

要求仕様書と設計書における語彙の差分に関する分析

若林 丈紘¹ 森崎 修司² 渥美 紀寿³ 山本 修一郎³

概要: ソフトウェアドキュメントのレビューにおける、要求仕様書と設計書の対応付けの候補となる単語を計算機支援により提示することを目指し、要求仕様書と基本設計書に現れる語彙の差分によって設計書で実現されていない要求を特定できるか分析した。まず、ある要求仕様書と基本設計書の組を対象に、基本設計書で出現しない単語の集合を取り出し、この単語を含む要求仕様書中の文の内容が基本設計書で実現されているかどうかを調べた。102語中20語の単語で、実現されていない要求を見つけることができた。また、語彙の差分から得られる候補について、基本設計書の章の分布にばらつきがあるかを調べた。要求仕様書に出現するが基本設計書の各章節を取り除くことで出現しなくなる単語の集合を取り出し、この単語を含む要求仕様書中の文の内容が基本設計書で取り除いた章節の内容と対応する割合を調べた。同一システムの要求仕様書と基本設計書4組を対象に単語の割合を調べた結果、対応付けができる単語の割合が高い章には共通点があることが分かった。

1. はじめに

ウォーターフォール型のソフトウェア開発では、要求定義書、基本設計書といったドキュメントが作成される。開発者は利用者の要求に基づいてこれらのドキュメントを作成し、その後ドキュメントに従ってコーディングやテストを行う。基本設計書は、その前段階で記述された要求事項を漏れなく満たしているかどうか、要求仕様書の中から内容を実現する部分を探し出す必要がある。多くのドキュメントは自然言語で記述されるため、これらは目視で評価することが一般的である。しかし、目視による評価は作業コストが大きいという問題がある。

目視による評価コストを小さくするために、自然言語処理を用いて計算機支援によりドキュメントを評価する手法を提案した研究がある [1]~[5]。文献 [1][4][5] では、あるドキュメントを流用・改変して作成したドキュメントの文章に矛盾がないか、目視で確認する作業を支援する。2つの異なるドキュメントの類似度を求めるために、単一の文、あるいは単語に分割して、一方のドキュメントの文に対応する候補を取り出す。取り出した候補を分析者に提示すること

により、対応付け作業を支援する。類似度はTF-IDF(Term Frequency-Inverse Document Frequency) といった手法を用いて求める。

文献 [1][2] では要求仕様書と設計書の共通語彙を使って、要求と設計のトレーサビリティを明らかにする支援をしている。しかし、ツールによる分析の対象は、要求仕様書と設計書のうち一部の文章に限定している。ドキュメントの文章全体についてトレーサビリティを明らかにすることは考慮していない。

本研究では、設計書で実現されるべき要求の候補となる単語を、計算機支援によって提示することを目指し、この単語を、要求仕様書の文章全体から取り出せるかどうかを確かめる。また、取り出した候補のうち実現されていない要求事項の割合に、基本設計書の章ごとの違いがあるかを次の手順で調べる。

- (1) ドキュメントから、要求仕様書に出現して設計書に出現しない単語を取り出す。取り出した単語を含む要求仕様書の文・図表を参照し、設計書に実現されている文・図表の割合を調べる。
- (2) ドキュメントから、ある章節を取り除くことで設計書に出現しなくなる単語を取り出す。この単語を含む要求仕様書の文・図表を参照し、取り除いた章節で実現されている文・図表の割合を調べる。

第2章では分析の手順を述べ、第3章では結果を述べる。第4章で分析結果を考察し、第5章でまとめる。

¹ 名古屋大学 工学部
Department of Electrical, Electronic Engineering and Information Engineering, school of Engineering Nagoya University, Japan

² 名古屋大学 大学院情報科学研究科
Department of Electrical, Electronic Engineering and Information Engineering, school of Engineering Nagoya University, Japan

