

分かりやすい利用規約の実現に向けたデータ処理・加工法の考案

五十嵐 太一[†] 小池 一路[†] 堀田 達琉[†] 松田 拓未[†] 山岸 芽生[†] 山崎 達也[†]
[†] 新潟大学工学部

1. はじめに

近年、利用規約には記載内容が複雑なものも多く、規約内容をよく読んでいない人が全体のおよそ 50%にも上る[1]. 本稿では、利用規約を読む人(読者)を増やすために、文章データの自然言語処理に注目して、規約文章を分かりやすくするデータ分析方法を考案した。

2. 利用規約

利用規約とは、あるサービスを提供する事業者側がユーザーに向けてサービスを利用する際のルールを定めたものである。具体的には、サービス利用時に生じる責任の所在などが明記されており、ユーザーが安全にサービスを利用するために重要となる情報が書かれている。

3. 規約文章の自然言語処理

利用規約が読まれない原因として「文章量の多さ」、「規約文章が難しい」の二つが挙げられる[1]. 以下では、これらの原因を解決する手法を説明する(データ処理には Python を用いた)。

3.1 難語の置き換え まず、文章内容の可読性を高めるために、規約中の難語を同義のより平易な単語へと置き換える。複数の利用規約を調査し、規約中に頻出する難語を挙げ、それぞれにより平易な単語を1対1対応させた図 1 に示す対応表を作成し、これに基づいて置き換え処理を行う。置き換え処理には replace 関数を使う。

replace(瑕疵 欠点)

3.2 ハイライト 規約中の重要単語をハイライト処理することによって、読者が注目すべき文章を誘導し、読者が読む文章量を減らす。まず、複数の利用規約を調査し、読者にとって重要となる分野(金銭や責任など)に関係する単語を挙げる。ここで得られた単語にハイライト処理を行う(図 2)。

3.3 リストアップ 規約文章は一般にいくつかの副題に分けられている。規約文章の文頭に副題のみを抽出して並べることで、読者の規約に対する全体的な理解を高めて読むべき文章を誘導し、読者が読む文章量を減らす。

規約文章から副題のみを抽出(リストアップ)するために、副題の識別を行う。識別には副題の文頭に付いている数字、句読点の有無、文字列の長さ(4~15 文字)に注目して処理を行う。最後に抽出された副題を図 3 の右図

に示すように規約文章の頭に並べる。

単語	置き換え先
お客様	あなた
瑕疵	欠点
当該	その
...	...

図 1 置き換え対応表

…トラブルについて、当社は一切責任を負わないものとし、利用者と他の利用者でトラブルになった場合でも、両者同士の責任で解決するものとし、一切の請求を受け付けません。

図 2 ハイライト処理の例

第 1 条(本規約への同意) 利用者は、本利用規約 に同意～ ...	第 1 条(本規約への同意) 第 2 条(規約の変更) 第 3 条(禁止行為) ...
---------------------------------------------	------------------------------------------------------

図 3 リストアップの例 左:処理前 右:処理後

4. 規約文章の処理結果

考案方法を実装した結果、ハイライトやリストアップは概ね想定通りの処理が実現でき、文章の可読性が高まった。特にリストアップに関しては、一目で全体像が把握しやすくなり文章全体の視認性の改善も見られた。

5. おわりに

通常読み飛ばされる利用規約を、自然言語処理を用いて分かりやすく表示する方法を考案し、実装することにより分かりやすく利用規約を表示することができた。しかしながら、難語の置き換えは、1対1に対応する単語群が少なく、平易化はまだ十分とは言えない。またハイライトされる単語数が多いときに文章が見づらくなるや、副題の頭に数字がない場合リストアップできないなど、いくつか解決すべき課題がある。今後、これらの課題を解決することにより、本手法の改善を行っていく。

参考文献

- [1] 消費者庁, “デジタル・プラットフォーム利用者の意識・行動調査”, https://www.caa.go.jp/notice/assets/consumer_system cms101_200520_03.pdf, 参照 Jan.26,2022.