十五条、第二十九条、第三十七条から第四十一条の八まで、第四十八条（第一項第三号、第七号及び第十号から第十二号まで並びに第二項を除く。）、第五十条（第四項を除く。）及び第五十一条の規定を適用する。この場合において、これらの規定中「信託業務」とあり、及び「信託業」とあるのは、「信託法第三条第三号に掲げる方法によってする信託に係る事務」とするほか、次の表の上欄に掲げる規定中同表中欄に掲げる字句は、同表下欄に掲げる字句と読み替えるものとする。

第二十三条第二項	本店	信託法第三条第三号に掲げる方法によってする信託に係る事務を行う主たる営業所
第二十三条第二項及び第三項	管理型信託会社登録簿	自己信託登録簿
第二十五条	業務方法書	信託法第三条第三号に掲げる方法によってする信託に係る事務の内容及び方法を記載した書類
第二十九条第一号	委託者又は受益者（これらの者から指図の権限の委託を受けた者を含む。）	受益者（当該者から指図の権限の委託を受けた者を含む。）

下線を引いた文言とその下にある表のように読み替え規則が与えられていることがわかる。

読み替える語等の指定に鍵括弧をしていないものの例としては消防法（昭和二十三年法律第百八十六号）第三十七条を挙げる。

特別区の存する区域においては、この法律中 市町村、市町村長又は市町村条例 とあるのは、^{それぞれ}夫々これを 都、都知事又は都条例 と読み替えるものとする。

下線を引いた箇所が読み替えられる語と読み替える語に該当するが、他の条文と違い鍵括弧による指定が無いことがわかる。

このように単語の出現するパターンのみで読み替え規定文のうち 97% 以上の読み替え規定文を解析することができた。今後の課題としては、括弧の対応がとれていない条文や表形式で与えられている条文の自動検出と表の解析機能を実装することである。また、後述する法令名の自動抽出機能と組み合わせることで、読み替え規定文適用後の条文の自動生成を行えるようになりたいと考えている。

3 略称・定義規定文の解析

3.1 背景

法律文書では文字で書き起こすと長くなる概念が頻繁に登場する。これを毎回書いていると文章量がとても長くなってしまい、読みにくくなる。そこで、それらに対して略称を与えてこの問題を解決することがある。例えば、航空法施行規則（昭和二十七年運輸省令第五十六号）の第二百三十六条の七十一第一項第四号では以下のような条文が書かれている。

国土交通省、防衛省、警察庁、都道府県警察又は地方公共団体の消防機関その他の関係機関の使用する航空機のうち搜索、救助その他の緊急用務を行う航空機の飛行の安全を確保する必要があるものとして国土交通大臣が指定する空域（以下「緊急用務空域」という。）

この例では、「国土交通省、防衛省、警察庁、都道府県警察又は地方公共団体の消防機関その他の関係機関の使用する航空機のうち搜索、救助その他の緊急用務を行う航空機の飛行の安全を確保する必要があるものとして国土交通大臣が指定する空域」という概念に対して、「緊急用務空域」という略称を定義している。そしてこの条文の直後である同条第二項では以下のような条文が書かれており、定義された略称が使われていることがわかる（下線筆者）。

国土交通大臣は、前項第四号の規定による指定をしたときは、インターネットの利用その他の適切な方法により、その旨及び当該指定に係る 緊急用務空域 を公示しなければならない。

もしこのような略称が定義できなければ、上の条文は

国土交通大臣は、前項第四号の規定による指定をしたときは、インターネットの利用その他の適切な方法により、その旨及び当該指定に係る 国土交通省、防衛省、警察庁、都道府県警察又は地方公共団体の消防機関その他の関係機関の使用する航空機のうち搜索、救助その他の緊急用務を行う航空機の飛行の安全を確保する必要があるものとして国土交通大臣が指定する空域 を公示しなければならない。

と書かねばならず、とても読みにくくなってしまうことがわかる。

このように略称はとても便利なものであるが、略称の元の定義を知る必要がある場合も出てくる。例えば、

国土交通大臣は、前項第四号の規定による指定をしたときは、インターネットの利用その他の適切な方法により、その旨及び当該指定に係る 緊急用務空域 を公示しなければならない。