³ 名古屋大学 情報連携統括本部 情報戦略室

表 1 対象ドキュメントと ID

ID	ドキュメント名	ページ数	単語数	単語種類	定義あり単語数	定義あり単語種類
R_1	雨量等防災情報提供システム 要求仕様書	12	2779	564	2460	491
R_2	図書館情報システム 基本計画書	45	16528	1466	12959	1167
R_3	住宅履歴情報管理システム 要件定義書	75	19712	1467	13434	1181
R_4	保証ケース統合作成支援ツール 要件定義書	5	388	83	380	79
A_1	雨量等防災情報提供システム 基本設計	58	9664	1142	6970	884
A_2	図書館情報システム 基本設計書	38	17916	2034	12258	1437
A_3	住宅履歴情報管理システム 外部設計書	136	28635	1697	21132	990
A_4	保証ケース統合作成支援ツール 基本設計書	14	1150	251	980	181

2. 分析

2.1 分析対象

対象ドキュメントは情報システムまたは単一ソフトウェアツールの要求仕様書、基本設計書とした。ドキュメントの一覧と、それらの規模や単語数を表 1 に示す。表中の R_i と A_i ($i = 1, 2, 3, 4$) は、それぞれ同一システムの要求仕様書と設計書である。

2.2 分析目的

2.2.1 分析 1 語彙差分の調査

要求仕様書で出現するが基本設計書で出現しない単語は、設計書で実現されなかった要求仕様書の記述内容の一部であると考えられる。そこで、同一システムの要求仕様書 R_i と基本設計書 A_i から、基本設計書に出現しない単語 $N_0(A_i)$ を取り出す。 $N_0(A_i)$ を含む要求仕様書の文・図表の内容が基本設計書に実現されているかを調べる。

2.2.2 分析 2 語彙差分の分布の調査

ある単語 n_k を含む文・図表は要求仕様書と基本設計書の両方に複数回出現することがある。複数回現れる文・図表が表す内容はそれぞれ異なることがある。そのため、前節で紹介した手法では、要求仕様書から単語を候補として取り出せたとしても、その候補の中から設計書において実現されていない要求を見つけられないことがある。

その例を図 1 に示す。図 1 の上部では、要求仕様書と基本設計書の両方において、“制限”という単語が同じ内容を表す文で出現している。どちらの文も“異常検知時のネットワーク利用制限”を表しており、要求仕様書の内容が設計書に実現されているといえる。一方、図 1 の下部では、“制限”という単語が要求仕様書と基本設計書の文の両方に出現するものの、その文が表す内容は異なる。この場合、“制限”という単語はどちらにも存在するため、“制限”という単語をもとにして“異常を検知した場合は回線の利用を制限すること”という要求を見つけ出すことができない。つまり、設計書に出現しない単語をもとに実現されていない要求を取り出せるかどうかは、設計書の文によって異なる。そこで、基本設計書 A_i からある章節を取り除き、新たに出てきた単語 $N_0(A_i)$ を含む要求仕様書の文・図表が、

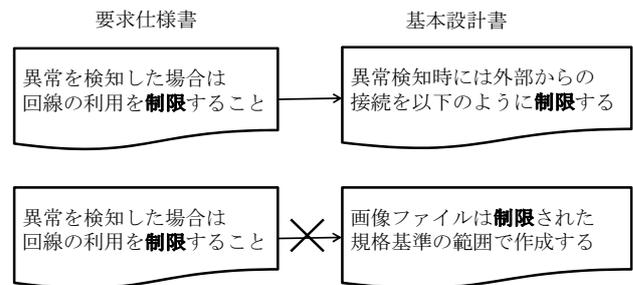


図 1 単語による要求仕様書と基本設計書の対応関係の例

取り除いた章節で実現されるべき内容であるかを調べる。

2.3 分析手順

2.3.1 準備

要求仕様書、基本設計書ともに同一の方法でドキュメントにおけるそれぞれの単語の出現割合を得る。

まず、あるドキュメント D から語彙を取り出す手順を示す。ドキュメント D は PDF 形式、または Word 形式で記録されている。PDF 閲覧アプリケーション、または Microsoft office Word により、ドキュメント D をテキストファイルとして保存する。保存したファイルを形態素解析ツール MeCab[6] に与え、MeCab が名詞と判断した単語の集合を得る。品詞体系は IPA 品詞体系日本語辞書 [7] に準ずる。MeCab で得られた単語の集合から、日本語語彙体系において定義のある単語の集合を得る。 $N_m(D)$ には“>”や“○”といった記号も含まれるため、これらを取り除くために日本語語彙体系での定義を用いた。日本語語彙体系に定義のある単語の集合から、単体で意味のないアルファベット 1 文字と数字を目視により取り除く。これにより、意味のない文字が取り除かれた単語の集合 $N(D) = \{n_1, \dots, n_k\}$ を得る。 $|N(D)|$ はドキュメント D に含まれる単語の種類数である。

