

出現単語の抽象度を用いたソフトウェアドキュメント評価の計算機支援にむけた分析

森崎 修司[†] 遠藤 充^{††} 杉山 岳弘^{††}

An Empirical Analysis toward Automated Software Document Evaluation by Abstraction Level of Included Words

Shuji MORISAKI[†], Mitsuru ENDO^{††}, and Takahiro SUGIYAMA^{††}

あらまし ソフトウェアドキュメントの具体化度合いをドキュメント全体にわたって網羅的に評価することを目指し、国語辞書で定義された抽象度を用いてドキュメントに出現する単語の抽象度から具体化度合いを評価できるか分析した。対象はインターネットで公開されている要求定義書、基本設計書、詳細設計書である。名詞に12段階の抽象度が定義されている日本語語彙大系を辞書として用いた。ドキュメント全体を網羅的に評価できるか調べるため、14ドキュメントを対象として辞書に抽象度が定義されている単語の割合とそれらの単語がドキュメントの特定箇所に偏って存在していないかを調べた。各ドキュメントにおいて出現する単語の30~50%程度が辞書で抽象度が定義されている単語であり、特定箇所に偏って存在していなかった。また、具体化することにより単語の抽象度の分布に変化があるかを調べるため、同一ソフトウェアの要求定義書、設計書の2組の間で抽象度ごとに単語の出現割合を比較したところ、要求定義書ではより抽象的な単語の割合が、設計書ではより具体的な単語の割合が大きかった。ドキュメントの具体化度合いを計算機支援により網羅的に評価できる可能性が示された。

キーワード ドキュメント定量化, 抽象度, 辞書

1. ま え が き

ドキュメントによる具体化を中心とするソフトウェア開発では、要求定義書、基本設計書、詳細設計書といったドキュメントを作成しながら、利用者の要求を段階的に具体化した上で、コーディング、テストを実施する。コーディングとテストにおいては実際にプログラムを動作させながらプログラムが期待どおりに作成されているかどうかを判断することができる。しかし、自然言語で記述されたドキュメントは動作させることができないので、内容を目視して評価することが一般的である [1], [2]。

目視によるドキュメントの評価は強力であるもののコストが大きいという問題があり、コストを小さくす

るためにドキュメントの一部を選択して目視評価する手法が提案されている [3]~[5]。Thelinらは全体の20%のページをサンプリングして、問題のありそうな箇所を選んだ上で、その箇所に限定して重点的な欠陥検出をする手法を提案している [5]。こうした手法では、目視評価のコストは小さくできるが、ドキュメント全体に対する網羅的な評価ができなくなる。

ドキュメントの一部を取り出すのではなく、ドキュメント全体にわたって偏りのない評価を小さいコストで実施するために、計算機支援によりドキュメント評価手法が提案されている [6], [7]。これらの研究は大きく二つのアプローチに大別できる。一つは、ドキュメント作成時に計算機が解釈しやすいルールを設け、作成者がそれに従うことにより、評価をしやすくするものである。UML等の記述形式をはじめとして、図表や体裁のチェック [8]がある。これらの研究はドキュメント全体の評価を低コストで実施できるというメリットがある一方で、作成者に記法やタグ付けなどの制約を強いるというデメリットがある。

[†] 名古屋大学大学院情報科学研究科, 名古屋市
Graduate School of Information Science, Nagoya University,
Nagoya-shi, 464-8601 Japan

^{††} 静岡大学大学院情報学研究科, 浜松市
Graduate School of Informatics, Shizuoka University,
Hamamatsu-shi, 432-8011 Japan

もう一つのアプローチは、ドキュメント作成には特に制約を設けず、事前に設定したルールを逸脱する部分を問題点の候補として作成者に提示するものである [6], [9]. Lehner はドキュメントの読みやすさを計測するため、一文当りの単語数といった事前のルール定義が必要ない評価方法によりドキュメントを評価した結果を報告している [6]. これらの研究は、ドキュメントが何らかの記法に沿っていなかったりタグ付けされたりしていなくても評価が可能であるため、適用範囲が広くドキュメントの作成コストを増やさないとはいえる。