という条文を読んでいるときに「緊急用務空域は誰が指定するのか」という情報も知る必要が出ると、「緊急用空域」という語の定義に戻る必要がある。このとき、自動で定義を提示することができるのと一々語を定義する条文に戻る必要が無くなり、便利である。

略称の定義文の自動抽出を行う手法は既に中村らによって提案されている [6]。この

論文では法律文コーパスを用いて解析を行う方法を提案している。しかし、この手法では登場する語がコーパスに登録されているものでないといけないため、定義された略称が登場する場合には対応できない。そのため

- 条文中に複数の定義文が登場する場合、先頭の物しか解析できない
- 括弧書き内の定義文は解析できない
- 条文の先頭から 100 字以内に書かれたものしか解析できない

という強い制約が存在している。そこで、より汎用的に略称・定義規定文の解析を行うための方法を模索することにした。

3.2 手法

略称・定義規定文は大きく分けて 2 つの形に分けられる。中村らの論文ではこれらをトイウ形・フィウ形と分類しており、本稿でもこの呼び名を採用する [6]。

トイウ形は「A、B 及び C (以下「〇〇〇」という。)」という形で書かれるものである。このとき、「A、B 及び C」が正式名称で「〇〇〇」が略称となる。具体例としては会社法（平成十七年法律第八十六号）第二十二條の以下のものを挙げる。

事業を譲り受けた会社（以下この章において「譲受会社」という。）が譲渡会社の商号を引き続き使用する場合には、その譲受会社も、譲渡会社の事業によって生じた債務を弁済する責任を負う。

この例では正式名称が「事業を譲り受けた会社」で、「譲受会社」が略称である。

フィウ形は「〇〇〇（A、B 及び C をいう。以下同じ。）」という形で書かれるものである。このとき、「A、B 及び C」が正式名称で「〇〇〇」が略称となる。具体例としては会社法（平成十七年法律第八十六号）第二百五十五條第二項の以下のものを挙げる。

証券発行新株予約権付社債に付された新株予約権の譲渡は、当該証券発行新株予約権付社債に係る新株予約権付社債券を交付しなければ、その効力を生じない。ただし、自己新株予約権付社債（株式会社が有する自己の新株予約権付社債をいう。以下この条及び次条において同じ。）の処分による当該自己新株予約権付社債に付された新株予約権の譲渡については、この限りでない。

この例では正式名称が「株式会社が有する自己の新株予約権付社債」で、「自己新株予約権付社債」が略称である。

トイウ形の略称とフィウ形の正式名称は正規表現で簡単に抽出することができるが、トイウ形の正式名称とフィウ形の略称部分の判定と抽出は自明ではない。特に正式名称部分に読点が含まれる場合や文頭から始まらない場合があり、これらの考慮が難しくなっている。例えば、青少年が安全に安心してインターネットを利用できる環境の整備等に関する法律（平成二十年法律第七十九号）第三条第一項は以下のような条文である。

青少年が安全に安心してインターネットを利用できるようにするための施策は、青少年自らが、主体的に情報通信機器を使い、インターネットにおいて流通する情報を適切に取捨選択して利用するとともに、適切にインターネットによる情報発信を行う能力（以下「インターネットを適切に活用する能力」という。）を習得することを旨として行われなければならない。

この条文において「インターネットを適切に活用する能力」という略称に対応する正式名称は「青少年自らが、主体的に情報通信機器を使い、インターネットにおいて流通する情報を適切に取捨選択して利用するとともに、適切にインターネットによる情報発信を行う能力」である。このようなものを正規表現で抽出することは不可能である。

略称・定義規定文を含む条文の構造は「○○○ △△△ (〜〜) ×××」のような形になっているとすることができる。トイウ形では「△△△」が正式名称で「〜〜」が略称に、ヲイウ形では「△△△」が略称で「〜〜」が正式名称にあたる。今解決しなければならないのは、「○○○」と「△△△」の境界がどこにあるのかを自動で判定することである。

本研究では係り受け解析を行うことでこの境界を自動で判定する手法を考案し、試した。条文のテキストを係り受け解析することで得られる、「どの単語がどの単語に係っているか」という関係の有向グラフを解析し