また、ドキュメント D における単語 n_k の出現回数を $C_D(n_k)$ とし、 $C_D(n_k) / \sum C_D(n_i) (i = 1, \dots, q)$ を出現割合 $P_D(n_k)$ とする。

単語 n_k を含む文・図表を $s(n_k)$ とする。

2.3.2 分析 1

要求仕様書 R で現れる語彙 $N(R)$ のうち、同一システムの基本設計書 A で現れない単語 $N_0(A)$ を取り出す。 $N_0(A)$

を含む文・図表が基本設計書 A で実現されているか確認する。

調査の対象は、表 1 中の要求仕様書 R_1 、基本設計書 A_1 とする。

まず、 $N(R_1)$ 、 $N(A_1)$ から $P_{A_1}(n_k) = 0$ となる n_k の集合 $N_0(A_1) = \{n_{01}, \dots, n_{0q}\}$ を得る。次に、 n_{0q} を含む要求仕様書 R_1 中の文 $s(n_{0q})$ を読み、 $s(n_{0q})$ を実現している文・図表を基本設計書 A 中から目視で確認する。確認の結果、 $s(n_{0q})$ が実現されていない設計書の漏れが認められれば 1 を、それ以外の場合に 0 の値をとる $L(n_{0q})$ を得る。ひとつの単語 n_{0q} が複数の $s(n_{0q})$ を持つ場合は、 $s(n_{0q})$ のうち少なくとも 1 つの漏れが認められれば $L(n_{0q}) = 1$ とする。

要求仕様書 R_1 のうち、設計書で考慮されないことが明らかである章がある場合は、その章を取り除いて上記と同様に分析をする。

2.3.3 分析 2

基本設計書 A_i からある部分 A_{i_j} (章・節) を取り除き、元の A_i との差分の単語 $N_0(A_{i_j})$ を取り出す。要求仕様書 R_i の文 $s(n_{0p})$ が取り除いた章節で実現されるべき要求かどうかを調べる。これにより、前節で行った対応付けがどのような章で適用できるかを調べる。

まず、前節と同様に、 $N(R_i)$ 、 $N(A_i)$ から $P_{A_i}(n_k) = 0$ となる n_k の集合 $N_0(A_i)$ を得る。次に、基本設計書 A_i の章節 A_{i_j} を取り除くことで得られる残りの部分 $A_i - A_{i_j}$ をもとに $N_0(A_i - A_{i_j})$ と $N_0(A_i)$ の差集合、つまり $N_0(A_{i_j}) = \{n_{01}, \dots, n_{0p}\}$ を得る。取り除く章節は設計書の目次に記述されているものを選んだ。 n_{0p} を含む要求仕様書 R_i 中の文 $s(n_{0p})$ を読み、 $s(n_{0p})$ が基本設計書の章 A_j で実現されるべき要求であるかどうかを目視で確認する。確認の結果、 $s(n_{0p})$ が A_{i_j} で実現されるべき要求である単語 n_{0p} の集合を $N_M(A_{i_j})$ とする。

章節 A_{i_j} を取り除いた時に A_i に出現しなくなる語の種類数 $|N_0(A_{i_j})|$ あたりの $|N_M(A_{i_j})|$ の割合 $Mp_{A_{i_j}}$ を得る。また、章節 A_{i_j} に含まれる語の種類数 $|N(A_{i_j})|$ あたりの $|N_M(A_{i_j})|$ の割合 $Mp'_{A_{i_j}}$ を得る。 $Mp_{A_{i_j}}$ が 0.5 以上、または $Mp'_{A_{i_j}}$ が A_i の平均値以上である章節 A_{i_j} を、 $N_0(A_{i_j})$ による対応付けがしやすい章節とみなし、記述されている内容を調べる。 $Mp_{A_{i_j}}$ が 0.5 未満かつ $Mp'_{A_{i_j}}$ が A_i の平均値未満である章節 A_{i_j} についても、同様に記述されている内容を調べる。

