

令和 6 年 6 月 10 日現在

機関番号：32613

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K11907

研究課題名（和文）要求仕様書の自動要約ツールの開発と適用評価

研究課題名（英文）Development and Evaluation of Automatic Summarization Tool for Requirements Specification

研究代表者

位野木 万里（Mari, Inoki）

工学院大学・情報学部（情報工学部）・教授

研究者番号：10739634

交付決定額（研究期間全体）：（直接経費） 3,300,000 円

研究成果の概要（和文）：要求仕様書は、システム化の背景や目的、機能要求、非機能要求から構成される。本研究では、機能要求に対しては、設計要素抽出要約、非機能要求に対しては定義状況を定量化し定義漏れリスクを浮かび上がらせるリスク要約、システム化の背景や目的には、重要文抽出の従来型の要約を対応づけ、要求仕様書の自動要約ツールを提案した。加えて、重要な概念が要求仕様のどこに記述されているかを示す記述状況可視化要約を活用し、文書の理解を支援する手法を考案した。適用評価により本手法は技術者の要求仕様書の理解の加速に有効であること、様々な技術文書への記述状況可視化の適用により、技術者への知識継承に有効であることを確認した。

研究成果の学術的意義や社会的意義

Digital Transformationへの取り組みが活発化し、技術者にとり、情報システムの要求仕様書の作成に加え、関連するシステムの要求仕様書等の技術文書を参照理解する機会が増加している。開発した自動要約ツールを活用して、要求仕様書の要約情報を用いることで、新規ビジネスアイデア創出、リスク予測、基本設計の効率化、仕様変更の影響箇所の特定の支援が可能になる。本技術を大規模かつ高度な内容の技術文書に活用し、要約情報を用いて、知識を形式知化・共有することで、組織の知識継承にも貢献し、IT人材不足の解決やDigital Transformationの社会実装の一層の加速への貢献が期待できる。

研究成果の概要（英文）：The requirements specification document includes the background, purpose, functional requirements, and non-functional requirements of target systems. We propose an automatic summarization support tool for these specifications. For functional requirements, our tool extracts design elements, while for non-functional requirements, it provides a risk summary indicating potential definition omissions. The background and purpose are summarized using a conventional method to extract key sentences. Additionally, we offer a method to support document comprehension by indicating where important concepts are located within the requirements specification. Evaluation confirmed that this method effectively accelerates engineers' understanding of the requirements specification. Furthermore, visualizing the description status in various technical documents is expected to be useful for knowledge transfer to engineers.

研究分野：ソフトウェア工学

キーワード：要求工学 要求仕様書 自動要約 理解支援 記述状況可視化 自然言語処理 AI

1. 研究開始当初の背景

情報システム開発において、要求定義工程で作成される要求仕様書は対象システムの開発範囲を確認するために重要な成果物である。旧型の基幹系情報システムであるレガシーシステムに対して、膨大な量の要求仕様書が資産として蓄積されている。レガシーシステムの保守作業のために、開発者は要求仕様書の内容を理解する必要がある。当該システムに精通していない初級技術者にとり、短期間に膨大な量の要求仕様書の内容を理解するのは困難である。

ところで、AI を用いた自然言語処理技術の開発が盛んである。これら技術には、自然言語で記述された文書をベクトル表現化し、文書中に出現する用語の出現頻度の高い用語を含む文を重要文として抽出し、文書全体を要約する技術が含まれる。自然言語処理技術の活用により、前述したレガシーシステムの要求仕様書の自動要約の実現も期待できる。

しかし、要求仕様書には対象とする業種業務領域に固有の用語が使われることや、一回のみ出現した用語を用いて記述された要求文であっても重要な要求となる場合があることから、用語の出現頻度に基づく重要度判定による要約のみでは、要求仕様書全体の意図をとらえた要約になっていない。要求仕様書を対象にした場合、膨大な量の要求仕様書を効率的に理解可能とする、要求仕様書の自動要約技術の具現化には至っていなかった。