1. 括弧書き（「(〜〜)」の部分）の直前の単語に係る語の最長部分
2. 括弧書き（「(〜〜)」の部分）より後に書かれている語（「×××」の部分）に係っている単語が一切出現しない連続する部分

の2つの情報を得る。求める略称・正式名称に対して、1番目の情報は「この部分の中に求める分は含まれる」という部分で2番目の情報は「この部分は求める分に含まれる」という部分になっていると考えた。そこで、1番目の部分が句読点を含まない場合には1番目の部分を採用し、句読点を含む場合には2番目の部分を採用するというルールで抽出を行ってみた。これは、文の区切りが存在しない場合には考えられる中で最長な部分を採用すると確からしいだろうという予想に基づくものである。

3.3 ソフトウェアの実装

日本語の係り受け解析部分の実装はPythonで行った。Pythonの自然言語処理ライブラリであるspaCyに基づいて作成された日本語用の自然言語処理ライブラリであるGiNZA^{*3}を使用するためである。これは事前学習モデルを導入した高性能な自然言語処理ライブラリである[7]。このライブラリを使用して条文を文節ごとにくぎり、その文節がどこに係っているのかの情報を取得し、JSON形式で吐き出すソフトウェアを作成した。これはMITライセンスの元GitHubで公開している^{*4}。

上記の係り受け解析を行うソフトウェアに対して渡す条文のテキストの生成や、係り受け解析の結果を元に略称・正式名称を取り出すソフトウェアはRustで実装した。これはMITライセンスの元GitHubで公開している^{*5}。

3.4 結果

全法令のデータに対して解析を行ったところ、128100個の定義を抽出することがで

^{*3} <https://github.com/megagonlabs/ginza>

^{*4} https://github.com/japanese-law-analysis/parse_japanese_dependency

^{*5} https://github.com/japanese-law-analysis/analysis_ryakusyou

きた。

生成された略称・正式名称と出現箇所の情報は JSON 形式で、CC0 ライセンスの元 GitHub で公開している*6。

3.5 まとめ

概ね定義文を抽出することができたものの、略称・正式名称の判定が間違っている箇所が少なからずあった。例えば、上に挙げた

青少年が安全に安心してインターネットを利用できるようにするための施策は、青少年自らが、主体的に情報通信機器を使い、インターネットにおいて流通する情報を適切に取捨選択して利用するとともに、適切にインターネットによる情報発信を行う能力（以下「インターネットを適切に活用する能力」という。）を習得することを旨として行われなければならない。

という文言に対しては、略称の「インターネットを適切に活用する能力」は正しく抽出できているものの、正式名称の部分は「自らが、主体的に情報通信機器を使い、インターネットにおいて流通する情報を適切に取捨選択して利用するとともに、適切にインターネットによる情報発信を行う能力」と判定して抽出してしまっており、先頭の「青少年」という文言が抜けている。

係り受け解析の結果をもとにした解析の精度は高そうであるが、精度をあげる余地がかなりあると感じている。品詞の情報も考慮に加えたり、係り受け解析の際に使用する辞書を法令文書用のものにしたことで、精度を上げられるのではないかと考えている。

4 法令名の自動抽出

4.1 背景

法令や判例からは他の法令やその中の条項を参照する事がとても多い。これは、法令では他の法令の規定の補足や条件の上書きなどを行い、判例では論拠を積み上げていく過程で実際の法令の条文に当たる必要があるからである。参照先の条文の内容を自動で取得できたり、条文間の依存関係を明らかにして影響範囲を調べることができたりするなど、条項への参照を抽出することはとても有意義である。

条項についての参照表現の抽出方法については木村によって分類され、形式的な抽出が可能なが示されている [3]。法令名および法令番号の自動取得についてはハッシュ法を用いた方法が関本によって提案されている [8]。

しかし、法令名については条文中で定義された略称によって参照されることもある。例えば郵政民営化法（平成十七年法律第九十七号）の第八十三条では冒頭では以下のような記述がある。

*6 https://github.com/japanese-law-analysis/data_set/blob/master/law/ryakusyou.json

