



上流工程文書の記述曖昧性チェックを主 目的とした「日本語文チェックツール」 の開発と運用

SPI Japan 2018 (2018年10月10日)

株式会社インテック

生産本部 生産技術部 新森 昭宏



目次



- 1. 背景
- 2. 改善したいこと
- 3. 改善策を導き出した経緯
- 4. 改善策(日本語文チェックツール)の内容
- 5. 「日本語文チェックツール」の実現方法
- 6. 「日本語文チェックツール」による変化や効果
- 7. 「日本語文チェックツール」の妥当性確認
- 8. 課題と今後の展望

1. 背景



問題意識

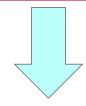
仮説?

- ・ 上流工程文書の記述に曖昧性があると、関係者間の誤解を招き、開発の生産性や品質の低下につながる
 - **要件定義書、基本設計書等**
- ・日本語はもともと曖昧な記述が多用されやすい言語である
 - **「○○と思われる」、「○○的には○○である」**
- レビューでは、表層的なチェックに時間が費やされがち
 - 「てにをは」、体裁

2. 改善したいこと



上流工程文書の日本語文の記述曖昧性を機械的に指摘するツールを提供する



- ・ 文書作成者に注意を促し、明確で厳密性の高い文書作 成ができるように支援したい
- レビューは、内容的なチェックに集中できるようにしたい
- お客様に提出する前に文書をチェックし、誤解を招く 表現を無くしたい

3. 改善策を導き出した経緯



ITエンジニア向け「文章の書き方」の書籍は多く出版されている

- ・2005年
 - ・「SEを極める 仕事に役立つ文章作成術」
- ・2007年
 - ・「SEのための 図解の技術 文章の技術」
- ・2010年
 - ・「要求を仕様化する技術 表現する技術」
- ・2015年
 - ・「ビジネスルールを可視化する要求定義の図解術」
- ・2016年
 - ・「『伝わる日本語』練習帳」
- ・2016年
 - 「ITエンジニアのための 伝わる文章カドリル」

<これほど多い理由>

- そもそも難しい
- ・精神論、お話だけでは徹底しない(コンピュータ支援が必要!?)

3. 改善策を導き出した経緯



Go Beyond

- Word/PowerPointには、文書校正機能が付いている
- ・ 商用の文書校正支援ツールもある

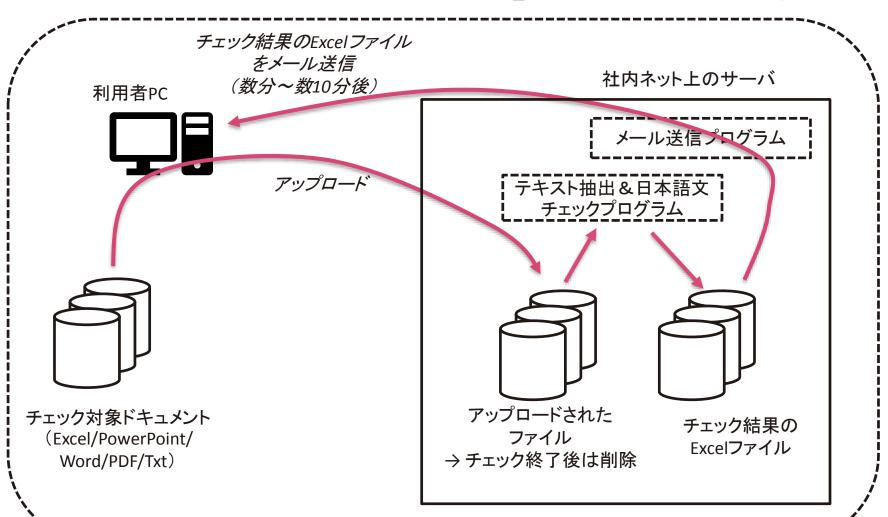
• • •

- **・ しかし、あまり使われていない**
- ・ そもそも一般文書向けであり、提供されている機能は、 誤字・脱字や「表記のゆれ」のチェック
- ・ (上流工程文書向け)記述曖昧性チェック機能は無い