本研究で目指す評価手法は、Lehner の手法と似ているがドキュメントの具体化度合いを評価する点で異なる。本研究では、ソフトウェアの要求は曖昧な状態から設計書の記述を経て、プログラミング可能な具体的な内容に変換されることを前提とし、その過程を記述したドキュメントに含まれる単語も具体的にないと考える。単語が具体的かどうかは国語辞書を用いて計算機支援により評価する。そのため、具体的な問題箇所を特定したり、評価結果のみで欠陥や問題を修正したりすることは想定していない。事前のルール定義といった準備コストを小さくし、進捗確認時の簡易評価やモニタリング、セルフチェック、検収時の一次検査といった目視による評価と組合せることを前提としている。

しかし、国語辞書で抽象度が定義された単語がドキュメントの特定箇所に偏っていないことやドキュメントの具体化度合いが単語の抽象度に表われるかどうかは明らかではない。ここで、ドキュメントの特定箇所に偏った評価とは、文献 [5] のようにドキュメントの一部を抜き取り、その箇所に限定した評価をすることである。本論文では、特定箇所に限定しないドキュメント評価を網羅的なドキュメント評価とし、作成時に記述方法等の制約を課すことなく、ドキュメントの具体化度合いを評価できるか実証的に分析する。具体的には、インターネットで公開されている 14 件の要求定義、基本設計、詳細設計書を対象とし、ドキュメントに含まれる単語のうち、国語辞書で抽象度が定義されている単語の割合、抽象度が定義されている単語がドキュメントの特定箇所に偏っていないかを調べる。また、同一ソフトウェアの前工程のドキュメントと後工程のドキュメントの両方が公開されている 2 組のドキュメントを対象として、前工程のドキュメントに含まれる単語よりも後工程のドキュメントに含まれる単語

の方が具体的か調べる。

2. 関連研究

ドキュメント評価の自動化は様々に研究されている。QuARS [7] 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 の評価指標として 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 を定義し、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。QuARS の評価には Flesch reading ease index [10] といった既存の読みやすさの指標も含まれている。これらの指標には、一文当りの単語数のような事前にルール定義を要しないものもあり、本論文の分析と共通部分が多い。しかしながら、QuARS では個々の問題を検出することを主眼としている。また、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 [11]。文献 [11] ではその例として “Display categorized instructions and documentation.” を例として挙げ、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が曖昧である点を指摘し、6 種類の方法での検出精度を比較している。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、本論文の分析で対象とする具体化の度合いへの言及はない。

Lehner は Flesch reading ease, Fog index, Cloze procedure といった単語数や音節数など文の構造や読みやすさを計測する既存の指標を用いて要求定義書全体を評価している [6]。Lehner が選んだ指標はルールの事前定義を必要としない点で本研究の分析と共通しているが、具体化度合いに関する言及はない。また、ドキュメントの読者にとっての読みやすさに関連するものであり、本研究の分析とは目的が異なる。

3. ドキュメント評価

3.1 従来の評価手法との比較

本研究で目指すドキュメント評価は、ドキュメントに出現する単語のうち国語辞書に抽象度の定義のあるものを全て利用する。従来の評価手法は、過去の実績等から問題となる可能性がある部分や表現を計算機

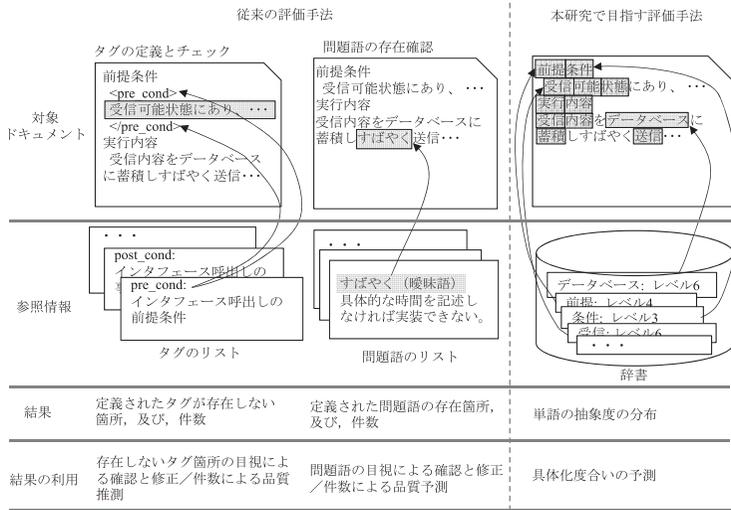


図 1 従来の評価手法との比較
Fig. 1 Comparison to existing approach.