郵便局株式会社の成立の際現に公社が郵政民営化法等の施行に伴う関係法律の整備等に関する法律（平成十七年法律第百二号。以下「整備法」という。）

「郵便局株式会社の成立の際現に公社が郵政民営化法等の施行に伴う関係法律の整備等に関する法律」という長い名前の法律に対して「整備法」という略称を定義している。実際に、そのあとの条文である第百七条第一項の口では「整備法附則第五条第一項の規定により」のように、「整備法」という略称が使われて条文を参照している。正確な参照を分析するためにはこの略称も含めた解析ができるようにする必要がある。今回は、これを行うための前処理として法令中で定義されている略称を自動で抽出することとした。

また、国が「略称法令名一覧」という web ページで事前に定義された略称の一覧を公開している（図2）^{*7}。ここに定義された略称は断りなく使用されることが考えられるため、これの一覧も取得する必要がある。

4.2 手法

略称の定義文は例にあげたように「＜法令の正式名称＞（＜法令番号＞。以下＜条件文＞において「＜略称＞」という。）」という形式で、

- 法令の正式名称
- 法令番号
- 条件文
- 略称

の4つの情報を含んで定義される。このとき、条件文とは例えば「以下次項及び第六項において」のような、略称が使われる条項を制限する文である。この文が書かれない場合もあり、その時は定義された後、同じ法令内であればいつでも使えるようになる。このうち、法令番号がわかれば法令の正式名称は自動で割り出すことができるため、抽出しなければならない情報は

- 法令番号
- 条件文
- 略称

の3つとなる。この3つの情報を含んだ文を表す

1 ((明治|大正|昭和|平成|令和)(一|二|三|四|五|六|七|八|九|十|〇)+年[^()、。↵
↵ あ-ん]+第(一|二|三|四|五|六|七|八|九|十|百|千|〇)+号。?(以下)?([^(↵
↵]+において)?、?「[^']*」*(法|令)」という。

という正規表現を作成し、これにマッチする文から上記の3つの情報を抜き出すこととした。

「略称法令名一覧」の web ページからのデータを抜き出す際には、web ページにアクセスし HTML データを解析することで情報を得ることとした。

^{*7} <https://elaws.e-gov.jp/abb/>

略称法令名一覧

登録略称法令名一覧

正式法令名	法令番号	略称法令名 1	略称法令名 2	略称法令名 3	略称法令名 4
明治五年太政官布告第二百三十七号（改暦ノ布告）	明治五年太政官布告第二百三十七号	改暦の布告			
明治八年太政官布告第五十四号（勲章制定ノ件）	明治八年太政官布告第五十四号	勲章制定の件			
明治八年太政官達第五十二号（不用物品等払下ノトモ官署ノ官吏入札禁止ノ件）	明治八年太政官達第五十二号	不用物品等払下ノトモ官署ノ官吏入札禁止の件			
明治十年太政官達第九十七号（大勲位菊花大綬章及副章製式ノ件）	明治十年太政官達第九十七号	大勲位菊花大綬章及副章製式の件			
明治十九年勅令第五十一号（本初子午線経度計算方及標準時ノ件）	明治十九年勅令第五十一号	本初子午線経度計算方及標準時の件			
行旅病人及行旅死尸人取扱法	明治三十二年法律第九十二号	行旅法			
明治三十二年法律第四十号（失火ノ責任ニ關スル法律）	明治三十二年法律第四十号	失火法	失火責任法		
明治三十二年勅令第二百三十七号（行旅病人死尸人等の引取及費用）	明治三十二年勅令第二百三十七号	行旅病人死尸人等の引取及費用			

図 2：略称法令名一覧という web ページ

4.3 ソフトウェアの実装

手慣れていることやすでに自身が作成した法令解析用のライブラリを利用することから、ソフトウェアの実装は Rust 言語で行った。作成したソフトウェアは MIT ライセンスの元 GitHub で公開している^{*8}。

「略称法令名一覧」ページのスクレイピングソフトウェアも Rust で実装した。作成したソフトウェアは MIT ライセンスの元 GitHub で公開している^{*9}。

