

# 共通語彙によるソフトウェア設計書の評価にむけた分析

奈加 大樹<sup>1,a)</sup> 森崎 修司<sup>1</sup> 渥美 紀寿<sup>1</sup> 山本 修一郎<sup>1</sup>

受付日 2015年7月24日, 採録日 2016年2月8日

**概要:** ソフトウェア設計書を目視を要さず網羅的に評価することを目指し, 14 件の基本設計書 (A 群), 14 件の要求仕様書 (R 群), 3 組の同一ソフトウェアの基本設計書と要求仕様書の組 (AR 群) を用いて実証的に評価した. 具体的には, 単語の最低出現頻度を変化させながら, A 群の基本設計書の間で共通語彙が存在するかを確かめた. 次に, AR 群の基本設計書において対となる要求仕様書よりも共通語彙が大きい割合で含まれるか確かめた. 評価の結果, 最低出現頻度を 0.05%, 0.10%, 0.15% に設定した場合でも共通語彙が存在した. 共通語彙に 10 単語以上が含まれる場合には, AR 群の基本設計書のほうが対になる要求仕様書よりも大きい割合で共通語彙を含むことが分かった. 基本設計書の評価を, 対になる要求仕様書に含まれる単語と共通語彙を用いて, 実施できる可能性が示唆された.

キーワード: ドキュメント定量化, 自然言語処理, 設計書

## Toward Software Design Document Evaluation with Common Terms

DAIKI NAKA<sup>1,a)</sup> SHUJI MORISAKI<sup>1</sup> NORITOSHI ATSUMI<sup>1</sup> SHUICHIROU YAMAMOTO<sup>1</sup>

Received: July 24, 2015, Accepted: February 8, 2016

**Abstract:** Toward automated comprehensive evaluation of software design document, we conducted an empirical evaluation with fourteen software architecture design documents (group A), fourteen requirement specification documents (group R), and three sets of architecture design document and requirement specification document of the same software (group AR). The evaluation investigates whether software architecture design documents (group A) have common terms, and whether each architecture design document includes larger percent of the common terms than the corresponding requirement specification in three sets of documents AR. The results of the evaluation show that common terms among the group A documents exist with minimum term frequency thresholds 0.05%, 0.10%, and 0.15%. The results also show that each architecture design document includes larger percentage of the common terms that have more than 10 terms. The results indicate that common terms have potential to evaluate level of fineness of description of software architecture design document.

**Keywords:** document metrics, natural language processing, software design document

### 1. はじめに

ウォーターフォール型ソフトウェア開発では, 要求仕様書, 基本設計書といったソフトウェアドキュメントを作成することで, 利用者の要求をより詳細に記述し, 段階的に詳細化する. 作成したソフトウェアドキュメントをもとにコーディングし, テストする. テストではプログラムを動作させながら期待どおりにコーディングできているかどうかを評価する. 多くのソフトウェアドキュメントは自然言語で

記述され, 実行可能な形式で記述されていないため, 動作させながら評価ができない. そのため, ソフトウェアドキュメントを目視で評価することが一般的である [4], [14]. 評価では, 誤りや抜けがないかといった欠陥がないか確認するとともにコーディングに必要な情報が記述されているかといったドキュメントの具体化度合いも確かめ, コーディング以降での開発計画に影響を与えないかどうかも確認する.

ソフトウェアドキュメントの目視による評価は一般的な手法であるが, 実施コストが大きいという問題がある. そこで, 自然言語処理を用いた計算機支援によるソフトウェアドキュメント評価手法を提案した研究がある [1], [3], [7], [8], [9], [10]. 文献 [7], [8] では, 要求仕様書と

<sup>1</sup> 名古屋大学  
Nagoya University, Nagoya, Aichi 464-8601, Japan  
<sup>a)</sup> naka.daiki@j.mbox.nagoya-u.ac.jp

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。これらの研究では要求仕様書と設計書を単一、または、少数の文に分割し、要求仕様書の文を実現した設計書の文を分析者が目視により対応づけることにより、トレーサビリティを明らかにする。具体的には要求仕様書の文に対応する設計書の文の候補を文に含まれる単語の類似度から求め、分析者に提示することにより、対応づけ作業を支援する。類似度はTF-IDF (Term Frequency-Inverse Document Frequency) といった特徴的な共通語彙を抽出する手法を用いて求める。要求仕様書と設計書の間のトレーサビリティが計算機支援により得られれば、要求仕様書に記述されている項目が設計書で実現されているかを調べたり、設計書の具体化度合いを知ったりすることができる。しかし、そうした類似度による設計書の文の候補には誤りも含まれているため、分析者の目視による判断を前提としている。

本研究では目視を要さず設計書の具体化度合いを評価することを目指す。設計書に含まれる共通の語彙があると仮定し、共通語彙を含むかどうかによって設計書の具体化度合いを評価できるか実証的に評価する。また、共通語彙が設計書の一部に偏在しているかどうかを確かめる。本論文では、まず、インターネットで公開されているソフトウェアドキュメントを対象とし、要求仕様書だけが公開されている群  $R$ 、基本設計書だけが公開されている群  $A$ 、両方公開されている群  $AR$  とに分ける。ドキュメント群  $A$  に共通語彙が存在するかどうかを確かめる。共通語彙が存在した場合には、ドキュメント群  $AR$  において、基本設計書、要求仕様書のそれぞれにおいて、共通語彙が含まれる割合を調べる。ドキュメントの共通語彙の割合が基本設計書と要求仕様書で異なり、対になる基本設計書と要求仕様書を与えられたときに、どちらが基本設計書かを区別できる場合には共通語彙による具体化度合いの評価が計算機支援により実現できる可能性があるとする。

以降、2章では関連研究をあげる。3章では分析の手順を述べ、4章で結果を述べる。5章で分析結果を考察し、6章でまとめる。

## 2. 関連研究

文に含まれる特徴的な共通語彙を用いて、ソフトウェアドキュメント、ソースコード、バグレポートを分析し、作業支援につなげようとする研究がある [1], [3], [6], [7], [8], [11], [13], [15]。文献 [7], [8] では、特徴的な共通語彙を共有する要求仕様書の文と設計書の文を求め、分析者が要求仕様書と設計書のトレーサビリティを明らかにする支援をしている。特徴的な共通語彙はTF-IDFを用いて求める。要求仕様書の文に対応する設計書の文のトレーサビリティが高い精度で求められれば、設計書の具体化度合いを知ることができる。しかし、現時点では必ずしも高い精度が得られるわけ

ではない。そのため、文献 [7] のように分析者の支援を前提としているため、本研究で目指す目視を要さない評価は実現できない。

文献 [11] では、6件のオープンソースソフトウェアを対象として、バグレポートの語彙とソースコードの語彙を分析している。ソースコードの語彙は、変数名、メソッド名、メソッド引数名、コメントから得る。分析の結果、バグレポートとソースコードの語彙が共通していることを示している。特に、あるバグレポートの語彙と対応するバグ修正のためのソースコードパッチの語彙が共通していることを報告し、バグレポートが報告された段階で、バグレポートの語彙から対応するソースコードの修正箇所を予測できる可能性を示している。

文献 [3], [6], [13], [15] では、大量のバグレポートの中で重複するバグレポートを検出し、バグ修正者を支援するための手法が提案されている。また、文献 [1] では、Issue Tracking System に登録された要求とバグレポートを区別しようとしている。共通語彙 (TF-IDF) を用いてドキュメントの分析を行う点は本研究と共通しているが、本研究では、設計書の具体化度合いを評価することを目的としている点で異なる。

目視によらずドキュメントを評価するツールや手法が提案されている [2], [9], [10]。QuARS [9] はNASAで開発された要求ドキュメントの評価ツール群である。QuARSでも目視によらないドキュメント評価を目指しているが、具体化度合いに関する言及がない点、個々の文章を分析する点で本研究とは異なる。具体的には、shall, must, should といった要求を表す表現 (imperative) 数、figure, table, for example といった具体例を指す表現 (directive), as appropriate, but not limited to といった weak phrase を事前に定義しておき、ドキュメントに含まれるそれらの数を用いて対象ドキュメントの品質を定量的に評価している。Lehner もドキュメントを構成する1文1文の読みやすさを計測し、目視によらないドキュメント評価に取り組んでいる [10]。しかし、評価の方法は、1文あたりの単語数といった読みやすさに関するものであり、具体化度合いに関する言及はない。

Chantree らは一般的な文章から作成したコーパスと比較することにより、誤解を招く可能性のある曖昧さを自動検出する方法を提案している [2]。文献 [2] では “Display categorized instructions and documentation.” を例としてあげ、categorized が instructions だけを説明しているのか documentation を含めて説明しているのか曖昧である点を指摘し、こうした曖昧さを検出する6種類の方法を比較している。Chantree らの方法では、and, or で接続される単語の曖昧さのみを対象としており、本論文で対象とするようなドキュメント全体にわたった評価は難しい。

### 3. 分析

#### 3.1 対象ドキュメント

対象ドキュメントはインターネットで公開されている情報システムの基本設計書、要求仕様書とした。記述言語は日本語とし、検索エンジンに“基本設計書”、“要求仕様書”といった検索キーワードを与え、検索対象のファイル形式としてPDF形式を指定し、検索結果に含まれたものから選んだ。ページ数の少ないドキュメントの場合、分量が少ないために共通語彙を含まない場合があると考え、基本設計書は30ページ以上のもの、要求仕様書は10ページ以上のものとした。ドキュメントの名称と本論文の以降で使うIDとの対応を表1に、各ドキュメントの規模や単語数を表2に示す。また、各ドキュメントを取得したURLを付録に示す。表中の  $A_a$  と  $R_a$ ,  $A_b$  と  $R_b$ ,  $A_c$  と  $R_c$  は

同一ソフトウェアの基本設計書と要求仕様書である。これら3組のドキュメントをAR群と呼ぶ。また、基本設計書  $A_1, \dots, A_{14}$  (基本設計書群  $A$ ) は対応する要求仕様書が検索エンジンではみつからなかった基本設計書である。要求仕様書  $R_1, \dots, R_{14}$  (要求仕様書群  $R$ ) も同様に対応する基本設計書がみつからなかった要求仕様書である。 $A_1$  と  $R_1$  は異なるシステムの基本設計書と要求仕様書である。AR群には  $A$  群、 $R$  群のドキュメントは含まれない。

#### 3.2 分析目的

##### 3.2.1 基本設計書における共通語彙

どのような利用用途のソフトウェアであっても、設計書の内容をもとにコーディングするためには、“入力”、“出力”、“表示”、“パラメータ”といった単語が必要になると考えられる。そこで、利用用途の異なる14件のソフトウェアの基

表1 対象ドキュメント名称とID  
Table 1 Titles and IDs of the documents.