3. 結果

3.1 分析 1

$N_0(A_1)$ を含む要求仕様書の文・図表のうち、基本設計書に実現されていないものが 102 語中 20 語であった。また、 R_1 のうち設計書で考慮されないことが明らかである 1 章を取り除いたところ、 $L(n_{0q}) = 1$ の単語は $N_0(A_1)$ 46 語

表 2 R_1 における $L(n_{0q}) = 1$ である単語 n_{0q} の数

$L(n_{0q})$	$ N_0(A_1) $
1	20
0	82

のうち 20 語であった。

見つかった $N_0(A_1)$ 102 語について、 $s(n_{0q})$ を確認し、 $L(n_{0q}) = 1$ となる単語 n_{0q} の語数を調べた。結果を表 2 に示す。表 2 の通り、 $L(n_{0q}) = 1$ となる単語は 20 語あった。見つけた記述の中には、“障害対応についての経費は運用経費に含む”といった、設計書とは異なる開発ドキュメントでの実現を想定している可能性のある要求もあった。

R_1 において、 $s(n_{0q})$ の出現箇所には章によって偏りがあった。設計書で実現されるべき要求は実現の有無に関わらず、すべて 2 章に出現していた。 R_1 の 1 章はシステムの構築をする目的や契約内容、構築業務を進める上での計画が記述されていた。2 章はシステムの機能と取り扱うデータの仕様が記述されていた。1 章を取り除いたとすると、 $N_0(A_1)$ 46 語のうち $L(n_{0q}) = 1$ の単語は 20 語である。この場合、 $L(n_{0q}) = 1$ の単語は $N_0(A_1)$ 全体の約 43% であり、 $N_0(A_1)$ 全体に占める割合が大きくなる。

3.2 分析 2

$Mp_{A_{i_j}}$ 、 $Mp'_{A_{i_j}}$ は、設計書の章節ごとにそれぞれ異なっていた。 $Mp_{A_{i_j}}$ とは、元の A_i との差分の単語 $N_0(A_{i_j})$ あたりの、 A_{i_j} で実現されるべき要求 $N_M(A_{i_j})$ の割合である。 $Mp'_{A_{i_j}}$ とは、 A_{i_j} に出現する単語 $N(A_{i_j})$ あたりの、 $N_M(A_{i_j})$ の割合である。また、 $Mp_{A_{i_j}}$ が 0.5 以上、あるいは $Mp'_{A_{i_j}}$ が平均値以上である章は、データ項目、システム運用、セキュリティ要件について記述されていた。 $Mp_{A_{i_j}}$ が 0.5 未満かつ $Mp'_{A_{i_j}}$ が平均値未満である章節は性能要件が記述されていた。

要求仕様書と基本設計書の組 R_i 、 A_i ($i = 1, 2, 3, 4$) の各章節について、 $N_0(A_{i_j})$ を得た。また、取り除いた章節 A_{i_j} で実現されるべき $N_0(A_{i_j})$ の割合を調べた。結果を表 3, 4, 5, 6 に示す。表 3, 4, 5, 6 の $|N_0(A_{i_j})|$ は、 A_{i_j} を取り除いたときに設計書に出現しなくなり、かつ要求仕様書に出現する単語の数である。 $|N_M(A_{i_j})|$ は $N_0(A_{i_j})$ のうち抜いた章節 A_{i_j} で実現される要求文に含まれる語彙の数である。 $Mp_{A_{i_j}}$ 、 $Mp'_{A_{i_j}}$ はそれぞれ $N_0(A_{i_j})$ 、 $N(A_{i_j})$ あたりの $N_M(A_{i_j})$ の割合である。 $|N_0(A_{i_j})| = 0$ の章節については、 $Mp_{A_{i_j}}$ 、 $Mp'_{A_{i_j}}$ ともに値を得られないため、表には“-”と表記する。

4 件の基本設計書とも、 $|N_0(A_{i_j})| = 0$ 、 $Mp_{A_{i_j}}$ が 0.5 以上の節があった。 A_4 は $|N_0(A_{i_j})| = 0$ である章節が 7 あり、全体で 13 ある章節の半分以上を占めていた。 A_4 はソフトウェアツールの設計書であるため、ある節で使われる文章表現や設計書に固有の用語が、隣接する節で多く使わ