4.4 結果

現在施工されている法令中全体に対して解析をかけたところ、3504 個の法令から合わせて 11193 の略称の定義を抽出することができた。抽出された略称一覧の JSON ファイルは CC0 ライセンスの元 GitHub で公開している^{*10}

「略称法令名一覧」の web ページで公開されているものについては、2412 個の法令について合わせて 3276 個の略称が定義されていた。抽出された略称一覧の JSON ファイルは CC0 ライセンスの元 GitHub で公開している^{*11}

*8 https://github.com/japanese-law-analysis/analysis_law_reference

*9 https://github.com/japanese-law-analysis/egov_law_abb

*10 https://github.com/japanese-law-analysis/data_set/blob/master/law/law_abb.json

*11 https://github.com/japanese-law-analysis/data_set/blob/master/law/egov_abb.json

4.5 まとめ

略称定義文を正規表現を使って解析することにより、合計で 1.1 万を超える略称のリストを生成することができた。この数は国が公開している略称一覧よりも 3 倍ほど多い。これにより、既に提案されている法令名抽出方法の手法や Aho-Corasick 法などの既知の効率的な文字列探索アルゴリズムを使用することで、より精密な法令名の出現箇所の解析を行うことができると期待される。

ただし、条項を含めたものについて略称を定義している場合の対応ができていないため、これについてどう略称情報を取得してデータを保持するのかを検討する必要がある。条項を含めたものについての略称とは、例えば恩給法（大正十二年法律第四十八号）の第二十四条の三第一項では「旧恩給法の特例に関する件の措置に関する法律による改正前の旧勅令第六十八号第八条第一項（以下「改正前の旧勅令第六十八号第八条第一項」という。）に規定する」と書かれており、法令名に加え「第八条第一項」まで指定したうえで略称を定義している。

5 判例の自動収集およびテキストデータ化

5.1 背景

法律分野の文書の中では法令だけではなく判例も重要である。法令が埋められない穴を判例によって埋めていき、次の類似事例の時の判断に活かされるからである。そのため、判例の解析も法令と同様に行う必要がある。

現状、判例のテキストがオープンデータとして公開されているものが無いため、作成することとした。

5.2 手法

裁判所は「裁判例検索」という Web サービスを提供している^{*12}。このサービスは単語検索や裁判年月日、事件番号等で該当する裁判例を検索し、その裁判の

- 事件番号
- 法廷名
- 種別
- 判示事項
- 裁判要旨
- 判決文全文

などが記載された個別の詳細ページへのリンクを取得することができるサービスである。ここで、判決文全文は PDF ファイルであるが、それ以外は単純な HTML によって提供される。

判例の信頼性と構築の容易さから、今回はこの裁判例検索サービスを利用して判例

^{*12} https://www.courts.go.jp/app/hanrei_jp/search1

データベースを作成することとした。

情報の抽出は主にスクレイピング技術を使用することで行った。まず、このサービスに対して取得したい期間を表す検索クエリが含まれた URL を自動で生成してアクセスし、詳細ページのリンクのリストを取得する。その後、詳細ページの HTML を解析して必要な情報を抜き出し、判決文全文の PDF ファイルをダウンロードする。最後に、その PDF ファイルからテキストを抽出することとした。

5.3 ソフトウェアの作成

並列並行計算が高速で安全に行うことができ、自身の手になじんでいるため、スクレイピングソフトウェアを Rust で実装した。作成したソフトウェアは MIT ライセンスの元 GitHub で公開している^{*13}。判例の一覧は JSON 形式で出力される。

判例 PDF ファイルからテキストの抽出を行う方法は 2 通り用意した。それぞれ

- poppler ライブラリ^{*14}が提供する pdftotext コマンドを用いてテキストを抽出した後に、ページ番号や行番号に相当する部分を削除する
- PDF ファイルを画像ファイルに変換した後に、本文テキストが書かれた部分のみを切り出し、tesseract^{*15}という OCR ソフトウェアによってテキストを抽出する

という方法である。作成したソフトウェアは MIT ライセンスの元 GitHub で公開している^{*16}。

5.4 結果

判例データの収集は公開されている中で最も古い 1926 年から実験を行った 2023 年 1 月 15 日までの期間で行った。その結果約 7 万件の判例データを収集することができた。生成された判例に関するデータは CC0 ライセンスの元 GitHub で公開している^{*17}。