ID	ドキュメント名
$A_1$	見守り情報管理システム基本設計書
$A_2$	消防業務支援システム基本設計書
$A_3$	年金業務システム基本設計書
$A_4$	汎用受付システム基本設計書
$A_5$	山梨県共同システム基本設計書
$A_6$	新山梨県立図書館情報システム基本設計書
$A_7$	JDR 医療チーム登録者管理システム基本設計書 (案)
$A_8$	病名検索システム基本設計書
$A_9$	雨量等防災情報提供システム基本設計書
$A_{10}$	多種文字コードと文字情報基盤文字コードとのコード変換ライブラリを介したデータ連携基本設計書
$A_{11}$	NEDO 旧保有鉱区管理システム基本設計書
$A_{12}$	Net Commons 2.0 WEKO Module 基本設計書
$A_{13}$	熊本市総合行政情報システム共通基盤システム基本設計書
$A_{14}$	日医標準レセプトソフト基本設計書
$A_a$	住宅履歴情報管理システム外部設計書
$A_b$	外部設計書制御系におけるセキュリティ機能共通基盤の開発
$A_c$	図書館情報システム基本設計書
$R_1$	自動車保有関係手続のワンストップサービスシステム要件定義の概要
$R_2$	次期 PARTNER システム要件定義書
$R_3$	旅費等内部管理業務共通システム (うち物品管理システム) 要件定義書 (案)
$R_4$	川口市共通基盤システム基本要件定義書
$R_5$	市川市共通基盤システム要件定義書
$R_6$	沖縄県保安林情報管理システム構築委託業務要件定義書及び仕様書
$R_7$	長野市就職情報システム要件定義書
$R_8$	周産期救急情報システム及び救急医療情報システムに関する共通業務支援システム要求定義書
$R_9$	電子図書館用電子計算機システム仕様書
$R_{10}$	つがる 西北五広域連合総合医療情報システム要求仕様書
$R_{11}$	佐賀中部広域連合地域包括支援センターシステム要件定義書
$R_{12}$	安全運転支援システム (DSSS レベル*) システム定義書
$R_{13}$	雨量等防災情報提供システム要求仕様書
$R_{14}$	福島赤十字病院医療基幹システム基本要件仕様書
$R_a$	住宅履歴情報管理システム要件定義書
$R_b$	要件定義書制御系におけるセキュリティ機能共通基盤の開発
$R_c$	図書館情報システム基本計画書

表 2 対象ドキュメントの詳細  
Table 2 Sizes of the documents.

ID	種別	ページ数	名詞数	名詞種類	定義あり名詞数	定義あり名詞種類
A <sub>1</sub>	基本設計書	57	9,592	1,367	6,271	575
A <sub>2</sub>	基本設計書	327	80,679	2,495	64,385	1,857
A <sub>3</sub>	基本設計書	71	14,245	1,424	11,319	972
A <sub>4</sub>	基本設計書	76	14,168	1,687	11,012	1,072
A <sub>5</sub>	基本設計書	56	12,579	1,336	9,000	774
A <sub>6</sub>	基本設計書	67	10,638	1,726	8,361	1,079
A <sub>7</sub>	基本設計書	44	6,112	746	4,684	471
A <sub>8</sub>	基本設計書	118	17,576	1,300	11,372	483
A <sub>9</sub>	基本設計書	60	7,020	1,088	5,437	704
A <sub>10</sub>	基本設計書	45	7,447	835	4,968	435
A <sub>11</sub>	基本設計書	36	6,220	1,138	4,255	576
A <sub>12</sub>	基本設計書	31	3,504	652	2,392	340
A <sub>13</sub>	基本設計書	168	36,884	2,903	28,301	1,638
A <sub>14</sub>	基本設計書	125	12,819	1,458	9,414	905
A <sub>a</sub>	基本設計書	138	24,514	1,210	18,647	754
A <sub>b</sub>	基本設計書	218	33,434	1,639	26,178	845
A <sub>c</sub>	基本設計書	337	59,745	2,481	47,780	1,428
R <sub>1</sub>	要求仕様書	59	12,425	1,267	9,965	867
R <sub>2</sub>	要求仕様書	100	36,419	2,347	26,521	1,074
R <sub>3</sub>	要求仕様書	177	38,390	2,089	29,707	1,267
R <sub>4</sub>	要求仕様書	39	5,315	744	4,130	482
R <sub>5</sub>	要求仕様書	16	2,810	634	2,117	435
R <sub>6</sub>	要求仕様書	23	3,179	786	2,490	562
R <sub>7</sub>	要求仕様書	14	1,766	550	1,317	408
R <sub>8</sub>	要求仕様書	20	3,689	507	3,077	354
R <sub>9</sub>	要求仕様書	114	22,707	2,651	14,747	1,313
R <sub>10</sub>	要求仕様書	46	8,097	1,557	6,231	1,083
R <sub>11</sub>	要求仕様書	16	3,136	856	2,226	564
R <sub>12</sub>	要求仕様書	47	7,782	707	5,737	464
R <sub>13</sub>	要求仕様書	13	2,439	565	1,936	430
R <sub>14</sub>	要求仕様書	18	4,295	1,018	3,324	695
R <sub>a</sub>	要求仕様書	74	13,806	1,520	10,631	1,006
R <sub>b</sub>	要求仕様書	181	38,581	2,000	25,758	1,118
R <sub>c</sub>	要求仕様書	47	12,897	1,484	10,624	987

本設計書群 A から頻出名詞を取り出し、それらの間に共通語彙 F があるかどうかを調べる。同様にして要求仕様書群 R の共通語彙 G を取り出す。また、3.3.2 項で述べる条件を満たす単語を F から取り除いた共通語彙 H があるかどうか確かめる。H があるか調べる理由は、F には設計書固有ではなくソフトウェア開発固有の単語が含まれる可能性があるためである。F、H ともに共通語彙が得られた場合には最低出現割合を設定し、それよりも出現割合の大きい単語を共通語彙とする。閾値を変化させ共通語彙の増減を調べる。

### 3.2.2 基本設計書と要求仕様書における共通語彙の割合

基本設計書のほうが要求仕様書よりも多く F、H を含めば、対になる基本設計書と要求仕様書が与えられたときに、どちらが基本設計書かを区別することができる。基本設計書は要求仕様書を具体化したものであるため、基本設計書の具体化度合いの評価につながることを期待できる。基本

設計書と要求仕様書の組である AR 群において、基本設計書と要求仕様書が共通語彙 F、H をどの程度含むかを確かめ、基本設計書のほうが対になる要求仕様書よりも共通語彙を含む割合が大きいことを確かめる。

### 3.2.3 基本設計書における共通語彙の分布

共通語彙により、基本設計書全体の具体化度合いを評価できるのか、あるいは特定箇所を評価できるのかを明らかにする。基本設計書 A<sub>a</sub>、A<sub>b</sub>、A<sub>c</sub> の章ごとに共通語彙の割合を求め、基本設計書の特定箇所において共通語彙に含まれる単語が偏って存在するか、あるいは、基本設計書全体に共通語彙が一様に存在するかを明らかにする。

## 3.3 分析手順

### 3.3.1 準備

基本設計書、要求仕様書ともに同一の方法で共通語彙を得



る。本項では手順の準備として、共通語彙の求め方を示す。まず、最初に単一のドキュメント  $D$  から語彙を取り出す手順を示す。ドキュメント  $D$  は PDF 形式で記録されている。PDF 閲覧アプリケーションにより、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 に与え、MeCab が名詞と判断した名詞の集合  $N_m(D)$  を得る。 $N_m(D)$  から日本語語彙大系 [16] において定義のある名詞の集合  $N(D) = \{n_1, \dots, n_q\}$  を得る。 $N_m(D)$  には“(”や“[”といった記号も含まれるため、これらを除外するために日本語語彙体系での定義を用いた。日本語語彙体系は 14 万語を収録する辞書である。また、ドキュメント  $D$  における名詞  $n_k$  の出現回数  $C_D(n_k)$  を求め、出現割合  $P_D(n_k)$  を求める。 $P_D(n_k) = C_D(n_k) / \sum C_D(n_i)$  ( $i = 1, \dots, q$ ) である。次に、名詞  $n_k$  の出現割合の閾値  $t_1$  を設定し、 $t_1$  を超えていれば 1 を、それ以外の場合に 0 の値をとる  $L_D(n_k)$  を得る。

$$L_D(n_k) = \begin{cases} 1 & (P_D(n_k) > t_1) \\ 0 & (P_D(n_k) \leq t_1) \end{cases} \quad (1)$$