研究代表者は、高品質な要求定義のための手法やツールの研究開発に従事してきた。要求仕様の一貫性検証支援ツールを複数企業に適用評価する過程で、レガシーシステムの保守業務がソフトウェア開発事業の大半を占めていること、当該システムの保守担当者は本来の開発者とは世代交代しており、レガシーシステムの要求仕様書の理解に多大なコストが発生していることが分かった。また、膨大な量の要求仕様書をコンパクトに要約した情報があれば、ステークホルダ間の合意形成に有効であることが分かった。

AI に基づく自然言語処理技術のアルゴリズムやオープンソースの研究成果が多数提供されているが、自然言語で記述された要求仕様書に着目した自動要約技術は、成熟した状況には至っていない。研究代表者らのこれまでの研究成果である「検証知識」の形式知化と検証自動化の知見を、AI に基づく自然言語処理技術と融合することで、要約自動化の実現に有効であると考え、要求仕様書の要約知識の形式知化と自然言語処理技術に基づく要約自動化ツールの開発に関する本研究の着想に至った。

2. 研究の目的

本研究ではシステム構築の背景や目的等の要点をまとめること、基本設計以降で必要となる設計情報を抽出すること、リスクのある要求を指摘することを自動化する要求仕様書の自動要約ツールを提供することで、要求定義のノウハウを共有し、レガシーシステムの要求仕様の理解や基本設計工程へのスムーズな連携を加速し、我が国のソフトウェアベンダーのレガシーシステムの保守業務への依存体質からの脱却と、要求定義技術のレベルアップを目指す。また、AI に基づく自然言語処理技術に関する研究成果を取り入れた要求仕様書の自動要約ツールを実システムの要求仕様書の要約に適用することで、従来は人手により実施していた要求定義のタスクの自動化の範囲を拡大し、デジタルビジネスへの変革等を実現する魅力あるソフトウェアサービスの要求定義の効率化、高品質化により、ソフトウェア開発技術の国際競争力向上に貢献する。

3. 研究の方法

(1) 実施項目 1: ベテラン技術者による要求仕様書の要約ノウハウの形式知化とルールベースによる要求仕様書の自動要約ツールの開発: アクター、画面、データ、振る舞い、デバイス等の設計要素に着目し、設計要素の定義や説明が含まれる要求文を基本設計に必要な情報として抽出する要約、リスクワードを含む要求文をリスクのある要求として抽出する要約等の要求仕様書の要約知識を有識者へのインタビューにより明らかにし形式知化する。抽出した知識は、辞書とルールの形式により定義し、要求仕様書の自動要約ツールを開発する。

(2) 実施項目 2: 自然言語処理技術の自動要約アルゴリズムを用いた要求仕様書の自動要約機能の開発: 要求仕様書を構成する各要求文を自然言語処理 AI 等によりベクトル表現に変換し、用語の出現頻度から重要と考えられる用語を含む要求文を特定する、抽出型の自動要約を生成する機能を開発する。

(3) 実施項目 3: [実施項目 1][実施項目 2]で開発した自動要約機能を組み込んだ要求仕様書の自動要約ツールを、実案件に適用し有識者により評価を行う。評価を通して、技術者の要求仕様書の理解のプロセスと要求仕様書の自動要約ツールの望ましい使われ方を明らかにし、開発した要約機能の改善項目を整理し、同ツールの改善、拡張を行う。

4. 研究成果

図 1 に要求仕様書の自動要約ツールの全体像を示す。図 1 の左側が、本研究で開発した要求仕

様書の自動要約ツールの機能構成を示している。図1の右側は、要求仕様書を読む目的別に技術者の理解のプロセスの詳細化とプロセスを加速するための支援技術の対応付けを示した参照モデルである。これは[実施項目3]の成果に対応する。本ツールは、U1とU2の理解の加速を支援し、技術者がU3の理解に集中し、基本設計以降の開発にスムーズに連携できることを目指している。以下、本研究の主な成果について実施項目と図1を対応づけて説明する。