PDF ファイルからテキストを抽出する手法は、pdftotext コマンドを使用する方が OCR を行うよりも精度が高かったため、pdftotext を使用する方法を採用した。生成されたテキストファイルは CC0 ライセンスの元 GitHub で公開している^{*18}。ただし、文字の抜け

^{*13} https://github.com/japanese-law-analysis/listup_precedent

^{*14} <https://poppler.freedesktop.org/>

^{*15} <https://github.com/tesseract-ocr/tesseract>

^{*16} https://github.com/japanese-law-analysis/pdf2txt_precedent

^{*17} https://github.com/japanese-law-analysis/data_set/blob/master/precedent/list.json

^{*18} https://github.com/japanese-law-analysis/data_set/tree/master/precedent/text_files

や行の順番のズレなどが存在することを確認しており、よりテキスト抽出の精度を上げる必要があると考えている。

5.5 発見した誤記・例外について

この収集および解析の過程で年月日にマイナスの値が出現している詳細ページが以下の7つ存在することを確認した（図3）。

- 昭和 46(あ)1051 ^{*19}
- 昭和 39(オ)1302 ^{*20}
- 昭和 35(オ)535 ^{*21}
- 昭和 36(オ)513 ^{*22}
- 昭和 36(オ)536 ^{*23}
- 昭和 34(オ)71 ^{*24}
- 昭和 24(れ)2055 ^{*25}

発見した翌日である 2023 年 1 月 16 日に最高裁判所に報告を行ったところその数日後に全てのページが修正された。例えば昭和 39(オ)1302 という判例では

- 修正前：https://web.archive.org/web/20230116041527/https://www.courts.go.jp/app/hanrei_jp/detail2?id=66298

*19 http://web.archive.org/web/20220606192511/https://www.courts.go.jp/app/hanrei_jp/detail2?id=51011

*20 https://web.archive.org/web/20230115065439/https://www.courts.go.jp/app/hanrei_jp/detail2?id=66298

*21 https://web.archive.org/web/20230115072821/https://www.courts.go.jp/app/hanrei_jp/detail2?id=63595

*22 https://web.archive.org/web/20230115073312/https://www.courts.go.jp/app/hanrei_jp/detail2?id=63715

*23 https://web.archive.org/web/20230118092926/https://www.courts.go.jp/app/hanrei_jp/detail2?id=63719

*24 http://web.archive.org/web/20200804085713/https://www.courts.go.jp/app/hanrei_jp/detail2?id=70761

*25 http://web.archive.org/web/20230118094001/https://www.courts.go.jp/app/hanrei_jp/detail2?id=70761

- 修正後： https://web.archive.org/web/20230118092109/https://www.courts.go.jp/app/hanrei_jp/detail2?id=66298

がそれぞれ URL のページであるが、タイムスタンプを確認すると修正前が 2023 年 1 月 16 日で修正後が 2023 年 1 月 18 日となっており、報告を受けた翌日に修正が行われていることがわかる。

また、裁判例に一意に割り振られる事件番号というものが割り振られていないで空白になっている事件が 2 つ存在することを確認した。それぞれ

- 昭和 23 年 10 月 13 日判決の「市税滞納処分取消請求事件」^{*26}^{*27}
- 平成 17 年 6 月 20 日判決の「新株予約権発行差止仮処分決定認可決定に対する保全抗告 (付更正決定)」^{*28}

である。事件番号がついていないことについて最高裁判所に問い合わせたところ、「何らかの事情によってつかないこともあり、それではないか」との回答を得た。

5.6 まとめ

公式に公開されている判例集から判例データベースを自動で構築することができるソフトウェアを作成し、実際に構築した。ただし、判決文全文を PDF ファイルから抽出する過程の精度が良くないため、改善の余地がある。

裁判所の公開している情報に誤りがあることを発見して最高裁判所に報告したところ、修正された。また、事件番号が振られていないものが存在するという例外事象について確認することができた。

6 おわりに

読み替え規定文の解析については情報処理学会第 85 回全国大会に論文を出し講演を行った [9]。また、優秀な発表であると認められ学生奨励賞^{*29}および大会奨励賞^{*30}を受賞することができた。

