

商用ツールを活用した要求仕様書中の 図表の自動要約手法の提案

永野 貴大† 長岡 武志‡ 北川 貴之‡ 位野木 万里†

工学院大学†

東芝デジタルソリューションズ株式会社‡

1. はじめに

デジタルトランスフォーメーション（以下、DX と略す）の加速が求められている[1]。DX の推進において、関連する情報システムの要求仕様書の内容理解は重要である。

情報システムの要求仕様書は、対象システムの開発の背景と目的、機能要求、非機能要求等が記述される[2]。機能要求の記述は、対象業務の流れの中で、対象システムのどの機能がどのように使われるのか、また、関連する他システムとの連携関係が記述される。そのような業務と機能の関係性は、業務フロー図や業務処理仕様一覧表等の図表の形式で複数ページに渡って記述されることが多い。また、開発当事者ではない場合、入手可能な情報システムの文書ファイルは、紙の仕様書をコピーして作成された pdf ファイルや画像ファイルのことがあり、必要な仕様を正確かつ迅速に読み取ることが困難な場合がある。

本稿では、Microsoft Azure Form Recognizer[3]等を活用した、要求仕様書の自動要約手法を提案する。これにより、図表形式で記述された要求仕様書を、初級技術者が効率的に理解することを支援する。以下、本稿の 2 章では本研究の課題とそれを解決するためのアプローチについて述べる。3 章では、ツールを用いた要約手法を提案し、4 章においては、本手法を具体事例に適用した結果を示し、5 章においては提案手法の有効性について考察する。6 章にて本稿のまとめを行う。

2. 課題と解決策へのアプローチ

情報システムの要求仕様書の理解において、2 つの課題が考えられる。1 つ目の課題は、初級技術者にとり、短時間での理解が困難という点である。とくに、要求仕様書には多くの複雑な図表が含まれる場合があり、それら図表から、重要情報を迅速に抽出して理解することは容易で

はないと考えられる。本課題の解決策として、図表の要約ツールを作成し、自動的に図表から要約情報を得ることで重要情報の取りこぼしの防止や作業コスト削減を図る。

2 つ目の課題は、要約ツールを一から作成するのに高いコストがかかる点にある。本課題への解決策として、事前学習済みの商用ツールを用いることを提案する。本研究では商用ツールである、Microsoft Azure Form Recognizer を利用する。当該ツールは、表形式の画像等からの情報抽出が可能であり、セル内の誤認識、誤検出が少ないという利点がある。加えて、当該ツールは、API を通してアドオンを追加し、要約情報の出力をカスタマイズできる柔軟性を備える。また、読み取り対象の表のセル内の記述が文の場合、Form Recognizer のみでの要約情報の生成は困難なため、文章を多角的に分析可能なテキストマイニング用のツール KH Coder[5]を利用する。

3. Form Recognizer を活用した図表の要約手法

図 1 に提案する図表の自動要約手法の処理の流れを示す。本図において、対象とする図表とは、業務の流れを記述する、業務フローや業務処理記述を対象とする。

図表の自動要約は次の手順で行われる。(1) 仕様書の設計要素が存在する列に応じた座標を定め、Form Recognizer のソースコード内で座標を設定し、図表中のセル内の要素を抽出する。

(2) アドオンによって、抽出したセル内の記述から設計要素を特定し、各設計要素の一覧表を作成する。(3) セル内が文章となっている場合、KH Coder を利用する。(2) によって出力された設計要素から出現回数の多いものをピックアップし、強制抽出対象に設定して該当列での出現回数を出力する。(4) KH Coder の共起ネットワーク機能により、出現語間での関連性を視覚化した図を出力することで設計要素間の関係性に関する要約情報を提供する。

4. 適用評価

4.1 評価概要

提案手法に基づいて生成した要約情報の活用の有効性について、具体事例を用いて評価する。

Proposal of Automatic Summarization Method of Figures and Tables in Requirements Specifications Using Commercial Tools