4. 改善策の内容



社内に「日本語文チェックツール」サイトを立ち上げる



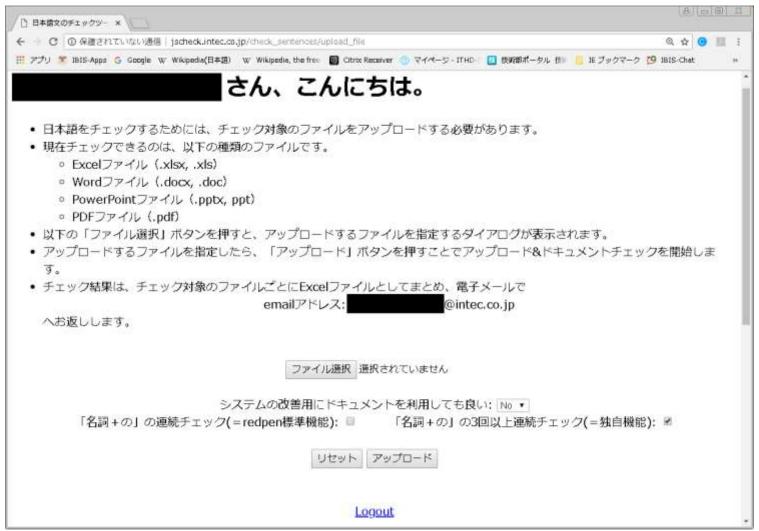
5. 改善策の実現方法



- OSSツールを最大限に活用
 - redpen
 - juman, knp
 - Tika
 - pythonとそのライブラリ(openpyxl, python-docx, 他)
 - Django

