

RedPen を使った文書自動添削ツールの導入提案

PM コース 矢吹研究室 1442031 氏名 小山 隆太郎

1. 研究の背景

ソフトウェアエンジニアはプログラムを組むだけではなく、たくさんの技術文書を書く。そして専門的なチュートリアル、マニュアル等を読み手に理解してもらうことが大切である。ソフトウェアを開発する際には多くの実装テストをする必要がある。そのため CheckStyle や lint 等の静的解析ツールを導入することで、フォーマットのエラーを自動で検知することが出来る。しかし静的解析ツールは文書のチェックを行えるものがないため、文中のミスを修正することに時間を割く等で作業の本質に取り組む時間以外に無駄が生じてしまうことがある。このような状況に対して RedPen と呼ばれる文書静的解析ツールが開発され、日々エンジニアが改良を加えている [1]。RedPen は一部の静的解析ツール (CheckStyle, lint 等) に相当する機能を文書に与えるものであり、文書作成でも最低限の検査を自動で行いたいという動機の下、改良が進んでいる。大学の授業等で文書を提出することがあるため、RedPen の開発に着目することでこれからの文書作成の質が向上するのではないかと考えた。

2. 研究の目的

RedPen を利用し、文書作成に割く時間を短縮できるようにすることで、もっと大きな粒度の問題 (実作業、着目すべき文や質等) に集中できるようにすることが目的である。また、文中のミスを少なくした状態で文書を提出できるようにする [2]。そのため、仮想環境上で RedPen が動作するよう設定を行う。

3. 研究の手法

本研究は以下の手法で行う。

- RedPen が動作する環境を仮想環境を用いて構築する [3]。
- 参考文献のサンプル文 [4] と Web ページからの引用文等で添削を行い、添削結果が本の改善文に準拠するよう設定ファイルを書き換える。
- 研究概要文を RedPen で添削しながら作成する。

4. 研究の結果

本研究は以下のように進んでいる。

- RedPen を仮想環境で起動した。
- RedPen の設定を編集し、添削機能を確立した。

4.1 研究概論の添削結果

執筆段階の本概論を RedPen で添削したところ、以下の文中ミスが指摘された。

- FrequentSentenceStart: 同じ文頭が連続して使用されている場合に指摘される。本概論の文頭に「RedPen」を過度に使用したため指摘された。
- JapaneseAmbiguousNounConjunction: 名詞前後の接続詞が不要であることを指摘する。「文章の間の句点」の一文を「文書間の句点」に修正したことで解消された。この文中ミスの指摘が一番多かった。
- SuccesiveWord: 同一単語が連続して使用されている場合に指摘される。文中に「マシンをを」と間違えて執筆

したからである。

- InvalidSymbol:設定通りではない句点が使用されている場合に指摘される。句点を半角で入力していたため指摘された。

この段階で計 11 箇所の文中ミスが指摘された。指摘箇所を修正した結果 3 件に文中ミスを減らすことが出来た。

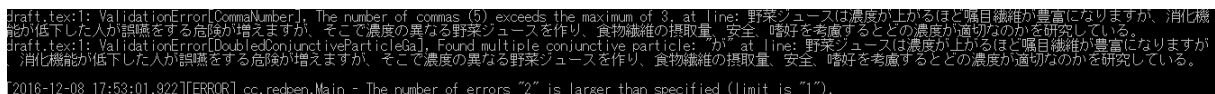
指摘箇所が残った原因として、設定が曖昧であることが考えられる。「文章の間の句点」を文中ミスで指摘しているが、「... の... の...」というパターンも文中ミスとして添削してしまう。例として、「私の研究室のパソコン」という文を、文中ミスとして指摘してしまう。

4.2 引用文の添削結果

参考文献 [4] の一文を以下に引用する。

野菜ジュースは濃度が上がるほど食物繊維が豊富になりますが、嚥下機能の低下した人が誤嚥する危険が増えますが、そこで濃度の異なる野菜ジュースを作り、食物繊維の摂取量、安全、嗜好を考慮するとどの濃度が適切なのかを研究している。

図 1 に引用文を添削した結果を載せる。



```
draft.tex:1: ValidationError(CommaNumber), The number of commas (5) exceeds the maximum of 3, at line: 野菜ジュースは濃度が上がるほど食物繊維が豊富になりますが、消化機能が低下した人が誤嚥する危険が増えますが、そこで濃度の異なる野菜ジュースを作り、食物繊維の摂取量、安全、嗜好を考慮するとどの濃度が適切なのかを研究している。
draft.tex:1: ValidationError(DoubledConjunctiveParticleGa), Found multiple conjunctive particle: 'が' at line: 野菜ジュースは濃度が上がるほど食物繊維が豊富になりますが、消化機能が低下した人が誤嚥する危険が増えますが、そこで濃度の異なる野菜ジュースを作り、食物繊維の摂取量、安全、嗜好を考慮するとどの濃度が適切なのかを研究している。
[2018-12-08 17:53:01.922][ERROR] cc.redpen.Main - The number of errors "2" is larger than specified (limit is "1").
```

図 1 引用文の添削結果

- DoubledConjunctiveParticleGa:一文に 2 回以上、接続助詞「が」が検出されると二重否定になるため検出される。この場合「野菜ジュースは濃度が上がるほど食物繊維が豊富になりますが、消化機能が低下した人が誤嚥をする危険が増えますが、」と内容が理解しにくくなる。参考文献の改善案では、二重否定の片方を「です。」で閉じるべきと指摘し、この通りに文書を書き換え再度添削した結果、指摘は出力されなかった。
- CommaNumber:文中の句点の数が必要以上である場合に指摘される。この場合「濃度の異なる野菜ジュースを作り、食物繊維の摂取量、安全、嗜好を考慮するとどの濃度が適切なのかを研究している。」中の「摂取量、安全、嗜好」の句点を「・」にすることが技術文章では相応しい。

4.3 考察

RedPen には設定ファイルが設けられており、設定を書き換えるごとに出力結果も変わるため、きちんとした設定が確立されてなく、現状の機能は不十分である。RedPen のファイル設定を確立し、文中のミスを抽出することは出来たが、中には正しい文中表記が間違いであることを指摘された。添削を正確に行うためには、設定ファイルを粒度の細かい部分まで作成する必要があると考察する。

5. 今後の計画

RedPen の設定を見直した上で、RedPen 導入後、文書の質と書く時間の変化を調査する予定である。

参考文献

- [1] Redpen を使って技術文書を気軽に校正しよう. <http://gihyo.jp/lifestyle/serial/01/redpen>.
- [2] Redpen でお手軽文書校正. <http://tech.respect-pal.jp/try-redpen/>.
- [3] Redpen ドキュメント. http://redpen.cc/docs/latest/index_ja.html.
- [4] シンプルに書く 伝わる文章術. 飛鳥新社, 2012.