表 3 A_1 の各章における対応付けできた語の割合

章 (j)	$ N_0(A_{1j}) $	$ N_M(A_{1j}) $	$Mp_{A_{1j}}$	$M'p_{A_{1j}}$
1.1.1	5	3	0.60	0.030
1.1.2	8	5	0.63	0.024
1.1.3	0	0	-	-
1.1.4	6	2	0.33	0.011
1.1.5	1	0	-	0
1.1.6	14	10	0.71	0.046
1.2.1	0	0	-	0
1.2.2	0	0	-	0
1.2.3	0	2	-	0
1.2.4	0	1	-	0
2.1.1	15	9	0.60	0.0363
2.1.2	4	3	0.75	0.024
2.1.3	0	0	-	-
2.2.1	6	2	0.33	0.015
2.2.2	3	3	1.00	0.022
2.2.3	1	0	-	-
3	13	6	0.46	0.026
4.1	0	0	-	-
4.2	3	1	0.33	0.008
4.3	3	3	1.00	0.055

れていた。

$Mp_{A_{ij}}$ が 0.50 以上, または $M'p_{A_{ij}}$ が A_i の平均値以上である章 A_{ij} の一覧を表 7 に示す. A_1 について, 1.1.1 節はシステムで扱う情報の種類, 1.1.6 節はネットワーク, 2.1.1 節はデータ処理部の設計, 3 節はシステムの運用に必要な事項, 4.3 節は RASIS に関するソフトウェア作成の注意点が記述されていた. A_2 について, 1.1 節, 1.2 節はシステムの目的, 3.2 節はデータ項目の一覧, 4.1 節は旧システムからの移行方法, 6.2 節は運用業務の内容, 7.3 節は冗長化, 多重化の仕様と範囲, 7.4 節はセキュリティ対策, 8.1 節はシステム構築の実施計画が記述されていた. A_3 について, 1 章はシステムが業務に対応する範囲, 2 章はデータ項目の一覧, 4.1 節は機能一覧, 6.1 節は平時の運用方法が記述されていた. A_4 について, 3.2.2 節はツールの機能が記述されていた. これらの章のうち, データ項目, システム運用, セキュリティ要件について記述された章が A_1 , A_2 , A_3 で共通して得られた.

また, $Mp_{A_{ij}}$ が 0.50 未満かつ $M'p_{A_{ij}}$ が A_i の平均値未満である章 A_{ij} の一覧を表 8 に示す. 表 8 に示す章のうち, 性能要件について記述された章が A_1 , A_2 , A_3 で共通して得られた.

4. 考察

4.1 分析 1

基本設計書で出現しない単語をもとに要求仕様書を読むことで, 基本設計書に実現されていない記述を見つけることができた. 見つけた記述の中には, 設計書とは異なる開発ドキュメントにおいて実現する可能性がある要求もあっ

表 4 A_2 の各章における対応付けできた語の割合

章 (j)	$ N_0(A_{2j}) $	$ N_M(A_{2j}) $	$Mp_{A_{2j}}$	$M'p_{A_{2j}}$
1.1	9	9	1.00	0.24
1.2	10	8	0.80	0.12
1.3	0	0	-	-
2	3	0	0	0
3.1	3	0	0	0
3.2	29	18	0.62	0.059
4.1	34	29	0.85	0.12
4.2	15	0	0	0
4.3	1	0	0	0
5.1	5	0	0	0
5.2	12	4	0.33	0.021
6.1	3	0	0	0
6.2	22	11	0.50	0.031
6.3	16	2	0.13	0.008
7.1	13	1	0.08	0.003
7.2	0	0	-	-
7.3	2	1	0.50	0.011
7.4	3	2	0.67	0.015
7.5	15	1	0.07	0.006
7.6	10	2	0.20	0.014
8.1	6	4	0.67	0.039
8.2	6	0	0	0

表 5 A_3 の各章における対応付けできた語の割合

章 (j)	$ N_0(A_{3j}) $	$ N_M(A_{3j}) $	$Mp_{A_{3j}}$	$M'p_{A_{3j}}$
1	4	2	0.50	0.047
2	100	59	0.59	0.13
3	9	3	0.33	0.021
4.1	34	24	0.71	0.10
4.2	6	0	0	0
5	0	0	-	-
6.1	8	6	0.75	0.11
6.2	7	3	0.43	0.041
6.3	24	8	0.33	0.072
7	25	10	0.40	0.034