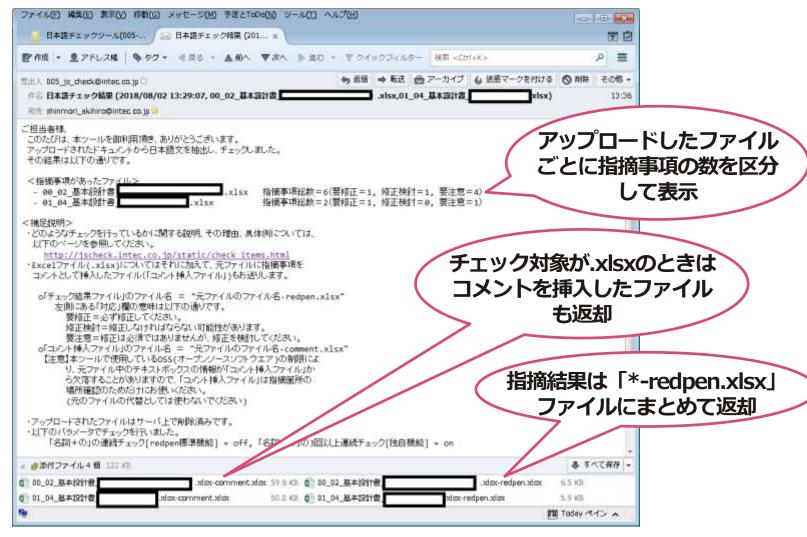


アップロード画面

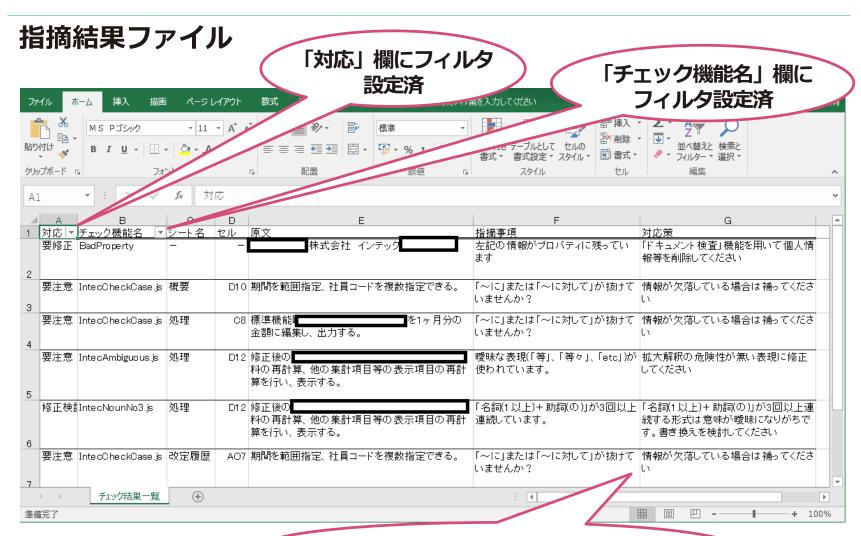




指摘結果のメール画面



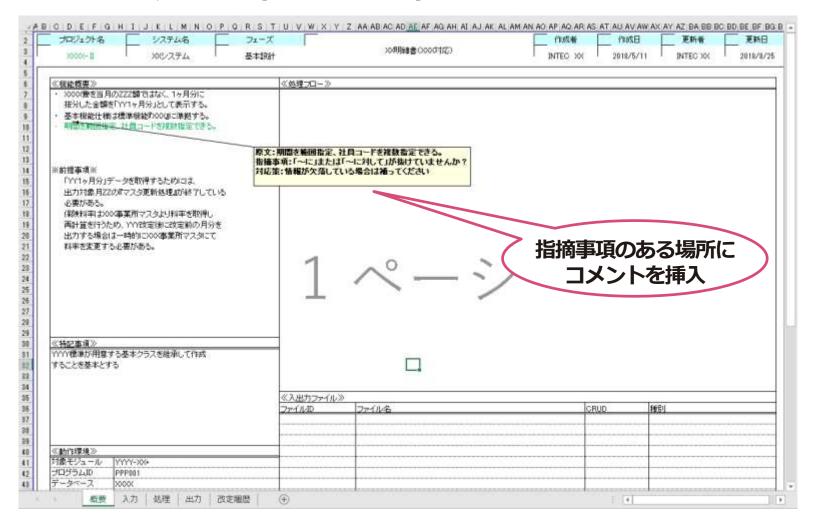




原文、指摘事項に加えて、対応策も提示



コメント挿入ファイル (.xlsxの場合)





<以下のような文の存在をチェック>

- ・ 上流工程文書として不適切な表現を含む文
 - **「非常に」「してない」「だと思い ます」「思っています」など**
- ・ 二重否定が使われている文
- ・ 接続助詞「が」が2回以上出現している文
 - **「~を行うが、~するが、~する。」など**
- ・ 文長が100文字を超えている文
- ・ 読点(、)の数が10を超えている文
- 必須格(「~を」「~に」)が欠落している文
- 否定表現
 - 「~の場合は、~しない」などの表現
- ・ 「名詞+の」が3回以上連続する文
- 助詞「は」「を」がそれぞれ4回以上出現している文
- 「等」「等々」「etc.」を含む文



<ファイルのプロパティ>

ファイルのプロパティに、作成者名・会社などの情報が残っていた場合に注意を喚起

<PowerPointのフッターのCopyright表示の年の部分>

・ Copyright表示の年の部分が、今年より前の年になっている場合に、 注意を喚起

指摘事項は、「要修正」・「修正検討」・「要注意」の3種 類に分類

- 指摘事項ファイル(.xlsx)のフィルタ設定により、利用者が自由 に表示を切り替えられる

6. 改善による変化や効果



<定量的成果>

- 総利用者数=240人以上
- ・ 毎日10回程度の利用(7月の平均)

<定性的成果>

- 要件定義書、基本設計書以外の文書のチェックに使う人が多い
 - **議事録、報告書、連絡文書、論文**
- 初期画面に以下の説明文を書いておいたところ、「Wordの文章 校正機能を把握していないので、それを把握したうえで使い分け たい」とアンケートに書いた人がいた
 - WordとPowerPointにはもともと「文章校正」機能が付いていますので、スペルチェックやスタイルチェックについてはそちらを活用してください。本ツールは、「文章校正」機能では実施されない「記述曖昧性チェック」を主に行います。