また、ドキュメント  $D_1, \dots, D_r$  において名詞  $n_j$  の出現割合が閾値を超えるドキュメントの数  $V_{D_1, \dots, D_r}(n_k)$  を次のように定義する。

$$V_{D_1, \dots, D_r}(n_k) = \sum L_{D_i}(n_k) \quad (i = 1, \dots, r) \quad (2)$$

$V_{D_1, \dots, D_r}(n_k)$  の値が大きいほど、名詞  $n_k$  がより多くのドキュメントに共通して出現することを示す。ドキュメント  $D_1, \dots, D_r$  のうち、2 件以上のドキュメントに出現する単語を共通語彙とし、ドキュメントの共通語彙  $F_{t_1}$  を次のように定義する。

$$F_{t_1} = \{n_1, \dots, n_w\} \quad (V_{D_1, \dots, D_r}(n_w) \geq 2) \quad (3)$$

また、名詞  $n_w$  が出現するドキュメントの件数に閾値  $t_2$  を設けた共通語彙  $F_{t_1, t_2}$  を次のように定義する。

$$F_{t_1, t_2} = \{n_1, \dots, n_w\} \quad (V_{D_1, \dots, D_r}(n_w) \geq t_2) \quad (4)$$

### 3.3.2 基本設計書における共通語彙

前項の手順を用いて、基本設計書群  $A$  に含まれる基本設計書の間で共通語彙  $F_{t_1}$  が存在するか調べる。まず、 $t_1 = 0$  として、名詞  $n_k$  が存在するかどうかを確かめる。また、次に述べる方法で定義される共通語彙  $H_0$  が存在するか確かめる。 $G_0$  は  $F_0$  を得た手順を用いて、要求仕様書群  $R = \{R_1, \dots, R_{14}\}$  から得る。 $F_0$  の要素  $n_k$  のうち、 $V_A(n_k) > V_R(n_k)$  となるものを  $H_0$  とする。 $F_0, H_0$  とも単語が存在した場合には、出現割合の閾値  $t_1$  を変化させ共通語彙数の変化と  $n_k$  を調べる。

### 3.3.3 基本設計書と要求仕様書における共通語彙の割合

同一ソフトウェアの基本設計書と要求仕様書の組である  $AR$  群 ( $A_\alpha, R_\alpha$  ( $\alpha = a, b, c$ )) において、共通語彙

$F_{t_1, t_2}, H_{t_1, t_2}$  がどの程度含まれるか確かめる。 $t_1$  の値は前項と同一とする。 $t_2$  の値は、3 から 14 までのすべての値とする。 $A_\alpha$  と  $R_\alpha$  ( $\alpha = a, b, c$ ) が、 $F_{t_1, t_2}, H_{t_1, t_2}$  を含む割合を比較する。具体的には、 $\sum P_{R_\alpha}(n_k) / |F_{t_1, t_2}|$ 、および、 $\sum_{n_k \in F_{t_1, t_2}} P_{A_\alpha}(n_k) / |F_{t_1, t_2}|$  ( $n_k \in F_{t_1, t_2}$ ) を  $\alpha = a, b, c$  それぞれについて求める。 $\sum P_{A_\alpha}(n_k) / |F_{t_1, t_2}|$  と  $\sum P_{R_\alpha}(n_k) / |F_{t_1, t_2}|$  を比較して、 $\sum P_{A_\alpha}(n_k)$  のほうが大きければ、共通語彙によって対になる要求仕様書と基本設計書を区別できる可能性があることが分かる。帰無仮説を“ $\sum P_{A_\alpha}(n_k) / |F_{t_1, t_2}|$  と  $\sum P_{R_\alpha}(n_k) / |F_{t_1, t_2}|$  は等しい”とし、二項検定により棄却できるか確かめる。帰無仮説が棄却できた場合には、 $\sum P_{A_\alpha}(n_k) / |F_{t_1, t_2}|$  と  $\sum P_{R_\alpha}(n_k) / |F_{t_1, t_2}|$  を比較する。また、 $n_k \in H_{t_1, t_2}$  についても同様に調べる。

### 3.3.4 基本設計書における共通語彙の分布

基本設計書  $A_a, A_b, A_c$  を章ごとに分割し、各章において共通語彙がどの程度の割合で含まれるか調べ、特定の章において共通語彙が多いかどうかを明らかにする。前項で得た共通語彙  $F_{t_1, t_2}, H_{t_1, t_2}$  のうち、基本設計書と要求仕様書の出現割合に最も大きな差があったものを用いる。

$A_a$  は住宅履歴情報管理システムの基本設計書で、7 章で構成されている。4 章の業務アプリケーション設計がドキュメント全体のページ数の約 5 割を占めており、システムの機能が詳細に説明されている。 $A_b$  は大規模プラント制御システムにおけるセキュリティ機能共通基盤の基本設計書で、8 章で構成されている。4, 5, 6 章でシステムの詳細を説明し、1, 2, 3 章, 7, 8 章はシステムの目的や開発指針が記述されている。 $A_c$  は図書館の情報システムの基本設計書で、8 章で構成されている。3, 5 章はシステムの機能が詳細に記述されている章であり、4, 6, 7 章はシステムの移行、運用、性能、構成について記述されている。

## 4. 分析結果

### 4.1 基本設計書における共通語彙

$F_0, H_0$  ともに存在した。基本設計書群  $A$  のすべてに出現する単語も存在した。そこで、複数の  $t_1$  を設定し  $F_{t_1}, H_{t_1}$  を求めた。 $t_1$  の値は、最も単語数の小さい  $A_{12}$  のおおよそ 1 語に該当する 0.0005、14 件すべての基本設計書に出現する単語がなくなる 0.0015、その間の値である 0.00010 とした。 $t_1 = 0.0005, 0.0010, 0.0015$  としたときの  $F_{t_1}, H_{t_1}$  の共通語彙の語数 ( $|F_{t_1, t_2}|, |H_{t_1, t_2}|$ ) を図 1, 図 2 に示す。図 1 のとおり、 $V_{A_1, \dots, A_{14}} = 14$  となる名詞が複数存在した。 $V_{A_1, \dots, A_{14}} \geq 10$  となる単語は  $t_1 = 0.0015$  で 29 語、 $t_1 = 0.0005$  で 67 語あった。 $V_{A_1, \dots, A_{14}} \geq 5$  となる単語は  $t_1 = 0.0015$  で 98 語、 $t_1 = 0.0005$  で 248 語あった。図 2 に示すとおり、 $|H_{t_1}|$  は同一の  $t_1, V_{A_1, \dots, A_{14}}$  における  $|F_{t_1}|$  に対して、約半分であった。 $V_{A_1, \dots, A_{14}}$  が 10 以上になると  $|F_{t_1}|$  は 20 語未満となった。 $V_{A_1, \dots, A_{14}}$  が 11 以上になると  $|H_{t_1}|$  は 10 語未満となった。

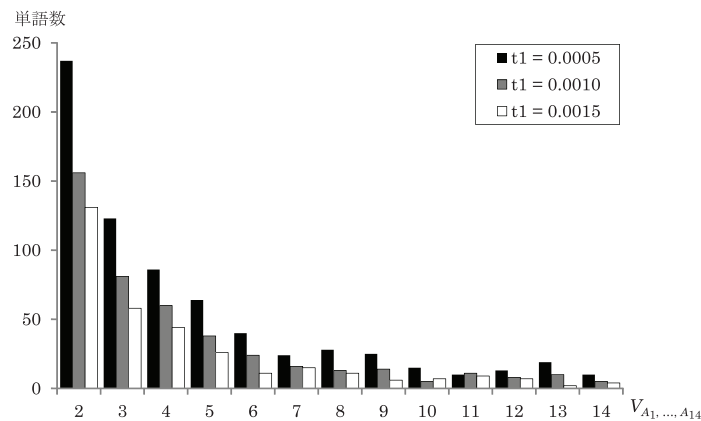


図 1  $F_{t_1}$  の単語数

Fig. 1 Number of words in  $F_{t_1}$ .

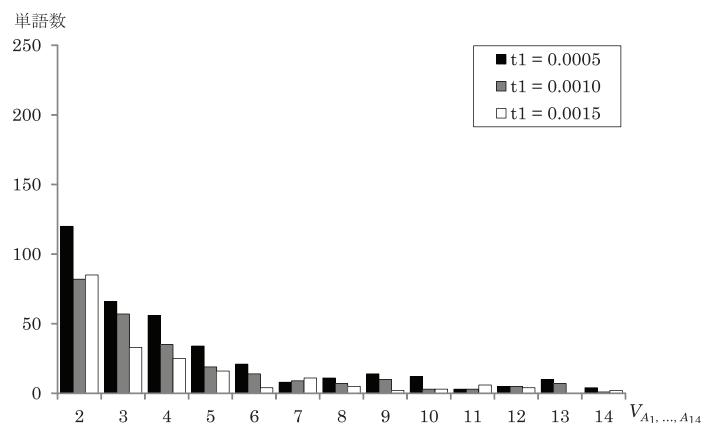


図 2  $H_{t_1}$  の単語数

Fig. 2 Number of words in  $H_{t_1}$ .