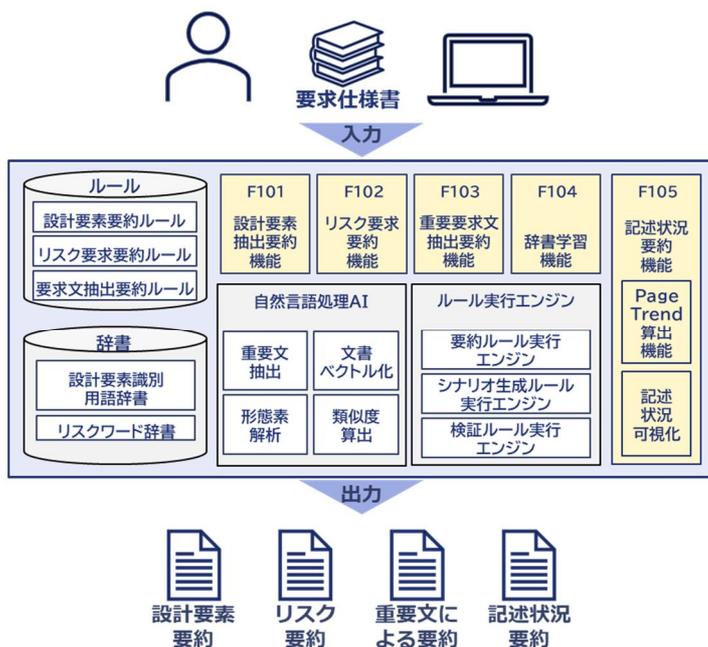


図1 要求仕様書の自動要約ツール全体像

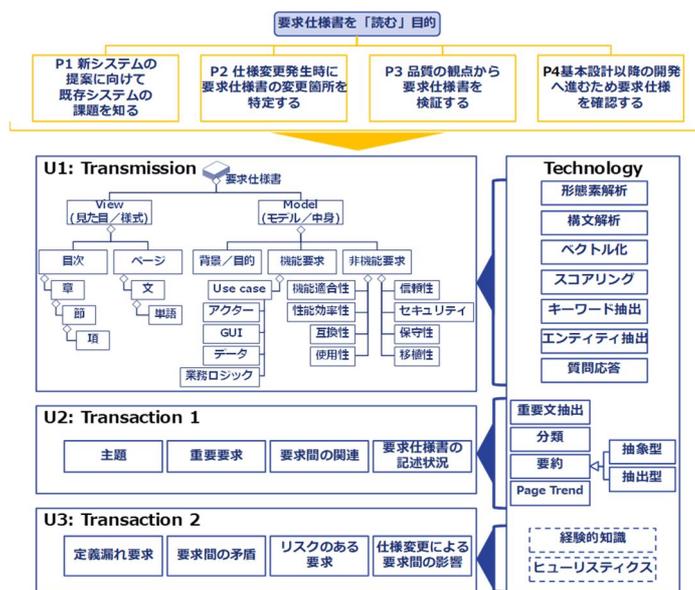


図2 要求仕様書の読解目的別の理解のプロセスと支援技術の参照モデル

(1) [実施項目1]の成果：設計要素抽出要約とリスク要求要約

基本設計にスムーズに接続するには、要求仕様書から設計対象となるアクター、データ、画面、振舞い等の設計要素を抽出し一覧化し、開発規模感を把握することが重要である。そこで、機能要求の記述から、アクター、データ、画面、振舞いを設計要素要約ルールに従って抽出し、基本設計に必要な設計要素一覧として要約化する仕組みを構築し、F101の機能を開発した[高橋宏季2020][位野木万里2021]。