†Nagano Takahiro, Inoki Mari, Kogakuin University

‡Nagaoka Takeshi, Kitagawa Takayuki, Toshiba Digital Solutions Corporation

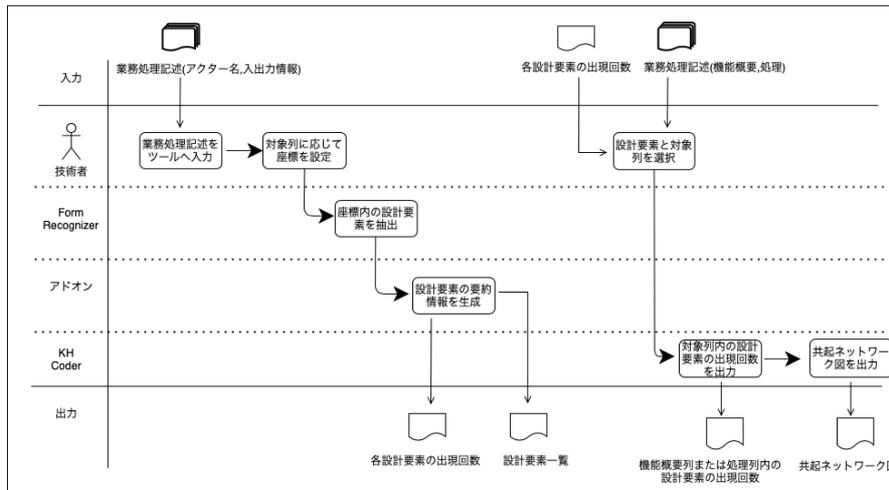


図1 Form Recognizer と KH coder を組み合わせた自動要約手法

要約対象とする文書は年金業務システム調達仕様書[4]の別添に含まれる業務処理記述(約60ページ)のうち、最初の10ページをとりあげて評価を行う。評価手順は次の通りである。(1)該当する要求仕様書中に定義された設計要素(名称と出現回数)を一覧化し、正解データとして用意する。(2)提案する手法により要約情報を生成する。(3)3人の被験者(工学院大学情報学部の3年生)により、要求仕様書から設計要素を手動抽出し、設計要素一覧としてまとめる。(4)(1)~(3)の結果を比較し評価する。

4.2 評価結果

表1に正解データと、手作業、ツールを用いた場合の正味と述べの件数結果及びツールの品質評価を示す。正解データと手作業を比較すると2件ほどの誤抽出が発生していた。正解データとツールを比較すると件数の差は見られなかった。ツールによる適合率、再現率は、90%以上で高い水準であった。

表2に、手作業とツールによる抽出にかかった平均時間と作業コストを示す。3人の被験者が平均0.63人時で抽出しているのに対し、ツールは0.28人時で抽出できた。

表1 適用評価結果

| | 述べ(件数) | 正味(件数) | 適合率 | 再現率 | F値 |
|-------|--------|--------|-------|-------|-------|
| 正解データ | 183 | 67 | | | |
| 手作業 | 185 | 67 | 0.994 | 0.983 | 0.988 |
| ツール | 182 | 67 | 0.998 | 0.994 | 0.995 |

表2 抽出時間及び作業コスト比較

| | 合計時間 | 平均時間 | 作業コスト |
|------|--------|-------|--------|
| 技術者A | 40分35秒 | 4分0秒 | 0.67人時 |
| 技術者B | 42分13秒 | 4分17秒 | 0.70人時 |
| 技術者C | 32分58秒 | 3分25秒 | 0.54人時 |
| ツール | 17分18秒 | 2分6秒 | 0.28人時 |

5. 考察

要求仕様書中に記述された図表から設計要素を効率的に自動抽出可能であることを確認できた。作業コストに関しては、人手で実施した場合と比較し、約50%以上の削減が確認できた。

6. まとめ

商用ツールを活用した図表からの設計要素の読み取りによる要求仕様書の要約と技術者支援手法を提案した。本研究では、業務処理仕様の定義表から設計要素とその出現回数を出力したが、今後はさらに様々な図表の読み取りと、要約生成の手法を検討し、初級技術者の要求仕様書の理解支援に貢献していく。

謝辞

本研究は JSPS 科研費 JP19K11907 の助成を受けた。

参考文献

- [1] 経済産業省, デジタルトランスフォーメーションに向けた研究会, DX レポート ~IT システム「2025年の崖」の克服と DX の本格的な展開~, 経済産業省, 2018, https://www.meti.go.jp/shingikai/mono_info_service/digital_transformation/pdf/20180907_03.pdf (参照 2021-01-16)
- [2] ISO/IEC/IEEE 29148:2018, Systems and software engineering — Life cycle processes — Requirements engineering, 2018.
- [3] Form Recognizer - 自動データ処理システム - <https://azure.microsoft.com/ja-jp/services/form-recognizer/> (参照 2021-10-03)
- [4] 厚生労働省, 年金業務システム(個人番号管理サブシステム等(2次開発情報連携))に関わる設計・開発等業務及びアプリケーションソフトウェア保守業務 調達仕様書(案) pp, 132-141 <https://www.mhlw.go.jp/sinsei/chotatu/chotatu/shiyousho-an/160428-1.html> (参照 2021-10-03)
- [5] KH Coder <https://kncoder.net/> (参照 2021-12-10)