が解釈できるルールとして定義しておき、計算機によりルールを逸脱している箇所を検出する。ルールには曖昧語や誤解されやすい表現を定義しておく。逸脱箇所が明確である点、計算機で解釈できるものであればルールを自由に定義できる点で優れているが、ドキュメントを網羅的にカバーするルールの定義が難しいため評価がドキュメントの一部にとどまる点、ルールを事前に定義するコストが大きい点が課題である。

本研究で目指すドキュメントの具体化の評価は、事前定義した単語と対応する抽象度をルールとみなせば、従来手法と同様に計算機で解釈可能なルールの定義とルールによるドキュメントの評価の二つから構成される。本研究では、ルールに該当する抽象度辞書の定義は国語辞書を流用することにより、ルールの構築コストを低減し、ルールがドキュメントを網羅的に評価できるようにすることを目指す。

図 1 は従来の計算機支援によるドキュメント評価と本研究で目指す評価手法を比較したものである。従来のドキュメント評価手法は、左側の“タグの定義とチェック”と真ん中の“問題語の存在確認”である。タグの定義とチェックは文献 [8] に代表される手法であり、文書作成の際にメタデータを付与しておき、評価の際にそのメタデータを利用する。問題語の存在確認は文献 [7] に代表される手法であり、誤りを引き起こしやすい典型的な表現を事前登録しておき、該当する箇所を問題の候補として開発者に提示し、確認と修正

を促す。QuARS では 5 カテゴリー、35 表現を事前に定義しているが [13]、これらの表現のみでドキュメント全体にわたって網羅的に評価できるわけではない。

3.2 対象ドキュメントと辞書

本研究で対象とするドキュメントは自然言語で記述されているものとし、特定の記法に従っていることを前提としない。以降では、対象ドキュメントを D とし、 D に現れる単語を (d_1, d_2, \dots, d_p) とする。 d_1 はドキュメントの先頭に現れる単語であり、 d_p はドキュメントの最後に現れる単語である。

単語 w の抽象度 $L(w)$ は国語辞書 W の定義を用いて求める。 $W = \{w_1, w_2, \dots, w_m\}$ とし、 w_i には抽象度が定義してあるものとする。抽象度は $L(w_i)$ で得られる。単語 w_i が u_i 個の語義をもつ多義語の場合には、 $L(w_i)$ は、それぞれの意味での抽象度 $(l_1, l_2, \dots, l_{u_i})$ となる。

本研究では単語 d は対象ドキュメントから計算機により抽出でき、 d の抽象度 $L(d)$ が辞書 W で定義されているかどうか計算機により判断できることを前提としている。計算機による抽出の方法には特に前提を置かないが、例えば、形態素解析ツールに対象ドキュメントを与えて得られた d を辞書 W の辞書検索ツールに与え、 $L(d)$ が定義されているかどうか人手を介さず判断できるものとする。 $L(d)$ が定義されている場合には抽象度を得る。図 1 右側の“本研究で目指す評価手法”はドキュメントとそれを構成する単語の例である。

3.3 課題

本研究で目指す評価手法の実現には以下を明らかにする必要があり、本研究では課題に対する分析を実施する。

- 課題 1: W による D の網羅

本研究で目指す評価手法は一般の国語辞書 W を用いることによりルール定義のコストを低減しつつ、ドキュメントの特定箇所に偏らない評価を目指している。しかしながら、ソフトウェアドキュメント D に出現する単語の抽象度が W によりどの程度定義されているかは明らかではない。また、 W によって定義されている単語 d_k がドキュメント D の特定の箇所に偏って存在していないか明らかではない。

- 課題 2: 多義語の抽象度

同一の単語でも文脈によって意味が変わることがある。文脈から単語の意味を推測し、適した語義の抽象度の選択を自動化することは難しい。多義語の場合に (l_1, \dots, l_u) から何らかの方法でその単語の抽象度を得なければならない。