表 3, 表 4 に  $F_{t_1}$ ,  $H_{t_1}$  において  $V_{A_1, \dots, A_{14}}$  が大きかった単語上位 20 件を示す. 表 3, 表 4 ともソフトウェアが対象とするドメイン (業務) 固有の単語は含まれておらず, “入力”, “出力”, “表示”, “更新”, “登録”, “検索”, “取得” といった単語が含まれていた.  $F_{t_1}$  に含まれるが  $H_{t_1}$  に含まれていない単語には “システム”, “情報”, “管理” があつた.

#### 4.2 基本設計書と要求仕様書における共通語彙の割合

基本設計書と要求仕様書の組  $AR$  群において  $F_{t_1, t_2}$  に含まれる単語の出現割合を調べた. 結果を表 5 に示す. 表中の “ $|F|$ ” は該当する  $t_1, t_2$  の組合せにおける共通語彙  $F_{t_1, t_2}$  の単語数である. すべての  $t_1, t_2$  の組合せにおいて, 帰無仮説 “ $\sum P_{A_\alpha}(n_k)/|F_{t_1, t_2}|$  と  $\sum P_{R_\alpha}(n_k)/|F_{t_1, t_2}|$  は等しい ( $n_k \in F_{t_1, t_2}$ )” が  $\alpha = a, b, c$  において棄却された ( $p < 0.05$ ).

$AR$  群のいずれのドキュメント,  $t_1$  の 3 つの値すべてにおいても,  $t_2$  の値が大きくなるにつれ, 共通語彙を含む割合が小さくなった. 同様に,  $AR$  群のいずれのドキュメント, 12 個の  $t_2$  の 3 つの値すべてにおいても,  $t_1$  の値が大きくなるにつれ, 共通語彙を含む割合が小さくなった.  $t_2 = 3, \dots, 11$  では, 3 組すべてにおいて  $F_{t_1, t_2}$  の単語の出

現割合は基本設計書のほうが大きかった.

共通語彙  $H_{t_1, t_2}$  ( $t_2 = 3, \dots, 14$ ) を得て,  $AR$  群における出現割合を調べた. 結果を表 6 に示す. 表中の “ $|H|$ ” は該当する  $t_1, t_2$  の組合せにおける共通語彙  $H_{t_1, t_2}$  の単語数である. すべての  $t_1, t_2$  の組合せにおいて, 帰無仮説 “ $\sum P_{A_\alpha}(n_k)/|H_{t_1, t_2}|$  と  $\sum P_{R_\alpha}(n_k)/|H_{t_1, t_2}|$  は等しい ( $n_k \in H_{t_1, t_2}$ )” が  $\alpha = a, b, c$  において棄却された ( $p < 0.05$ ).

共通語彙  $H_{t_1, t_2}$  においても,  $AR$  群のドキュメント,  $t_1$  の 3 つの値すべてにおいて,  $F_{t_1, t_2}$  と同様の傾向を示した.  $t_2 = 3, \dots, 12$  としたとき,  $AR$  群のすべてのドキュメントにおいて, 基本設計書のほうが  $H_{t_1, t_2}$  の出現割合が大きかった.

3 組の基本設計書と要求仕様書の間で共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  を含む割合の差 ( $\sum P_{A_\alpha}(n_k)/|F_{t_1, t_2}| - \sum P_{R_\alpha}(n_k)/|F_{t_1, t_2}|$  ( $n_k \in F_{t_1, t_2}$ )) が最も大きくなった  $t_1, t_2$  の組合せを表 7 に示す. 割合の差が大きくなる  $t_1, t_2$  の値は 3 組の基本設計書と要求仕様書の間で異なっていた.  $A_a$  と  $R_a$  の組,  $A_c$  と  $R_c$  の組では, 同一の  $t_1, t_2$  に対して,  $H_{t_1, t_2}$  のほうが  $F_{t_1, t_2}$  よりも割合の差が大きくなっていった.  $A_b$  と  $R_b$  の組では,  $F_{t_1, t_2}$  のほうが  $H_{t_1, t_2}$  よりも割

表 3  $F_{t_1}$  における  $V_{A_1, \dots, A_{14}}$  の値が上位 20 件となった単語

Table 3 Top 20 words with the largest  $V_{A_1, \dots, A_{14}}$  values in  $F_{t_1}$ .

$t_1 = 0.0005$		$t_1 = 0.0010$		$t_1 = 0.0015$	
単語	V	単語	V	単語	V
入力	14	機能	14	情報	14
機能	14	情報	14	システム	14
システム	14	システム	14	データ	14
基本	14	データ	14	可能	14
場合	14	可能	14	管理	13
データ	14	管理	13	ため	13
対象	14	設計	13	作成	12
可能	14	作成	13	機能	12
結果	14	場合	13	登録	12
情報	14	登録	13	処理	12
登録	13	一覧	13	表示	12
管理	13	変更	13	場合	12
出力	13	対象	13	変更	12
画面	13	結果	13	入力	11
表示	13	ため	13	出力	11
項目	13	項目	12	設計	11
設定	13	出力	12	一覧	11
取得	13	表示	12	対象	11
変更	13	処理	12	基本	11
設計	13	画面	12	設定	11

表 4  $H_{t_1}$  における  $V_{A_1, \dots, A_{14}}$  の値が上位 20 件となった単語

Table 4 Top 20 words with the largest  $V_{A_1, \dots, A_{14}}$  values in  $H_{t_1}$ .

$t_1 = 0.0005$		$t_1 = 0.0010$		$t_1 = 0.0015$	
単語	V	単語	V	単語	V
入力	14	可能	14	データ	14
基本	14	作成	13	可能	14
対象	14	登録	13	作成	12
結果	14	設計	13	処理	12
設計	13	一覧	13	表示	12
表示	13	変更	13	変更	12
画面	13	対象	13	設計	11
出力	13	結果	13	出力	11
一覧	13	処理	12	一覧	11
変更	13	表示	12	名	11
名	13	画面	12	設定	11
設定	13	出力	12	対象	11
使用	13	設定	12	内容	10
取得	13	更新	11	画面	10
更新	12	名	11	結果	10
処理	12	確認	11	検索	9
検索	12	条件	10	確認	9
用	12	用	10	状況	8
詳細	12	詳細	10	追加	8
番号	11	検索	9	削除	8

合の差が大きくなっていった。

AR 群のドキュメントにおける頻出単語を表 8, 表 9, 表 10 に示す。表の 1 行は 1 つの単語に対応する。“回数”列はそのドキュメントにおける出現回数である。“語彙”列

表 5 同一ソフトウェアの基本設計書と要求仕様書における  $F_{t_1, t_2}$  の割合

Table 5 Percentages of words in  $F_{t_1, t_2}$  to all words in architecture documents and the corresponding requirement specification documents.

		$A_a$ と $R_a$		$A_b$ と $R_b$		$A_c$ と $R_c$		
$t_1$	$t_2$	$A_a$	$R_a$	$A_b$	$R_b$	$A_c$	$R_c$	$ F $
0.0005	4	0.62	0.58	0.71	0.58	0.70	0.67	86
	6	0.54	0.49	0.54	0.47	0.58	0.54	40
	8	0.46	0.43	0.42	0.39	0.51	0.47	28
	10	0.38	0.31	0.34	0.25	0.41	0.39	15
	12	0.35	0.28	0.29	0.20	0.34	0.33	13
14	0.13	0.15	0.16	0.10	0.14	0.15	10	
0.0010	4	0.55	0.50	0.59	0.49	0.60	0.57	60
	6	0.46	0.41	0.44	0.36	0.49	0.47	24
	8	0.39	0.32	0.36	0.27	0.42	0.39	13
	10	0.33	0.28	0.28	0.19	0.33	0.32	5
	12	0.26	0.21	0.21	0.14	0.22	0.21	8
14	0.10	0.13	0.14	0.09	0.11	0.12	5	
0.0015	4	0.48	0.41	0.52	0.44	0.54	0.50	44
	6	0.41	0.34	0.36	0.30	0.43	0.41	11
	8	0.36	0.30	0.31	0.21	0.36	0.35	11
	10	0.30	0.24	0.25	0.17	0.30	0.28	7
	12	0.18	0.19	0.17	0.13	0.17	0.17	7
14	0.10	0.13	0.07	0.07	0.08	0.09	4	

表 6 同一ソフトウェアの基本設計書と要求仕様書における  $H_{t_1, t_2}$  の割合

Table 6 Percentages of words in  $H_{t_1, t_2}$  to all words in architecture documents and the corresponding requirement specification documents.