非機能要求記述を種類別に自動分類し定義状況を定量的視点で可視化する要約手法を考案した。本手法は、要求仕様書に含まれる非機能要求を入力として、Convolutional Neural Networkとドメインオントロジーに基づく2つの自動文書分類技術を組み合わせ、非機能要求の自動分類結果を定量的観点で視覚化することで要約データを生成する。本研究では、公開されている要求データを取りあげて、F104の学習機能により、非機能要求の分類ノウハウを学習させ、分類精度と定量要約手法の可視化方法を評価した。非機能要求の分類については、Convolutional

Neural Network、ドメインオントロジーのそれぞれ単体での結果と、Convolutional Neural Networkとドメインオントロジーを組み合わせた合議アルゴリズムを適用した結果を比較し、合議アルゴリズムによる方法の方が、分類精度を向上させることを確認した。また、有識者インタビューにより、非機能要求の定義状況を定量要約により可視化することは、定義が不十分な非機能要求の種類を特定しやすく、要求仕様書中のリスクの発見と防止に有効活用可能であることを確認できた[位野木万里 2022][YutaSaikawa2021]。

(2) [実施項目 2]の成果：重要要求文抽出要約

要求仕様書の目的や背景の書き方や記述内容は一般文書と同様であり、重要な用語は複数回記述されていることから、用語の出現頻度に応じたスコア算出と、高スコアな用語を含む代表文抽出による抽出型要約を実行する F103 の機能を開発した。本機能の開発においては既に公開されている OSS を活用した[石井達樹 2020][位野木万里 2021]。

(3) [実施項目 3]の成果：理解のための知識の参照モデルと記述状況可視化の開発

複数の要約情報の自動生成の検討において、要求仕様書の読解の目的、理解のプロセスを詳細化し、各プロセスをサポートするテクノロジーを対応づけ、形式知化した。それを、要求仕様書の理解を支援する自動要約技術の参照モデルとして定義した(図 1 右)。本参照モデルを活用すれば、要求仕様書を読む目的を達成するために、どのようなテクノロジーをどこで活用すべきか、自動要約ツールの生成結果から、理解を加速するためにどのようなノウハウが必要となるのかをフィードバックする起点として活用することが可能になる。

ところで、要求仕様書中に記載の文と、上記の用語の出現頻度に応じたスコア算出に基づく重要文抽出の過程で、重要文がどのあたりでどの程度強調されているのかが分かると、文書全体の記述状況を把握でき、重要文抽出がピンポイントの要約であるならば、ページ全体の状況に関する線の要約重要であることに気付いた。そこで、重要概念が要求仕様書内にどのように記述されているかを示す、記述状況を要約する手法を新たに考案し、機能 F105 を開発した。記述状況の要約手法は、重要文抽出の技術に基づき、要求仕様書に記述された話題を自動抽出し、要求仕様書の各ページの構成要素特性と、ページ間類似度を算出し、これらにより生成したデータを可視化し記述状況として提供する。図 2 に本機能に用いた記述状況可視化要約の手順を示す。

記述状況可視化要約の手順は 4 つのステップからなる。Step1 では、入力された要求仕様書から要求文を抽出する。Step2 は、抽出された要求文の中から重要文を抽出する。これは[実施項目 2]で実現した重要要求抽出要約機能を適用する。Step3 では、対象ページにおける当該話題の出現率を示す Page Trend を算出する。なお、Page Trend は、対象ページの要求文数によって補正する PageTrend2 と、補正は行わない、PageTrend1 の 2 つの指標を定義した。Step4 において、PageTrend1、PageTrend2 の結果を、ページ毎重要要求毎の折れ線グラフやデータバー等で可視化する。



図 2 記述状況可視化要約の手順

本研究では、記述状況の要約手法を実案件の要求仕様書に適用しケーススタディを行った。ケーススタディでは、有識者インタビューと、初級技術者による理解度テストを行い、記述状況の要約情報が要求仕様書の読解に与える影響について分析した。分析により、記述状況の要約情報は、要求仕様書の読解の目的を十分に理解したベテラン技術者にとり、対象の要求仕様書を読みながら、要求の分析、検証、基本設計の実施方針の立案等に関する理解の加速に有効であるという効果を確認できた[中村雄太郎 2024]。

