

文書間差分抽出における生成AI活用の考察

第40年度(2024年度)ソフトウェア品質管理研究会 研究コース5「人工知能とソフトウェア品質」 D2E-AIチーム

- 研究員 : 益子なるみ(個人参加)
主査 : 石川 冬樹(国立情報学研究所)
副主査 : 徳本 晋(富士通株式会社)
アドバイザー : 栗田 太郎(ソニー株式会社)

目次

- 背景
- 課題
- 検証の焦点
- 研究内容
- まとめと今後の課題

背景

- 生成AIを品質面で活用をもっとおこないたい
 - ふわっとした個人的な目標
 - ChatGPT(またはCopilot)さん教えて、ではなくもう少しきちんと文書を取り扱う
- ソフトウェア開発における品質向上の活用
 - 品質管理や保証などの自動化
 - 障害予測
 - テストの自動化
 - レビュー
 - フィードバックの分析
 - などなど

背景

- 生成AIを品質面で活用をもっとおこないたい
 - ふわっとした個人的な目標
 - ChatGPT(またはCopilot)さん教えて、ではなくもう少しきちんと文書を取り扱う
- ソフトウェア開発における品質向上の活用
 - 品質管理や保証などの自動化
 - 障害予測
 - テストの自動化
 - ✓ レビュー
 - フィードバックの分析
 - などなど

課題

- レビューの課題
 - 工数の確保
 - 指摘事項のばらつき
 - 見逃しや漏れによる手戻り
 - 人材を育てるにも時間がかかる

課題

■ レビューの課題

✓ 工数の確保

- 指摘事項のばらつき
- 見逃しや漏れによる手戻り
- 人材を育てるにも時間がかかる

■ レビューの観点のうち対象としたもの

No.	観点	説明	検証対象
1	網羅性	必要な項目やセクションの網羅	
2	正確性	記載情報やデータが正確	
3	一貫性 整合性	ドキュメント内やドキュメント間での矛盾, 用語の一貫性	○
4	相違点	追加・削除された情報や変化点, 情報の過不足	○
5	適合性	想定している文書の利用対象者に適しているか	
6	形式・書式	文法見出し等スタイルの適切さ, 記載粒度・表現	
7	視覚的要素	フォーマットや図表の形式や色遣い	
8	参照関係	ドキュメント間の参照や依存関係の正しさ	

課題

■ レビューの課題

- 過去の文書整合性をチェックするツールはベンダーロックインやEXCELフォーマットに依存
 - ⇒ 活用できればとても有効だが、習熟に時間を要したりバージョンアップ追従が大変
 - ⇒ EXCELの関数、マクロ、方眼紙・・・