		$A_a$ と $R_a$		$A_b$ と $R_b$		$A_c$ と $R_c$		
$t_1$	$t_2$	$A_a$	$R_a$	$A_b$	$R_b$	$A_c$	$R_c$	$ H $
0.0005	4	0.32	0.18	0.22	0.15	0.32	0.25	56
	6	0.28	0.14	0.14	0.11	0.27	0.20	34
	8	0.23	0.11	0.13	0.10	0.24	0.17	11
	10	0.19	0.06	0.10	0.06	0.17	0.14	12
	12	0.16	0.04	0.07	0.03	0.13	0.11	5
14	0.03	0.01	0.02	0.01	0.02	0.03	4	
0.0010	4	0.33	0.18	0.16	0.14	0.31	0.23	35
	6	0.28	0.14	0.14	0.13	0.26	0.18	14
	8	0.23	0.08	0.12	0.09	0.19	0.13	11
	10	0.18	0.06	0.06	0.03	0.14	0.09	3
	12	0.14	0.04	0.04	0.02	0.10	0.07	5
14	0.00	0.00	0.00	0.00	0.01	0.00	1	
0.0015	4	0.25	0.11	0.18	0.15	0.28	0.21	25
	6	0.20	0.08	0.14	0.09	0.21	0.15	4
	8	0.18	0.06	0.11	0.06	0.16	0.12	5
	10	0.12	0.04	0.08	0.04	0.14	0.08	3
	12	0.04	0.02	0.04	0.03	0.07	0.04	4
14	0.01	0.01	0.03	0.02	0.03	0.02	2	

は含まれていた共通語彙である。共通語彙  $F, H$  のいずれにも含まれていなかった場合，“語彙”列は空欄である。3 組のドキュメントのいずれにおいても，要求仕様書におけ

表 7 基本設計書と要求仕様書の割合の差が最大となった  $t_1$ ,  $t_2$  の組合せ

Table 7 Pairs of  $t_1$  and  $t_2$  with the largest differences between architecture documents and the corresponding requirement specification documents.

組合せ	$F_{t_1, t_2}$			$H_{t_1, t_2}$		
	$t_1$	$t_2$	差	$t_1$	$t_2$	差
$A_a$ と $R_a$	0.0015	7	7.8	0.0010	5	15.3
$A_b$ と $R_b$	0.0005	4	13.7	0.0005	3	7.6
$A_c$ と $R_c$	0.0005	7	4.5	0.0015	5	8.1

表 8  $A_a$  と  $R_a$  における頻出単語

Table 8 Frequent words in  $A_a$  and  $R_a$ .

$A_a$			$R_a$		
単語	回数	語彙	単語	回数	語彙
情報	1,572	$F$	情報	992	$F$
住宅	1,479		住宅	816	
画面	1,142	$F, H$	履歴	340	
登録	497	$F, H$	システム	255	$F$
変更	443	$F, H$	所有	248	
履歴	434		機関	146	
申請	402		管理	141	$F$
アカウント	382		サービス	133	$F$
プレイヤー	330		活用	129	
図	303	$F, H$	蓄積	126	
確認	299	$F, H$	登録	121	$F, H$
所有	273		ため	99	$F$
更新	252	$F, H$	生成	99	
完了	235		項目	95	$F$
書	233	$F$	場合	95	$F$
管理	229		必要	90	$F$
表示	227	$F, H$	的	89	$F$
検索	198	$F, H$	アカウント	86	
アクセス	151	$F, H$	利用	79	$F$
削除	150	$F, H$	書類	75	

る出現回数が上位である単語は、対応する基本設計書でも上位であるものが多かった。

表 8 のとおり、基本設計書  $A_a$  に含まれ要求仕様書  $R_a$  に含まれない単語は 13 種類あった。13 種類のうち、共通語彙  $F$  に含まれる単語が 10 種類、 $H$  に含まれる単語が 9 種類であった。表 9 のとおり、基本設計書  $A_b$  に含まれ要求仕様書  $R_b$  に含まれない単語は、8 種類あった。8 種類のうち、共通語彙  $F$  に含まれる単語が 2 種類、 $H$  に含まれる単語が 1 種類であった。表 10 のとおり、基本設計書  $A_c$  に含まれ要求仕様書  $R_c$  に含まれない単語は、9 種類あった。9 種類のうち、共通語彙  $F$  に含まれる単語が 5 種類、 $H$  に含まれる単語が 4 種類であった。

### 4.3 基本設計書における共通語彙の分布

$A_a$ ,  $A_b$ ,  $A_c$  の各章に対して、表 7 に示した  $t_1$ ,  $t_2$  の値における共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  を含む割合を求めた。結

表 9  $A_b$  と  $R_b$  における頻出単語

Table 9 Frequent words in  $A_b$  and  $R_b$ .

$A_b$			$R_b$		
単語	回数	語彙	単語	回数	語彙
機能	1,650	$F$	情報	771	$F$
情報	1,008	$F$	管理	636	$F$
機器	939		アクセス	617	$F, H$
認証	864		セキュリティ	585	
データ	769	$F$	制御	569	
通信	750		データ	561	$F$
制御	697		利用	444	$F$
フレーム	651		機能	432	$F$
ワーク	645		システム	409	$F$
セキュリティ	551		機器	383	
監視	529		通信	350	
アクセス	509	$F, H$	番号	320	$F, H$
利用	456	$F$	脅威	307	
秘匿	374		ネットワーク	307	$F$
管理	367	$F$	監視	295	
系	346		サービス	272	$F$
メッセージ	312		系	255	
サーバ	261	$F$	許可	254	
専用	259		不正	233	
出力	257	$F, H$	プロ	223	

表 10  $A_c$  と  $R_c$  における頻出単語

Table 10 Frequent words in  $A_c$  and  $R_c$ .

$A_c$			$R_c$		
単語	回数	語彙	単語	回数	語彙
機能	1,551	$F$	システム	463	$F$
システム	1,300	$F$	図書館	311	
資料	1,179		機能	299	$F$
名	1,033	$F, H$	情報	293	$F$
データ	1,020	$F$	資料	245	
情報	902	$F$	利用	241	$F$
指定	809	$F, H$	業務	237	$F$
表示	805	$F, H$	データ	175	$F$
項目	793	$F$	予約	160	
館	767		検索	157	$F, H$
予約	747		登録	139	$F, H$
利用	745	$F$	一覧	107	$F, H$
図書館	657		別	101	$F, H$
業務	634	$F$	サービス	97	$F$
出力	607	$F, H$	書	94	$F$
区分	580		統計	87	
書誌	558		表示	85	$F, H$
検索	546	$F, H$	化	85	$F$
修正	470		処理	79	$F, H$
発注	454		リスト	79	

果を表 11, 表 12, 表 13 に示す。3 件の基本設計書とも章ごとの  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  の出現割合は異なっており分布に偏りがあった。

$A_a$ ,  $A_b$ ,  $A_c$  とも  $F_{t_1, t_2}$  の割合と  $H_{t_1, t_2}$  の割合の差が大



表 11  $A_a$  における共通語彙  $F_{0.0015,7}$ ,  $H_{0.0010,5}$  の割合

Table 11 Common terms in  $F_{0.0015,7}$  and  $H_{0.0010,5}$  in  $A_a$ .

章	章タイトル	$F$ の割合	$H$ の割合
1	システム概要	0.49	0.17
2	データ設計	0.23	0.17
3	業務設計	0.35	0.24
4	業務アプリケーション設計	0.52	0.44
5	システム設計	0.52	0.09
6	システム運用	0.30	0.15
7	画面イメージ	0.32	0.25

表 12  $A_b$  における共通語彙  $F_{0.0005,3}$ ,  $H_{0.0005,3}$  の割合

Table 12 Common terms in  $F_{0.0005,3}$  and  $H_{0.0005,3}$  in  $A_b$ .

章	章タイトル	$F$ の割合	$H$ の割合
1	はじめに	0.70	0.24
2	システムの概要	0.73	0.25
3	設計指針	0.67	0.18
4	通信仕様	0.67	0.29
5	システム構造	0.71	0.14
6	機能仕様	0.74	0.27
7	設計・製造指針	0.61	0.14
8	おわりに	0.66	0.24

表 13  $A_c$  における共通語彙  $F_{0.0005,7}$ ,  $H_{0.0015,5}$  の割合

Table 13 Common terms in  $F_{0.0005,7}$  and  $H_{0.0015,5}$  in  $A_c$ .

章	章タイトル	$F$ の割合	$H$ の割合
1	システム概要	0.53	0.09
2	システム機能設計	0.66	0.08
3	論理データ設計	0.37	0.17
4	移行方式設計	0.55	0.24
5	外部インタフェース設計	0.44	0.20
6	運用設計	0.50	0.16
7	基盤環境設計	0.45	0.10
8	システム構築実施計画	0.49	0.16
	添付資料	0.55	0.26

大きい章と小さい章があった。 $A_a$  では、1, 4, 5 章において  $F_{t_1,t_2}$  の割合が他の章より大きく、0.49, 0.52, 0.52 であったが、 $H_{t_1,t_2}$  の割合は 1 章では 0.17, 4 章では 0.44, 5 章では 0.09 であり、章によって割合の差が異なっていた。1 章はシステムの目的、4 章は機能定義、5 章は機器、ネットワーク構成が記述されていた。 $A_b$  では、どの章も  $F_{t_1,t_2}$  の割合が 0.60 から 0.75 の間の値であったが、 $H_{t_1,t_2}$  の割合は 3, 5, 7 章が他の章と比較して小さかった。3, 7 章は開発プロセス、5 章は機器、ネットワーク構成が記述されていた。 $A_c$  では、1, 2, 4 章と添付資料において  $F_{t_1,t_2}$  の割合が他の章より大きかった。1, 2 章では  $F_{t_1,t_2}$  の割合は 0.53, 0.66 であったが、 $H_{t_1,t_2}$  の割合は 0.09, 0.08 であった。4 章と添付資料では  $F_{t_1,t_2}$  の割合はともに 0.55 であり、 $H_{t_1,t_2}$  の割合は 0.24, 0.26 であった。1 章は対象業務の記述、2 章は添付資料への参照がほとんどであった。4 章は