文書の構造やページ間の類似性からなる要求仕様書全体の記述状況の可視化による自動要約手法について、実装方法のバリエーションを洗い出し、試作プログラムにより実現可能性について検証した。具体的には、自然言語処理モデル BERT と、GPT-3 などの大規模言語モデルを用いて手法を実現した。本手法のバリエーションとしては、技術文書の全体要約において、前者の場合には重要文抽出を、後者の場合には抽象型要約によってトピックを特定することで区別した。また、特定したトピックを用いて要求仕様書の各ページの記述状況を自動可視化する実現方法

に関して、大規模言語モデルの種類によらない汎用アーキテクチャを明らかにした。

本適用結果をまとめ、実技術者や管理者等にヒアリングを行った。ヒアリングにより次の点を確認した[位野木万里 2023]。(1)要求仕様書に加えて、論文、特許明細書、技術レポート、企画書等のさまざまな技術文書へ適用可能である。(2)組織に蓄積されている多数の技術文書を人手のみで参照し理解することは困難なため、記述状況の要約技術を用いて、ベテラン技術者が指定した重要概念が文書のどこに記述されているかを可視化し共有することで、若手技術者へのノウハウの伝達を加速し、組織の知識継承に有効であると考えられる。

F104 の学習機能に関しては、要求の分類に関する一部の学習機能を試作するにとどまっている。本研究を進める過程で、大規模言語モデルとそれを用いた生成 AI 等のサービス提供が台頭し、学習機能をスクラッチで開発するよりも、生成 AI へ入力するプロンプトを工夫することで、要約情報に加えて、要求仕様書の自動生成の精度も向上することを確認している[中野丈 2024]。今後は、様々な大規模言語モデルと連携し、技術文書理解、要約、分析、生成、検証、管理等の要求定義タスクを自動化に対して本研究成果の適用を継続していく。

補足：文中の参考文献の著者とタイトルのみを示す。書誌情報詳細は「本報告書 5 . 主な発表論文等」を参照されたい。

[中村雄太郎 2024] 中村雄太郎, 長岡武志, 北川貴之, 位野木万里, 本位田真一 Page Trend による記述状況の可視化を用いた要求仕様書の理解手法の提案と tf-idf による記述状況の比較評価

[中野丈 2024] 中野丈, 島川遼太郎, 杉村康気, 宅間健生, 村野遼, 奥田博隆, 位野木万里, 要求工学の知見をプロンプト設計に組み込んだ生成 AI を活用した要求定義支援手法

[位野木万里 2023] 位野木万里, 技術文書の記述状況を瞬時に把握する自動要約

[位野木万里 2022] 位野木万里, 技術文書の記載漏れを指摘する定量化自動要約

[位野木万里 2021] 位野木万里, 技術文書の理解を高速化する記述状況の自動要約技術

[YutaSaikawa2021] Yuta Saikawa, Takeshi Nagaoka, Takayuki Kitagawa, Mari Inoki Non-functional Requirements Summarization Method Using Convolutional Neural Network and Domain Ontology 大規模情報システムの要求仕様書の自動要約手法の提案-背景や目的部分に対するマルコフ連鎖を用いた抽出型要約の適用

[石井達樹 2020] 石井達樹, 阿部養太, 大工原正哲, 中島千壽, 位野木万里

[高橋宏季 2020] 高橋宏季, 野村典文, 近藤公久, 位野木万里, 要求仕様書中のアクター名の定義漏れパターンと組織変更がもたらす影響 - 実案件分析と得られた教訓 -