機能定義	
機能-A	○○
機能-B	△△

詳細設計	
FUNC-AA	機能-A
呼出し	機能-Z
出力	

トレーサビリティがしっかりしているフォーマット。

詳細設計	
FUNC-AA	機能-Aを呼出す。戻り値を機能-Zにわたし画面に出力する。

- ⇒ 作りこまなくても、関係性が確認できたら便利

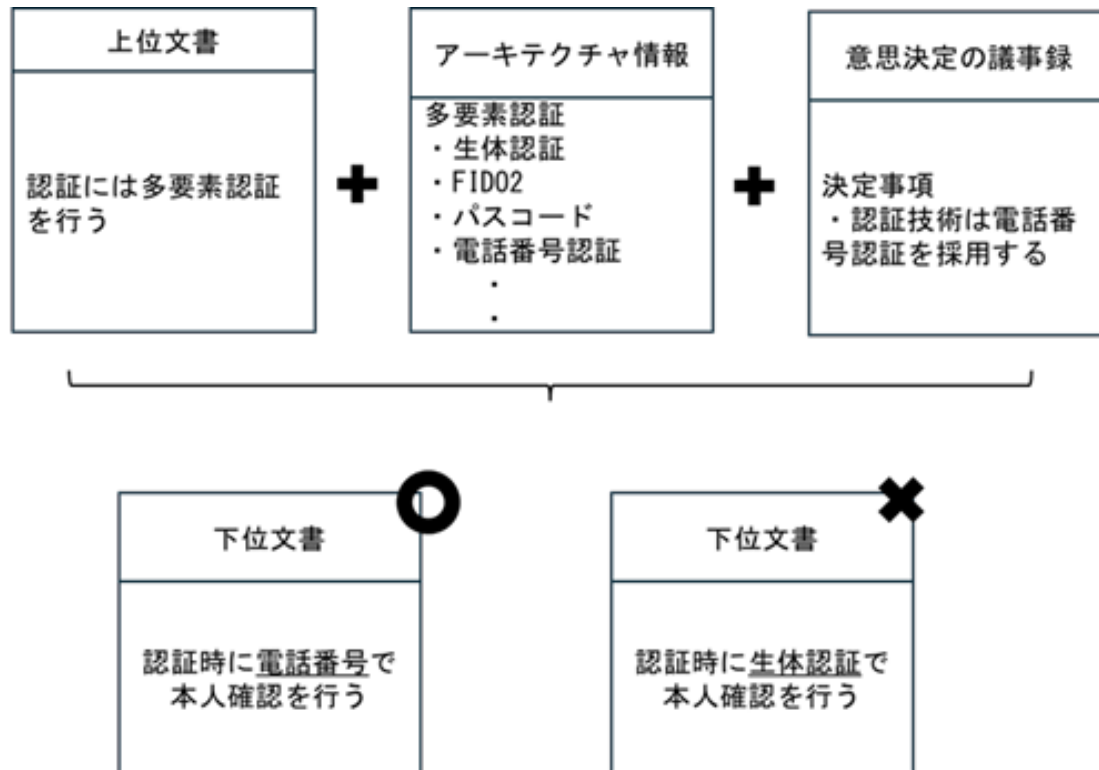
課題

■ 生成AIの課題

- 学習にかかるコスト
 - 最近は軽量版も出てきているがそれでも安価とは言えない
- ハルシネーション、誤りの混入
 - ハルシネーションに気が付かず判断を下す・・・
- 著作権侵害
- 人権侵害
- レイテンシー
- 説明性の担保の難しさ
- など・・・

検証の焦点(1/2)

- 生成AIでできればよいなと思ったこと
 - 最終的には相互関係、依存関係のある文書間レビューができることが理想



検証の焦点(2/2)

■ 今回の研究内容

- 対象文書を減らす
 - 2文書に絞る
- 確認ポイントを差分に絞る
 - そもそも同一性と相違・言葉のずれとは
 - これらが正しいとは人間が判断するしかない
 - 比較すべき正解があるなら普通に自動化してML
- 生成AIならではのメリット
 - 差異が出力されるだけでは、差分ツールと変わらない

機能-Aを呼出す。戻り値を機能-Zにわたし画面に出力する。



機能-Aを呼出す。戻り値を機能-Yにわたし画面に出力する。

生成AIならではの期待

機能-Zから機能-Yになったことにより〇〇である

研究内容

■ 今回の研究内容

– 利用文書

- 政府CIOポータルの『デジタル・ガバメント推進標準ガイドライン』より一部（第2章 プロジェクトの管理）を抜粋

– 利用プロンプト

- 青字箇所は任意で付与

ドキュメントのレビューをしています。以下の2つの資料の確認項目について抽出して比較してください。

資料

『文書①の資料名』

『文書②の資料名』

確認項目(任意)

・ ターゲットとする節名

比較項目(任意)

・ 用語の変更/文章の流れ

文書①の資料名

～省略

文書②の資料名

～省略

研究内容

– 利用ツール

- AzureOpenAI

- モデル gpt-4o (version:2024-08-06)を使用

- RAGは使わないの？というご意見もありそうなので

- 最初はRAG(Azure Ai Search)を使うつもりだった

- 参照文書のリンク先があっていない場合があり今回の研究では、ツールを使いこなすことが目的ではないため、差異の検証に絞った

研究内容

■ 今回の比較パターン

- 変更量や位置などで精度の違いが出るかを確認
- 変更についての追加のコメントが出るか確認
- 1文字、単語、箇条書きの順番や章の入れ替え
- 1文字は埋もれてしまうことを懸念して箇条書きなど1行 1～2文で構成されているものを選択

結果

■ 差分出力の評価（抜粋）

- 差分と総合的なコメントは出力された
- タイトル行や主語・述語は比較的変更あると判断されやすい
- 文中や複数個所に分散すると出力されづらいように見える
- 同じ単語の変更を単独で行うと抽出される