旧システムからの移行方法と運用方法が記述されていた。添付資料には機能名称と機能説明の一覧が記載され、機能定義の概要が列挙されていた。

## 5. 考察

### 5.1 基本設計書における共通語彙

$F_{t_1}$ ,  $H_{t_1}$  の両方において、 $V_{A_1, \dots, A_{14}}$  が大きい単語が存在した。異なる分野の基本設計書であっても共通して一定割合出現するような単語が存在した。 $V_{A_1, \dots, A_{14}}$  が大きい単語として“入力”, “出力”, “表示”, “取得”, “検索”, “更新”があった。これらの単語を含む機能の入出力や振舞いを定義する文が増え、具体化されていた。これらの単語の増加により、基本設計書が対になる要求仕様書と比較して具体化されている可能性があると考えられる。また、“基本”や“設計”といった基本設計書に固有と考えられる単語や“可能”, “対象”といった単語も  $V_{A_1, \dots, A_{14}}$  が大きい単語に含まれていた。これらの単語が具体化度合いとは直接的な関係がないかどうかを明らかにし、必要ない場合には、共通語彙から取り除くための方法を検討することは今後の課題である。

### 5.2 基本設計書と要求仕様書における共通語彙の割合

$|F_{t_1,t_2}|$ ,  $|H_{t_1,t_2}|$  が 10 よりも大きくなるような  $t_1$ ,  $t_2$  の組合せにおいて、要求仕様書よりも基本設計書のほうが  $F_{t_1,t_2}$ ,  $H_{t_1,t_2}$  を含む割合が大きかった。共通語彙を適切に選び、要求仕様書が共通語彙を含む割合を求め、基本設計書が共通語彙を含む割合と比較することにより、基本設計書の具体化度合いを評価できる可能性がある。

3 組のドキュメントの出現回数上位 20 位の単語を調べたところ、要求仕様書で頻出する単語の多くは基本設計書でも頻出していた。同一の単語であっても基本設計書のほうが出現回数が大きかった。要求仕様書では頻出していなかった単語が基本設計書において頻出しているものもあり、“画面”, “出力”といった機能定義のために使われている単語であった。要求仕様書では頻出していないが、基本設計書において頻出する単語に着目することで、基本設計書の具体化度合いを評価できる可能性がある。

一方で  $F_{t_1,t_2}$  における  $A_a$  と  $R_c$  や  $H_{t_1,t_2}$  における  $A_b$  と  $R_c$  のように異なるソフトウェアの基本設計書と要求仕様書の間では、基本設計書よりも要求仕様書のほうが  $F_{t_1,t_2}$ ,  $H_{t_1,t_2}$  を含む割合が大きいものもあった。対象ソフトウェアや要求仕様書の内容によらず、共通語彙を含む割合と閾値だけで、基本設計書の具体化度合いを評価するためにはさらなる調査と検討が必要である。

### 5.3 基本設計書における共通語彙の分布

3 件の基本設計書において、共通語彙  $F_{t_1,t_2}$  と  $H_{t_1,t_2}$  が

含まれる割合が章によって異なることが分かった。機器、ネットワーク構成を記述した章、開発プロセスや開発方針を記述した章で共通語彙  $H_{t_1, t_2}$  が含まれる割合が小さくなっていた。共通語彙  $F_{t_1, t_2}$  はそれらの章でも割合が大幅に小さくなることはなかった。機能定義や機能概要が記述された章では、 $F_{t_1, t_2}$ 、 $H_{t_1, t_2}$  とも割合が大きくなり、そうでない章では、 $F_{t_1, t_2}$  と比較して  $H_{t_1, t_2}$  の割合が小さくなった。

共通語彙を用いた具体化度合いの評価は機能定義の章を中心として実施することでより大きな効果が得られると考えられる。評価を実施する際にはあらかじめ機器、ネットワーク構成の章や設計指針のような開発プロセスに関する章を除外することで、より適切な評価ができる可能性がある。

また、 $F_{t_1, t_2}$  と  $H_{t_1, t_2}$  の割合の差 ( $\sum P_A(n_k)/|F_{t_1, t_2}|(n_k \in F_{t_1, t_2}) - \sum P_A(n_k)/|H_{t_1, t_2}|(n_k \in H_{t_1, t_2})$ ) を章の間で比較することにより、機能定義や機能概要の章の記述の具体化度合いを評価できる可能性がある。

$A_c$  の旧バージョンのシステムからの移行に関する章は  $F_{t_1, t_2}$ 、 $H_{t_1, t_2}$  とも大きな値となった。この章では機能定義や機能概要は記述されていなかったが、移行前の機能に関する記述があった。こうした旧バージョンの機能に関する記述と要求仕様に沿って記述された機能定義や機能概要とを区別するための方法を考案することは今後の課題である。

#### 5.4 ドキュメントの自動評価にむけて

設計書の共通語彙  $F_{t_1, t_2}$ 、 $H_{t_1, t_2}$  の抽出や基本設計書、要求仕様書における出現頻度の計数は目視を要さず、計算機支援により自動化できる。共通語彙だけで設計書を評価するには解決しなければならない課題がいくつかあるが、これまで目視でしか実施できなかった設計書全体を網羅的に評価できることには大きなメリットがあると考えられる。その1つは低コストで評価できる点である。ドキュメントレビューをはじめとする目視評価はコストが大きいため、ドキュメント作成後に1回だけ実施することが多かった。そのため、ドキュメント作成者が問題に気づくのはドキュメント作成後のことが多かった。しかし、計算機支援により低コストで評価できれば、作成途中に複数回評価することができる。早期に問題に気づき、是正できる可能性が高まる。ドキュメントの作成途中にドキュメントのごく一部を抜き取って目視評価するアジャイルインスペクション [5] においても、抜き取りの対象にならなかった箇所は作成後にしか目視評価が実施されず、同様の問題があった。

作成途中の設計書の評価は、セルフチェックや進捗確認ミーティングで活用できる。事前に共通語彙を過去の設計書と要求仕様書から得ておき、たとえば次のようにセルフチェックを実施する。作成が完了した要求仕様書をリポジトリに登録し、共通語彙が含まれている割合を計測し、その値を基準値とする。設計書の作成がはじまったら、設計書の追記、変更のたびにリポジトリに登録する。リポジト

リに登録するたびに設計書に含まれる共通語彙の割合を章ごとに計測し、要求仕様書から得た基準値とともにリポジトリの登録者（設計書作成担当者）に提示する。設計書作成担当者は未記入の項目、ページ数といった情報とともに提示された割合を見て設計書の作成が順調かどうか検討する。作成が完了したと思っていた章において、明らかに共通語彙を含む割合が小さい場合には、その章を目視で確認するきっかけにできる。進捗確認では、文献 [12] のように共通語彙を含む割合を進捗確認項目に含め、問題がないかどうかを関係者で検討することができる。

#### 5.5 妥当性

同一ソフトウェアの基本設計書と要求仕様書の両方が公開されているものがほとんどみつからず、共通語彙の割合の比較は3組のドキュメントのみを対象とした。より多数の基本設計書と要求仕様書の組を対象とし、本論文の結果と同様の傾向があるかどうかを確かめることは今後の課題の1つである。また、公共性の高いシステムであることや公開を前提に作成されたドキュメントであると考えられるため、3組のドキュメントは要求仕様書から基本設計書に順調に詳細化されたことを前提として評価を実施した。公開を前提としていないドキュメントをはじめとして他のドキュメントにおいて共通語彙を含む割合の差がどのように変化するかを調べることも今後の課題の1つである。

本論文では評価対象をドキュメントのタイトルに付けられた“基本設計書”としたが、これらが必ずしも合意された基本設計書の定義に沿った内容というわけではない。ドキュメントによって記述内容のばらつきがあったり、記述の範囲が異なったりしている。しかし、AR群の基本設計書の章別の評価で示したように、機能定義や機能概要が記述された章では  $F_{t_1, t_2}$ 、 $H_{t_1, t_2}$  とも割合が高くなり、開発方針、ネットワーク、機器構成が記述された章においては  $F_{t_1, t_2}$  の割合が大きく、 $H_{t_1, t_2}$  の割合は小さくなっていたことが分かった。こうした記述内容によって対象とする章を選択することにより、より適切な評価につながると考えられる。

分析では、本研究で目指すドキュメントの自動評価を想定したため、PDF からテキスト形式で取り出せる単語を対象とした。そのため、本文と表のテキストは対象であるが、図に含まれるテキストは対象になっていない。図に含まれているテキストを評価対象とした場合に評価結果にどのような影響があるかを調べることは今後の課題である。なお、今回評価の対象としたドキュメントでは、図に含まれるテキストは分量が少なく、図示された情報は本文で何らかの形で言及があるため、評価結果に大きな影響は与えないと判断している。

本論文の分析で対象としたドキュメントは特定のドメインに限定したものではないため、分析で得た共通語彙  $F$ 、 $H$  にはソフトウェアが対象とするドメイン固有の単語は

含まれていなかった。しかし、ドメイン固有の単語の増減により設計書の具体化度合いを評価できる可能性がある。類似のドメインの設計書を多数準備して共通語彙を集めれば、 $F$ ,  $H$ に含まれるような単語に加えてドメイン固有の単語も共通語彙として獲得できることが期待される。本論文ではインターネットで公開されているドキュメントを対象としたため、類似のドメインのドキュメントが多くても3件しかなく分析できていないが、ドメイン固有の単語による設計書の評価は重要な今後の課題の1つである。