5. 主な発表論文等

〔雑誌論文〕 計13件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 中村雄太郎, 長岡武志, 北川貴之, 位野木万里, 本位田真一	4. 巻 Vol.41 No.3
2. 論文標題 Page Trendによる記述状況の可視化を用いた要求仕様書の理解手法の提案とtf-idfによる記述状況の比較評価	5. 発行年 2024年
3. 雑誌名 コンピュータソフトウェア	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nakamura Yutaro, Nagaoka Takeshi, Kitagawa Takayuki, Inoki Mari, Honiden Shinichi	4. 巻 -
2. 論文標題 Understanding Support Method for Requirements Specification Using Description Status Based on Page Trend	5. 発行年 2023年
3. 雑誌名 2023 8th International Conference on Information and Network Technologies (ICINT)	6. 最初と最後の頁 43-48
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICINT58947.2023.00016	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 中野丈, 島川遼太郎, 杉村康気, 宅間健生, 村野遼, 奥田博隆, 位野木万里	4. 巻 2024-SE-216
2. 論文標題 要求工学の知見をプロンプト設計に組み込んだ生成AIを活用した要求定義支援手法	5. 発行年 2024年
3. 雑誌名 情報処理学会, 研究報告ソフトウェア工学 (SE)	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 高橋宏季, 野村典文, 近藤公久, 位野木万里	4. 巻 Vol.11 No.2
2. 論文標題 要求仕様書中のアクター名の定義漏れパターンと組織変更がもたらす影響 - 実案件分析と得られた教訓 -	5. 発行年 2020年
3. 雑誌名 情報処理学会 デジタルプラクティス	6. 最初と最後の頁 367 - 388
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計48件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 位野木万里
2. 発表標題 技術文書の記述状況を瞬時に把握する自動要約
3. 学会等名 大学見本市2023～イノベーション・ジャパン
4. 発表年 2023年

1. 発表者名 中村雄太郎, 長岡武志, 北川貴之, 位野木万里
2. 発表標題 要求仕様書における話題の記述状況の可視化 -BERTを用いたPage Trend算出手法の提案-
3. 学会等名 情報処理学会 第85回全国大会 5L-05
4. 発表年 2023年

1. 発表者名 位野木万里
2. 発表標題 技術文書の記載漏れを指摘する定量化自動要約
3. 学会等名 イノベーション・ジャパン2022
4. 発表年 2022年

1. 発表者名 齋川祐太, 長岡武志, 北川貴之, 位野木万里
2. 発表標題 Convolutional Neural Networkとドメインオントロジーを用いた定量的視点による非機能要求の要約手法の提案
3. 学会等名 情報処理学会 第84回全国大会 7L-01
4. 発表年 2022年

1. 発表者名 Yuta Saikawa, Takeshi Nagaoka, Takayuki Kitagawa, Mari Inoki
2. 発表標題 Non-functional Requirements Summarization Method Using Convolutional Neural Network and Domain Ontology
3. 学会等名 The 20th International Symposium on Advanced Technology (ISAT-20) (国際学会)
4. 発表年 2021年

1. 発表者名 位野木万里
2. 発表標題 技術文書の理解を高速化する記述状況の自動要約技術
3. 学会等名 イノベーション・ジャパン2021
4. 発表年 2021年

1. 発表者名 阿部養太, 石井達樹, 大工原正哲, 中島千壽, 位野木万里
2. 発表標題 大規模情報システムの要求仕様書における自動要約技術の適用-自動要約結果と先入観との差異を活用した理解支援
3. 学会等名 情報処理学会第82回全国大会 (6K-02)
4. 発表年 2020年

1. 発表者名 石井達樹, 阿部養太, 大工原正哲, 中島千壽, 位野木万里
2. 発表標題 大規模情報システムの要求仕様書の自動要約手法の提案-背景や目的部分に対するマルコフ連鎖を用いた抽出型要約の適用
3. 学会等名 情報処理学会第82回全国大会 (6K-01)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

高信頼ソフトウェア開発工学研究室
<https://www.ns.kogakuin.ac.jp/relab/>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	本位田 真一 (Honiden Shinichi)		
研究協力者	中村 雄太郎 (Nakamura Yutaro)		
研究協力者	長岡 武志 (Nagaoka Takeshi)		
研究協力者	北川 貴之 (Kitagawa Takayuki)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------