表 6 A_4 の各章における対応付けできた語の割合

章 (j)	$ N_0(A_{4j}) $	$ N_M(A_{4j}) $	$Mp_{A_{4j}}$	$M'p_{A_{4j}}$
1	1	0	0	0
2	4	1	0.25	0
3.1	0	0	-	-
3.2.1	2	0	0	0
3.2.2	1	1	1.00	0.17
3.2.3	0	0	-	-
3.2.4	0	0	-	-
3.3	0	0	-	-
3.4.1	1	0	0	0
3.4.2	0	0	-	-
3.5.1	1	0	0	0
3.5.2	0	0	-	-
3.5.3	0	0	-	-

表 7 $Mp_{A_{ij}}$ が 0.50 以上, または $Mp'_{A_{ij}}$ が平均値以上の章節

章 (j)	章のタイトル	Mp_{A_j}
$A_{1.1.1}$	システムで取り扱う情報	0.60
$A_{1.1.6}$	放送事業者等情報利用者との接続方法	0.71
$A_{1.2.1.1}$	県管理のデータ	0.60
$A_{1.3}$	運用管理の検討	0.60
$A_{1.4.3}$	ソフトウェアの機能要件	1
$A_{2.1.1}$	図書館情報システム再構築の方向性	1
$A_{2.1.2}$	方針実現のための方策	0.80
$A_{2.3.2}$	主要データ項目一覧	0.62
$A_{2.4.1}$	システム移行	0.81
$A_{2.6.2}$	システム運用管理	0.50
$A_{2.7.3}$	セキュリティ要件	0.67
$A_{2.7.4}$	アベイラビリティ要件	0.50
$A_{2.8.1}$	マスタスケジュール	0.81
$A_{3.1}$	システム概要	0.50
$A_{3.2}$	データ設計	0.59
$A_{3.4.1}$	機能一覧	0.71
$A_{3.6.1}$	運用設計	0.75
$A_{3.6.3}$	セキュリティ	0.33
$A_{4.3.2.2}$	分解	1

表 8 $Mp_{A_{ij}}$ が 0.50 より小さい, かつ $Mp'_{A_{ij}}$ が平均値より小さい章節

章 (j)	章のタイトル	Mp_{A_j}
$A_{1.1.4}$	システムからの情報自動取得方法	0.33
$A_{1.1.5}$	データ提供により放送事業者でできること	0
$A_{1.2.2.1}$	防災情報ステーション	0.33
$A_{1.2.2.3}$	土砂災害通報システム	0
$A_{1.4.2}$	機器仕様 (ハードウェアの性能要件)	0.33
$A_{2.2}$	図書館情報システム要求機能定義	0
$A_{2.4.2}$	データ移行	0
$A_{2.4.3}$	本番移行スケジュール	0
$A_{2.5.1}$	外部インタフェースの定義	0
$A_{2.6.1}$	システム運用フロー	0
$A_{2.6.3}$	システム維持管理	0.13
$A_{2.7.1}$	性能要件	0.08
$A_{2.7.2}$	データ要件	0
$A_{2.7.5}$	システム構成設計	0.07
$A_{2.7.6}$	ネットワーク構成設計	0.20
$A_{2.8.2}$	実施体制	0
$A_{3.3}$	業務設計	0
$A_{3.4.2}$	画面一覧	0
$A_{3.6.2}$	信頼性、性能、スケーラビリティ	0.43
$A_{3.6.3}$	セキュリティ	0.33
$A_{3.7}$	画面イメージ	0.40
$A_{4.1}$	はじめに	0
$A_{4.2}$	参考資料	0
$A_{4.3.2.2}$	プロジェクト	0
$A_{4.3.4.1}$	メニュー	0
$A_{4.3.5.1}$	構成定義	0