## 6. おわりに

ソフトウェアドキュメントの評価コストを下げることを目的とし、目視を要さない方法によるソフトウェアドキュメントの具体化度合いの評価にむけた分析を実施した。対象としたソフトウェアドキュメントは、公共団体等がインターネットで公開している基本設計書と要求仕様書である。検索エンジンを用いて基本設計書のみがみつかった  $A$  群 (14 件)、要求仕様書のみがみつかった  $R$  群 (14 件)、基本設計書と要求仕様書の両方がみつかった  $AR$  群 (3 組) を分析に用いた。 $AR$  群のドキュメントは  $A$  群、 $R$  群には含まれていない。

まず、基本設計書群  $A$  における共通語彙  $F$  が存在することを確かめた。また、 $F$  から要求仕様書群  $R$  で頻出する単語を除いた共通語彙  $H$  が存在することも確かめた。ここで、共通語彙は個々の基本設計書に割合  $t_1$  以上含まれる単語のうち、 $t_2$  件以上の基本設計書に共通して出現するものとした。 $t_1 = 0.0005, 0.0010, 0.0015$ ,  $t_2 = 3, \dots, 14$  において、共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  の変化を調べたところ、 $t_1 = 0.0015$ ,  $t_2 = 14$  においても  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  がともに存在することが分かった。

次に、 $AR$  群を対象に共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  がどの程度含まれているか割合を調べた。割合は、共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  に含まれる単語の出現回数をドキュメントに出現する単語の出現回数で割ったものである。3 組のドキュメントごとに  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  が含まれる割合は異なったが、共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  の単語数が 10 より大きくなる  $t_1$ ,  $t_2$  において、基本設計書に含まれる  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  の割合 (出現頻度) が要求仕様書に含まれる  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  の割合よりも大きかった。

最後に、 $AR$  群の基本設計書の章ごとに共通語彙  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  が含まれる割合を調べた。章ごとに  $F_{t_1, t_2}$ ,  $H_{t_1, t_2}$  が含まれる割合は異なったが、機能定義や機能概要に関する章では  $F_{t_1, t_2}$  の割合と  $H_{t_1, t_2}$  の割合の差が小さくなった。また、機器やネットワーク構成、開発プロセスに関する記述等、機能に関する記述が少ない章において  $H_{t_1, t_2}$  の値が小さくなった。基本設計書の機能定義や機能概要に関する章を対象として、章ごとに  $F_{t_1, t_2}$  や  $H_{t_1, t_2}$  の割合を要求仕様書と比較することにより、基本設計書の具体化度合いを

評価できる可能性が示された。

本論文で実施した分析をもとに目視評価を要さないソフトウェア設計書の評価手法を考案することは今後の課題である。計算機支援によりソフトウェア設計書の評価を自動化できれば、作成途中で複数回評価を実施することができ、作成途中の評価が可能になれば、問題の早期発見につながる場合があり、手戻り削減やリスク低減といった大きな効果が期待される。

## 参考文献

- [1] Antoniol, G., Ayari, K., Penta, M.D., Khomh, F. and Gueheneuc, Y.: Is it a Bug or an Enhancement?, A text-based Approach to Classify Change Requests, *Proc. 2008 Conference of the Center for Advanced Studies on Collaborative Research: Meeting of Minds*, pp.23:1–23:15 (2008).
- [2] Chantree, F., Nuseibeh, B., Roeck, A. and Willis, A.: Identifying Nocuous Ambiguities in Natural Language Requirements, *Proc. International Conference on Requirements Engineering*, pp.59–68 (2006).
- [3] Cleland-Huang, J., Dumitru, H., Duan, C. and Castro-Herrera, C.: Automated Support for Managing Feature Requests in Open Forums, *Comm. ACM*, Vol.52, No.10, pp.68–74 (2009).
- [4] Fagan, M.E.: Design and Code Inspections to Reduce Errors in Program Development, *IBM Systems Journal*, Vol.15, No.3, pp.182–211 (1976).
- [5] Gilb, T.: Agile Specification Quality Control: Shifting Emphasis from Cleanup to Sampling Defects, *Proc. INCOSE Conference*, Vol.18, No.1, pp.1165–1175 (2005).
- [6] Gu, H., Zhao, L. and Shu, C.: Analysis of Duplicate Issue Reports for Issue Tracking System, *International Conference on Data Mining and Intelligent Information Technology Applications*, pp.86–91 (2011).
- [7] Hayes, J.H., Dekhtyar, A. and Sundaram, S.: Advancing Candidate Link Generation for Requirements Tracing: The Study of Methods, *IEEE Trans. Softw. Eng.*, Vol.32, No.1, pp.4–19 (2006).
- [8] Heck, P. and Zaidman, A.: 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).
- [9] Lami, G.: QuARS, A Tool for Analyzing Requirement, Carnegie Mellon University Technical Report, CMU/SEI-2005-TR-014 (2005).
- [10] Lehner, F.: Quality Control in Software Documentation Based on Measurement of Text Comprehension and Text Comprehensibility, *Journal of Information Processing and Management*, Vol.29, No.5, pp.551–568 (1993).
- [11] Moreno, L., Bandara, W., Haiduc, S. and Marcus, A.: On the Relationship between the Vocabulary of Bug Reports and Source Code, *Proc. International Conference on Software Maintenance*, pp.452–455 (2013).
- [12] 松村知子, 森崎修司, 勝又敏次, 玉田春昭, 吉田則裕, 楠本真二, 松本健一: 問題の早期発見・改善を支援するインプロセスプロジェクト管理手法の実プロジェクトへの適用, *電子情報通信学会論文誌 D*, Vol.J92-D, No.11, pp.1974–1986 (2009).
- [13] Runeson, P., Alexandersson, M. and Nyholm, O.: Detection of Duplicate Defect Reports Using Natural Language Processing, *Proc. International Conference on*



*Software Engineering*, pp.499–510 (2007).

[14] Shull, F., Rus, I. and Basili, V.: How Perspective-based Reading Can Improve Requirements Inspections, *IEEE Computer*, Vol.33, No.7, pp.73–79 (2000).

[15] Sun, C., Lo, D., Wang, X., Jiang, J. and Khoo, S.: A Discriminative Model Approach for Accurate Duplicate Bug Report Retrieval, *Proc. International Conference on Software Engineering*, pp.45–54 (2010).

[16] 日本語語彙大系 CD-ROM 版, 岩波書店 (1999).

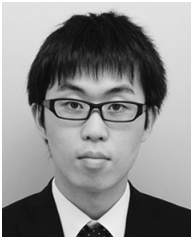
付 録

A.1 分析に用いたドキュメントの URL

表 A.1 分析に用いたドキュメントの ID と URL  
Table A.1 IDs and URL of the target documents.