略称・定義規定文の解析はもっと改良の余地があると言えるものの、ある程度目指し

^{*26} http://web.archive.org/web/20230118094715/https://www.courts.go.jp/app/hanrei_jp/detail3?id=22208

^{*27} http://web.archive.org/web/20230118094829/https://www.courts.go.jp/app/hanrei_jp/detail4?id=22208

^{*28} http://web.archive.org/web/20230118094842/https://www.courts.go.jp/app/hanrei_jp/detail4?id=9

^{*29} <http://www.ipsj.or.jp/award/taikaigakusei.html>

^{*30} <http://www.ipsj.or.jp/award/taikaisyorei.html>

原審裁判所名	仙台高等裁判所
原審事件番号	
原審裁判年月日	<u>昭和-954年3月25日</u>

図 3：昭和 -954 年という表記（下線筆者）

原審裁判所名	仙台高等裁判所
原審事件番号	
原審裁判年月日	<u>昭和46年3月25日</u>

図 4：昭和 -954 年という表記が修正されたあとのページ（下線筆者）

ているものを形にすることができた。

他の法令・条項への参照の解析については、まだ法令の略称を抽出できたところであり、先が長いと感じている。

法令への解析手法を判例にも適用する試みについては、判例のテキストデータの抽出に未だ手間取っており、頑張っていく必要があると考えている。しかし、この抽出の過程で裁判所の web ページにある誤記を発見することができ、ささやかながらも国に貢献することができたことは喜ばしい。

全体として着実に進んでいるものの、まだまだ取り組む余地が多数残っていると自分で評価している。また、研究の過程で作成したソフトウェアはオープンソースソフトウェアとして公開しており、得られたデータも CC0 のオープンデータとして公開している。多くの人がデータをより発展的なプロジェクトに応用してくれることが望ましく、より多くの人にこのプロジェクトを知ってもらえるよう広報にも力を入れたいと考えている。

謝辞

本プロジェクトを進めていくうえで産学間連携推進室の根本晃輔さん・西山大輝さん・太田弘樹さん・広瀬智之さん・服部真吾さんより多数のアドバイスを頂きました。また、産学間連携推進室の担当をしてくださっている新城靖先生・情報科学類長の伊藤誠先生・研究に対してアドバイスを頂いた山本幹雄先生にもお世話になりました。この場をお借りしまして、深く感謝を申し上げます。

参考文献

- [1] 竹井 直樹, 柴田 文明. 保険約款と保険商品のわかりやすさの向上について ― 最近の損害保険業界の取り組みと各損害保険会社の商品改定から考察する ―. 損害保険研究. 2010, 72 巻, 2 号, pp.109-127
- [2] 信岡俊祐, 中村誠, 島津明. 法例文の論理式への変換. 言語処理学会第 13 回年次大会発表論文集. 2007, pp.254-257
- [3] 木村祐介. 号の列举や参照表現をもつ法令文への論理式への変換. 北陸先端科学技術大学院大学情報科学研究科情報処理学専攻修士論文. 2008.
- [4] 磯崎陽輔. 分かりやすい法律・条例の書き方. ぎょうせい, 2011.
- [5] 日本政府. “e-Gov 法令検索”. e-Gov ポータル. <https://elaws.e-gov.go.jp/>, 2022 年 12 月 4 日確認.
- [6] 中村誠, 小川泰弘, 外山勝彦. 法令文中において括弧書きで定義されている法令用語とその語釈文の抽出. 言語処理学会第 19 回年次大会発表論文集. 2013, pp.670-673
- [7] 松田寛. GiNZA - Universal Dependencies による実用的日本語解析. 自然言語処理 Volume 27 Number 3. 2020
- [8] 関本大樹. 租税法令の自動機械処理に関する一つの試み: ハッシュ法に基づく条項名の実用的な抽出方法について. 久留米大学法学 79. 2019. 180-147.
- [9] 金子尚樹. 単語の出現規則に着目した法令中の読み変え規定文の解析手法の提案. 情報処理学会第 85 回全国大会論文集 (2). 2023. pp.729-730.