粒度	結果	文書①	文書②
1文字	×	設計・開発工程に入る前に要件定義の確定を行う前	設計・開発工程に入る後に要件定義の確定を行う前
	○	設計・開発工程に入る前に要件定義の確定を行う前	設計・開発工程に入る前に要件定義の確定を行う後
複数・異なる段落を同時に変更して実施	○	～「2)ウ 工程管理」、「2)エ 指標管理」及び「2)ク 品質保証」に記載した項目とする	～「2)ウ 工程管理」、「2)エ 指標管理」及び「2)ク 品質管理」に記載した項目とする
	×	～実施内容等をPMOと調整し、デジタル庁が別途定める手順に基づき、それぞれの場面(以下「レビューポイント」という。)において、次のとおり工程レビューを実施するものとする。～	～実施内容等をPMOと調整し、デジタル庁が別途定める手順に基づき、それぞれの場面(以下「レビュー観点」という。)において、次のとおり工程レビューを実施するものとする。～
	×	～プロジェクト推進責任者は、レビューポイントにおいて自己点検を行い、その結果をPMOに送付するものとする。～。	～プロジェクト推進責任者は、レビューポイントにおいて内部監査を行い、その結果をPMOに送付するものとする。～。

結果

■ 出力イメージ全体

- 変更箇所だけ出力されるのではなく、前後の文章を踏まえてコメントがある
- 若干の違い・・・は毎回出るのであまり参考にはならない

■ 例①

- レビューのタイミングが異なる点に言及されている

相違点

・ 工程レビューの内容

- 『検証資料 1-4-1』では、工程レビューの具体的なレビューポイントが「調達仕様書に添付する要件定義書の作成終了前」「設計・開発工程に入る前に要件定義の確定を行う前」「総合テスト計画書の確定を行う前」となっている。
- 『検証資料 1-4-2』では、レビューポイントが「調達仕様書に添付する要件定義書の作成終了前」「設計・開発工程に入る前に要件定義の確定を行う前」「結合テスト計画書の承認を行う後」となっており、最後のレビューポイントが異なる。

・ プロジェクトのモニタリング及び停止・改善

- 両資料ともにモニタリングの方法や停止・改善の基準は共通しているが、具体的な記載内容や手順に若干の違いが見られる可能性がある。

結論

両資料は、プロジェクト管理に関する基本的なフレームワークや手順が共通している一方で、特定のプロセスやレビューのポイントにおいて若干の違いが見られます。特に、工程レビューの具体的なポイントが異なるため、これらの違いを考慮してプロジェクトの実施や管理を行う必要があります。

結果

■ 出カイメージ

- 例②
- プロジェクト管理のアプローチが違うことに言及

相違点

- 用語の変更:
 - 資料『検証資料2-3-1』では「品質管理」という用語が使用されていますが、『検証資料2-3-2』では「品質保証」という用語に変更されています。
 - その他の用語や表現については、基本的に同じ内容が記載されていますが、細かな表現の違いがある可能性があります。

比較項目: 用語の変更

- 品質管理 vs 品質保証:
 - 資料『検証資料2-3-1』では「品質管理」という用語が使われているのに対し、『検証資料2-3-2』では「品質保証」という用語に変更されています。この変更は、プロジェクトの品質に対するアプローチや考え方の違いを反映している可能性があります。

結論

両資料はプロジェクトの立ち上げ及び初動に関する基本的な枠組みや手順については共通しているものの、用語の変更が見られます。特に「品質管理」と「品質保証」の用語の違いは、プロジェクトの品質に対するアプローチの違いを示唆しているかもしれませぬ。これらの違いを踏まえ、今後のプロジェクト管理においてどのように適用していくかが重要です。

まとめと今後の課題

■ まとめと今後の課題

- 単純な差分比較としてではなく、その差分の意味を解釈を期待する
- 元の期待する内容には程遠いのでさらに深掘して確認していく必要がある
 - 同一の意味で異なる表現
 - 類似語の違いによる推論された意味の相違
 - ソフトウェア開発設計書などにおける抽象度の異なる記述の同一性・相違点の抽出
 - 三つ以上の文書や形式の違うもので実施
 - 意図的に変更している個所の比較対象としての除外

ご清聴ありがとうございました。

謝辞

論文作成にあたり、石川冬樹主査、徳本晋副主査、栗田太郎アドバイザーにご指導賜りました。深くお礼申し上げます。