ID	URL
A <sub>1</sub>	<a href="http://ecom-plat.jp/fbox.php?eid=16496">http://ecom-plat.jp/fbox.php?eid=16496</a>
A <sub>2</sub>	<a href="http://www.city.yokohama.lg.jp/shobo/koukai/ippan-nyuusatsu-pdf/shoubou.kihonnsekkeisyotou.pdf">http://www.city.yokohama.lg.jp/shobo/koukai/ippan-nyuusatsu-pdf/shoubou.kihonnsekkeisyotou.pdf</a>
A <sub>3</sub>	<a href="http://www.mhlw.go.jp/sinsei/chotatu/chotatu/shiyousho-an/dl/090327-1d.pdf">http://www.mhlw.go.jp/sinsei/chotatu/chotatu/shiyousho-an/dl/090327-1d.pdf</a>
A <sub>4</sub>	<a href="http://www.pref.shimane.lg.jp/johoseisaku/elg/krs/krs4/index.data/sekkei.pdf">http://www.pref.shimane.lg.jp/johoseisaku/elg/krs/krs4/index.data/sekkei.pdf</a>
A <sub>5</sub>	<a href="http://www.ya-chos.gr.jp/php/upf/kj1000192tmp1.pdf">http://www.ya-chos.gr.jp/php/upf/kj1000192tmp1.pdf</a>
A <sub>6</sub>	<a href="http://www.pref.yamanashi.jp/toshokan/documents/system-kihonsekkei/documents/system-kihonsekkeisyo_2.pdf">http://www.pref.yamanashi.jp/toshokan/documents/system-kihonsekkei/documents/system-kihonsekkeisyo_2.pdf</a>
A <sub>7</sub>	<a href="http://www.jica.go.jp/chotatsu/buppin/ku57pq0000127z3w-att/temp3of412087.pdf">http://www.jica.go.jp/chotatsu/buppin/ku57pq0000127z3w-att/temp3of412087.pdf</a>
A <sub>8</sub>	<a href="http://www.dis.h.u-tokyo.ac.jp/byomei/dispack/latest/doc/disPACK-spec-031002.pdf">http://www.dis.h.u-tokyo.ac.jp/byomei/dispack/latest/doc/disPACK-spec-031002.pdf</a>
A <sub>9</sub>	<a href="https://www.pref.nagano.lg.jp/gijukan/kensei/nyusatsu/kokyokoji/bukyoku/documents/101014sankousyo.pdf">https://www.pref.nagano.lg.jp/gijukan/kensei/nyusatsu/kokyokoji/bukyoku/documents/101014sankousyo.pdf</a>
A <sub>10</sub>	<a href="http://mojikiban.ipa.go.jp/wp-content/uploads/2013/06/pdf/Hitachi_Spec.pdf">http://mojikiban.ipa.go.jp/wp-content/uploads/2013/06/pdf/Hitachi_Spec.pdf</a>
A <sub>11</sub>	<a href="http://www.nedo.go.jp/content/100462261.pdf">http://www.nedo.go.jp/content/100462261.pdf</a>
A <sub>12</sub>	<a href="http://forge.at.nii.ac.jp/attachments/13/NC2_Repository%E5%9F%BA%E6%9C%AC%E8%A8%AD%E8%A8%88%E6%9B%B8.pdf">http://forge.at.nii.ac.jp/attachments/13/NC2_Repository%E5%9F%BA%E6%9C%AC%E8%A8%AD%E8%A8%88%E6%9B%B8.pdf</a>
A <sub>13</sub>	<a href="http://www.city.kumamoto.jp/hpKiji/pub/detail.aspx?c_id=5&amp;id=1891&amp;class_set_id=3&amp;class_id=537">http://www.city.kumamoto.jp/hpKiji/pub/detail.aspx?c_id=5&amp;id=1891&amp;class_set_id=3&amp;class_id=537</a>
A <sub>14</sub>	<a href="http://ftp.orca.med.or.jp/pub/data/receipt/tec/">http://ftp.orca.med.or.jp/pub/data/receipt/tec/</a>
A <sub>a</sub>	<a href="http://www.cbl.or.jp/slc/pdf/rireki02.pdf">http://www.cbl.or.jp/slc/pdf/rireki02.pdf</a>
A <sub>b</sub>	<a href="https://www.ipa.go.jp/security/fy12/contents/crack/sekitoku/gaibusekeisyo.pdf">https://www.ipa.go.jp/security/fy12/contents/crack/sekitoku/gaibusekeisyo.pdf</a>
A <sub>c</sub>	<a href="http://www.city.yokohama.lg.jp/kyoiku/library/nyuusatsu/05.pdf">http://www.city.yokohama.lg.jp/kyoiku/library/nyuusatsu/05.pdf</a>
R <sub>1</sub>	<a href="http://www.mlit.go.jp/jidosha/topics/oss/yoken_teigi.pdf">http://www.mlit.go.jp/jidosha/topics/oss/yoken_teigi.pdf</a>
R <sub>2</sub>	<a href="http://www.jica.go.jp/announce/public/pdf/110809_opi_sum11.pdf">http://www.jica.go.jp/announce/public/pdf/110809_opi_sum11.pdf</a>
R <sub>3</sub>	<a href="http://www.meti.go.jp/information_2/downloadfiles/kokuji201202170202.pdf">http://www.meti.go.jp/information_2/downloadfiles/kokuji201202170202.pdf</a>
R <sub>4</sub>	<a href="http://www.city.kawaguchi.lg.jp/kbn/Files/1/04150034/attach/kihonteigi.pdf">http://www.city.kawaguchi.lg.jp/kbn/Files/1/04150034/attach/kihonteigi.pdf</a>
R <sub>5</sub>	<a href="http://www.city.ichikawa.chiba.jp/net/legacy/kiban/koukokubun/teigi_02.pdf">http://www.city.ichikawa.chiba.jp/net/legacy/kiban/koukokubun/teigi_02.pdf</a>
R <sub>6</sub>	<a href="http://www.pref.okinawa.lg.jp/site/norin/shinrin/hozen/koubo_hoanrin_system/documents/teigisyo_siyousyo.pdf">http://www.pref.okinawa.lg.jp/site/norin/shinrin/hozen/koubo_hoanrin_system/documents/teigisyo_siyousyo.pdf</a>
R <sub>7</sub>	<a href="http://www.city.nagano.nagano.jp/uploaded/life/90421_89748_misc.pdf">http://www.city.nagano.nagano.jp/uploaded/life/90421_89748_misc.pdf</a>
R <sub>8</sub>	<a href="http://www.nss-med.co.jp/PDF/project1/3define.pdf">http://www.nss-med.co.jp/PDF/project1/3define.pdf</a>
R <sub>9</sub>	<a href="https://www.tulips.tsukuba.ac.jp/portal/spec/spec2010.pdf">https://www.tulips.tsukuba.ac.jp/portal/spec/spec2010.pdf</a>
R <sub>10</sub>	<a href="http://www.aoi-mori.net/tsugaru/nintei/pdf/2011_04/siyousyo.pdf">http://www.aoi-mori.net/tsugaru/nintei/pdf/2011_04/siyousyo.pdf</a>
R <sub>11</sub>	<a href="http://www.chubu.saga.saga.jp/new2006/houkatusisutemu/%E4%BD%90%E8%B3%80%E4%B8%AD%E9%83%A8%E5%BA%83%E5%9F%9F%E9%80%A3%E5%90%88%E5%9C%B0%E5%9F%9F%E5%8C%85%E6%8B%AC%E6%94%AF%E6%8F%B4%E3%82%BB%E3%83%B3%E3%82%BF%E3%83%BC%E3%82%B7%E3%82%B9%E3%83%86%E3%83%A0%E8%A6%81%E4%BB%B6%E5%AE%9A%E7%BE%A9%E6%9B%B8.pdf">http://www.chubu.saga.saga.jp/new2006/houkatusisutemu/%E4%BD%90%E8%B3%80%E4%B8%AD%E9%83%A8%E5%BA%83%E5%9F%9F%E9%80%A3%E5%90%88%E5%9C%B0%E5%9F%9F%E5%8C%85%E6%8B%AC%E6%94%AF%E6%8F%B4%E3%82%BB%E3%83%B3%E3%82%BF%E3%83%BC%E3%82%B7%E3%82%B9%E3%83%86%E3%83%A0%E8%A6%81%E4%BB%B6%E5%AE%9A%E7%BE%A9%E6%9B%B8.pdf</a>
R <sub>12</sub>	<a href="http://www.npa.go.jp/koutsuu/kisei/DSSS/sys_teigisyo.pdf">http://www.npa.go.jp/koutsuu/kisei/DSSS/sys_teigisyo.pdf</a>
R <sub>13</sub>	<a href="http://www.pref.nagano.lg.jp/gijukan/kensei/nyusatsu/kokyokoji/bukyoku/documents/100830youkyusiyosyo.pdf">http://www.pref.nagano.lg.jp/gijukan/kensei/nyusatsu/kokyokoji/bukyoku/documents/100830youkyusiyosyo.pdf</a>
R <sub>14</sub>	<a href="http://www.fukushima-med-jrc.jp/hospital/pdf/20131216_kihon_shiyou.pdf">http://www.fukushima-med-jrc.jp/hospital/pdf/20131216_kihon_shiyou.pdf</a>
R <sub>a</sub>	<a href="http://www.cbl.or.jp/slc/pdf/rireki01.pdf">http://www.cbl.or.jp/slc/pdf/rireki01.pdf</a>
R <sub>b</sub>	<a href="http://www.ipa.go.jp/files/000013611.pdf">http://www.ipa.go.jp/files/000013611.pdf</a>
R <sub>c</sub>	<a href="http://www.city.yokohama.lg.jp/kyoiku/library/nyuusatsu/05.pdf">http://www.city.yokohama.lg.jp/kyoiku/library/nyuusatsu/05.pdf</a>





奈加 大樹

1991年生。2014年名古屋大学工学部電気電子情報工学科卒業。2014年同大学大学院情報科学研究科情報システム学専攻入学。



森崎 修司 (正会員)

2001年奈良先端科学技術大学院大学情報科学研究科博士課程修了。同年(株)インターネットイニシアティブ入社。2007年奈良先端科学技術大学院大学情報科学研究科助教。2013年4月静岡大学大学院情報科学研究科准教授。2013年10月名古屋大学大学院情報科学研究科准教授。ソフトウェアレビュー、エンピリカルソフトウェア工学の研究に従事。博士(工学)。IEEE, 電子情報通信学会, プロジェクトマネジメント学会各会員。



渥美 紀寿 (正会員)

2007年名古屋大学大学院工学研究科情報工学専攻博士課程修了。2005年南山大学数理情報学部情報通信学科講師。2010年名古屋大学大学院情報科学研究科附属組込みシステム研究センター研究員。2011年同センター特任助教。2012年より同大学情報連携統括本部情報戦略室特任助教。現在に至る。博士(工学)。プログラム解析, ソフトウェア保守, ソフトウェア再利用等の研究に従事。日本ソフトウェア科学会, 電子情報通信学会, IEEE, ACM各会員。



山本 修一郎 (正会員)

1977年名古屋工業大学情報工学科卒業。1979年名古屋大学大学院工学研究科情報工学専攻修了。同年日本電信電話公社入社。2002年(株)NTTデータ技術開発本部副本部長。2007年同社初代フェロー, システム科学研究所所長。2009年東京工業大学統合研究院医療情報プロジェクト特任教授。同年名古屋大学情報連携統括本部情報戦略室教授。ソフトウェア工学, 要求工学, ICカードプラットフォーム, ユビキタスコンピューティング, 知識創造デザインの研究に従事。情報処理学会業績賞, 電子情報通信学会業績賞, 通信協会前島賞等受賞。博士(工学)。著書に、『要求定義・要求仕様書の作り方』(ソフト・リサーチ・センター, 2006), 『ゴール指向によるシステム要求管理』(ソフト・リサーチ・センター, 2007), 『すりあわせの技術』(ダイヤモンド社, 2009), 『CMCで変わる組織コミュニケーション』(NTT出版, 2010)等。2000~2002年電子情報通信学会知能ソフトウェア工学研究会研究専門委員長, 2007年人工知能学会知識流通ネットワーク研究会主査。電子情報通信学会, 日本ソフトウェア学会, 人工知能学会, 日本情報経営学会, プロジェクトマネジメント学会, ACM, IEEE各会員。