た。しかし、設計書で考慮されない記述だとしても、設計書の記述漏れが見つかることは他の開発工程における漏れの事前防止やドキュメントの見直しに繋がる。このため、設

計書で考慮されない要求もまた、ドキュメントのレビューにおいて価値があると考ええる。

R_1 において、設計書で考慮されないことが明らかである 1 章を取り除いたところ、 $L(n_{0q}) = 1$ の単語は $N_0(A)$ 全体の約 43% であった。設計書で考慮されないことが明らかである章が要求仕様書にある場合は、この章をあらかじめ取り除くことで、取り出した候補のうち $L(n_{0q}) = 1$ である単語の割合を高めることができる。

4.2 分析 2

4.2.1 対応付けをできる対象ドキュメント

A_4 はソフトウェアツールの設計書であるため、ある節で使われる文章表現や設計書に固有の用語が、隣接する節で多く使われていた。これが、 A_4 で $|N_0(A_{ij})| = 0$ の章節が多かった理由であると考えられる。 $|N_0(A_{ij})| = 0$ である章節が設計書に多い場合、本研究の手法では、設計書で要求を漏れなく実現しているか判断することが難しい。よって、本研究の手法は、内容が異なる章節を多く含む情報システムのドキュメントに対して行うことが望ましいと考えられる。

4.2.2 N_0 を用いて対応付けできる割合の高い章節

要求仕様書 R_i と基本設計書 A_i ($i=1, 2, 3, 4$) の組において、 $Mp_{A_{ij}}$ が 0.50 以上または $Mp'_{A_{ij}}$ が A_i の平均値以上である章 A_{ij} を得られた。これらの章のうち、 A_1, A_2, A_3 で共通して、データ項目、システム運用、セキュリティ要件が記述された章が得られた。これは、上記の章がそれぞれその章に唯一である単語を多く含むためであると考えられる。例えば、運用に関する章では“可用性”、“保守”、“障害”といった単語が存在した。これらの単語は要求仕様書から基本設計書で実現される過程で別の単語に置き換わる可能性が低い。

一方、4 件の基本設計書において、 $Mp_{A_{ij}}$ が 0.50 未満かつ $Mp'_{A_{ij}}$ が A_i における平均値未満である章 A_{ij} を得られた。これらの章のうち、 A_1, A_2, A_3 で共通して、性能要件が記述された章が得られた。これは、要求仕様書における“遅延なく動作する性能”といったあいまいな記述が、基本設計書で具体的な性能の値や機器名に置き換わるためだと考えられる。

n_{0p} を含む $s(n_{0p})$ を候補として、設計書で実現されていない記述を要求仕様書から見つけないときは、データ項目、システム運用、セキュリティ要件の章について特に大きな効果が得られると考えられる。性能要件が記述された章については、候補を取り出した際に、設計書で実現されているかを確認するのが難しいことをレビューアがあらかじめ意識しておくことで、より短時間で作業できる可能性がある。

4.2.3 妥当性

同一ソフトウェアの要求仕様書と基本設計書の両方が公

開されているものがほとんど見つからなかった。このため、章節の除外による検証結果の比較対象は4組のドキュメントのみとした。より多くの要求仕様書と基本設計書の組を対象として、本論文の結果と同様の傾向があるか確かめることは、今後の課題の1つである。

本論文では、1つの単語 n_{0q} に対して複数の $s(n_{0q})$ が現れた場合、 $s(n_{0q})$ が少なくとも1つ基本設計書で実現されるならば n_{0q} は実現されているものとしてみなした。ある要求仕様書で同じ単語が異なる文脈で使われることは頻繁に起こり得ることである。

しかし、 $s(n_{0q})$ の数が多いとしても、レビューの作業コストが増えることはないと考えられる。その理由は、本研究が想定するレビューの目的が、レビューアが設計書の記述漏れに気付くことであるからだ。この場合、レビューアは候補として取り出した要求仕様書内の文を必ずしも全て設計書と照らし合わせて確認する必要はない。設計書に実現したことをレビューアが把握している要求 $s(n_{0q})$ については、レビューアは読み飛ばすことができる。このため、実際のレビューでは $s(n_{0q})$ の数が多いとしても、作業コストは増えることはないと考えられる。

分析では、本研究が目指す対応付け候補の自動生成を想定した。そのため、IPA品詞体系辞書を使った形態素解析器 MeCab で名詞と判定され、日本語語彙体系 [8] に定義のある単語を対象とした。日本語語彙体系は14万語を収録した辞書であるが、“ファイアウォール”、“DMZ”といった情報システムのドキュメントで出現頻度の高い単語が載っておらず、単語として取り出すことができなかった。