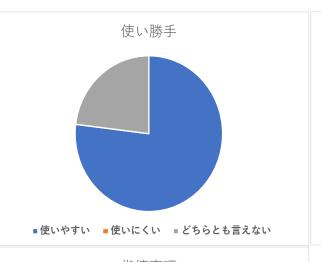
6. 改善による変化や効果(続)

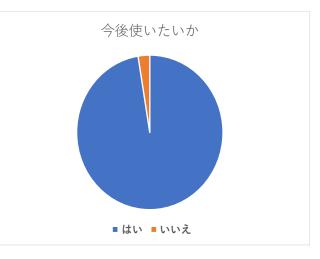


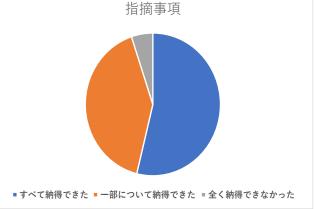
アンケート(無記名)の実施

- 総回答数=76件 (総利用者数の31%)

回答総数		39	
1. 使	い勝手		
	使いやすい	30	77%
	使いにくい	0	0%
	どちらとも言えない	9	23%
2. 指	摘事項		
	すべて納得できた	21	54%
	一部について納得できた	16	41%
	全く納得できなかった	2	5%
3. 今	後使いたいか		
	はい	38	97%
	いいえ	1	3%







回答者の97% 以上が継続利用 の意向を示して いる

6. 改善による変化や効果(続)



改善要望

- 高速化を希望します。
- ファイルサイズごとのざっくりとした処理時間の目安が欲しい。
- 可能であれば修正例がほしい。
- メールによる回答でなく、アドオン等その場でチェックできないか?対象個所が反転するなど。
- 今後は「障害報告書」の文章向けのチェック項目カスタマイ ズも期待します。

• 肯定的意見

- このようなツールがないかと探していました。作ろうかなあ と思っていました。ありがとうございます。
- 一 今後とも活用して行き、チェックの負荷を減らして行きたいと思います。
- 設計書等の自己レビューツールとして部内メンバーに展開したいと感じました。

7. 改善活動の妥当性確認



- アンケートで継続利用の意向を示している人が97% ある
- ・ 継続して、毎日5人程度の利用がある

8. 課題と今後の展望(願望?)



- 課題:現在のツールはあらかじめ組み込んだルールに 従ってチェックを行っているだけであり、このルール にはずれた表現のチェックは行えない
- 今後の展望(願望?):良い例となる日本語文、悪い例となる文を大量に収集し、機械学習の技術を用いた エンジンを開発したい
- ・ 課題:現在のツールは、一般の日本語文の多義性 チェックができる訳ではない
- ・今後の展望(願望?):一般の日本語文の多義性 チェックは本ツールの対象とするところではないが、 機械学習の技術でどこまでできるのかを見極めたい

参考文献、参考URL



- 「システム開発文書品質モデル Ver.1.0」, システム開発文書品 質研究会(ASDoQ), 2015年, https://asdoq.jp/research.html
- 清水吉男著,「要求を仕様化する技術 表現する技術」(改訂第2版),技術評論社,2016年
- 3. 上田志雄, 島田悟志, 「ITエンジニアのための伝わる文章カドリル」, 日経BP社, 2016年
- 4. 独立行政法人 情報処理推進機構, 「ユーザのための要件定義ガイド」(「4.3.3 要件定義文章の品質向上」), 2017年
- 5. 伊藤敬彦, 吉村孝広, 「ドキュメント作成システム構築ガイド」, 技術評論社, 2016年
- 6. RedPen A document checker, http://redpen.cc/
- 7. 山本和英, 鄭育昌, 「3.15 日本語校正」, 情報処理 「特集 自然言語処理技術の現状と展望」, 情報処理学会誌, Vol.57, No.1, 2016年

ご清聴ありがとうございました