前節の通り、基本設計書の章のうち、その章で実現されるべきである要求事項の割合が大きい章には共通点があった。それは、要求仕様書の単語が設計書で実現される際に、別の単語に置き換わる可能性の低い専門用語を多く含んでいることである。このため、幅広い分野、特に情報系の専門用語を網羅する辞書を使用することで、実現されていない要求の候補をより多く得られると考えられる。

5. おわりに

計算機支援によって、設計工程で実現されるべき要求の候補となる単語を要求仕様書の文章全体から取り出せるかどうかを確かめるために分析を実施した。

まず、情報システムの要求仕様書 R_1 と基本設計書 A_1 を対象に、要求仕様書で出現するが基本設計書で出現しない単語 $N_0(A_1)$ を取り出した。この単語を含む文・図表に記述される要求事項を設計書で実現しているかどうか調べたところ、取り出した単語102語のうち20語で要求が実現されていないことが分かった。また、要求を実現する記述がないことが明らかである章が設計書にある場合、これをあらかじめ取り除くことで、取り出した候補のうち要求が実現されていない単語の割合を43%に高めることができ

た。単語 $N_0(A_1)$ を用いることにより、設計書で実現されていない要求事項の候補を計算機支援により提示できる可能性が示された。

次に、4組の要求仕様書と基本設計書を対象に、基本設計書 A_i の各章節を取り除くことで出現しなくなる単語 $N_0(A_{i_j})$ を取り出した。この単語を含む文・図表に記述される要求事項が設計書の取り除いた章節 A_{i_j} で実現されるべき内容であるか調べた。設計書のある章節 A_{i_j} で実現されるべき要求事項に含まれる単語 $N_M(A_{i_j})$ について、章節 A_{i_j} を取り除いた時に A_i に出現しなくなる語の種類数 $|N_0(A_{i_j})|$ あたりの割合 $Mp_{A_{i_j}}$ と、章節 A_{i_j} に含まれる語の種類数 $|N(A_{i_j})|$ あたりの割合 $Mp'_{A_{i_j}}$ は、章節ごとにそれぞれ異なっていた。データ項目、システム運用、セキュリティ要件が記述された章節において $Mp_{A_{i_j}}$ 、あるいは $Mp'_{A_{i_j}}$ が大きくなった。また、性能要件が記述された章節において $Mp_{A_{i_j}}$ と $Mp'_{A_{i_j}}$ がともに小さくなった。ある単語が要求仕様書から設計書にかけて別の単語に置き換わらない要求事項については、取り出した単語 $N_0(A_{i_j})$ により設計書に実現されているかどうかを判断できることが期待される。

参考文献

- [1] J.H.Hayes, A.Dekhtyar, S.Sundaram, Advancing Candidate Link Generation for Requirements Tracing: The Study for Methods, IEEE Transactions on Software Engineering, vol. 32, no. 1, pp. 4-19(2006)
- [2] P.Heck, A. Zaidman, Horizontal Traceability for Just-In-Time Requirements: The Case for Open Source Feature Requests, Journal of Software Evolution and Process, vol. 26, no. 12, pp. 1280-1296(2014)
- [3] G. Lami, QuARS, A Tool for Analyzing Requirement, Carnegie Mellon University Technical Report, CMU/SEI-2005-TR-014(2005)
- [4] 竹中要一, 若尾岳志, 地方自治体の例規比較に用いる条文対応表の自動生成, 自然言語処理, vol. 19, no. 3, pp. 193-212. (2012)
- [5] 丹治広樹, 山本和英, 保険約款と派生書類の自動対応付け, 言語処理学会 第17回年次大会発表論文集, pp. 868-871 (2011)
- [6] MeCab: Yet Another Part-of-Speech and Morphological Analyzer. , <http://mecab.sourceforge.net/>.
- [7] IPA品詞体系日本語辞書 IPAdic ver.2.7.0, 奈良先端科学技術大学院大学 松本研究室, <http://chasen.naist.jp/stable/ipadic/>
- [8] 池原 悟, 宮崎 正弘, 白井 諭, 横尾 昭男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦, 日本語語彙体系 CD-ROM 版, 岩波書店 (1999)