- 課題 3: ドキュメントの具体化度合いによる単語の抽象度の変化

本研究で目指す評価手法は抽象度が異なるドキュメントに含まれる単語の抽象度の分布は異なることを前提としている。具体化度合いが異なるドキュメントに含まれる単語の抽象度が異なっている必要がある。

4. 分析

4.1 対象ドキュメントと辞書

対象ドキュメントはインターネットで公開されている日本語で記述された要求定義書、基本設計書、詳細設計書とした。検索エンジンに“設計書”、“要求定義書”といった検索キーワードを与え検索結果を PDF 形式、Microsoft Word 形式に限定することを指定し、検索結果に含まれたものとした。また、分析 3 で用いる同一ソフトウェアの異なるフェーズのドキュメント以外は 50 ページ以上のドキュメントとした。50 ページ以上とした理由は W による網羅性を確認するためにはある程度の規模以上のドキュメントが必要と考えたからである。同一ソフトウェアの異なるフェーズのドキュメントは公開されているものがほとんどなく、規模による選択はしなかった。ドキュメントの一覧を表 2 に示す。

抽象度を定義する辞書 W として日本語語彙大系 [12] を用いた。日本語語彙大系には 141857 語の名詞が収

表 1 日本語語彙大系段別収録単語例

Table 1 Examples of abstract level and words in the Japanese dictionary.

抽象度 (段)	単語例
1	いずれ、これ、何
2	具体、個体、主体
3	事、現象、主体
4	進捗、データ、組織
5	選択肢、数値、目安
6	程度、条件、順番
7	取得、必須、実装
8	四捨五入、日本語、暗号
9	交付、上書き、定義
10	禁止、メンテナンス、押下
11	採取、男女、同士
12	発行、レイアウト、掲載

録され、それぞれの名詞に 1~12 段の抽象度が定義されている。1 段が最も抽象的、12 段が最も具体的な単語である。多義語には、それぞれの語義に対応する段が定義されている。日本語語彙大系の各々の段に含まれる単語の一例を表 1 に示す。対象ドキュメントを閲覧するアプリケーションが提供する“テキスト形式で保存”機能を用いて本文と表のテキストを得た。得られたテキストから MeCab^(注1)を用いて名詞を抽出した。なお、図中のテキストは対象外としている。

4.2 目的と手順

課題 1, 2, 3 に対応する分析 1, 2, 3 を実施する。分析 1 は網羅性に関する分析、分析 2 は多義語の抽象度に関する分析、分析 3 はドキュメントの具体化度合いの評価に関する分析である。MeCab により名詞と判断された単語を出現順に並べたものを $D' = (d'_1, \dots, d'_q)$ とする。そのうち辞書 W に抽象度 $L(w)$ の定義があるものを $N = (n_1, \dots, n_r), n_i \in W$ とする。

- 分析 1: W による D の網羅率

目的 ドキュメント D の単語 d_k が国語辞書 W によって定義されている割合を明らかにする。また、 W において定義のある単語がドキュメントの特定の箇所に偏っていないか明らかにする。

手順 ドキュメントファイルごとに、出現する単語のうち、日本語語彙大系 W で抽象度が定義されている単語の割合 r/p を計数する。また、ドキュメントの特定箇所に偏った評価になっていないかを明らかにするため、ドキュメントの先頭から出現単語を 300 語、600 語、1200 語で区切り、その区間において辞書で定義されている単語の割合の分布を明ら

(注1) : <https://code.google.com/p/mecab/>

表 2 ドキュメントの URL
Table 2 Documents and URLs.

ID	名称 (上段) URL (下段)
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 http://www.mlit.go.jp/jidosha/topics/oss/yoken_teigi.pdf
R-2	次期 PARTNER システム要件定義書 http://www.jica.go.jp/announce/public/pdf/110809_opi_sum11.pdf
R-a	住宅履歴情報管理システム http://www.cbl.or.jp/slc/pdf/rireki01.pdf
R-b	被災者支援システム改修 要件定義書 http://www.soumu.go.jp/main_content/000120786.pdf
A-1	山梨県共同システム基本設計書 http://www.ya-chos.gr.jp/php/upf/kj1000192tmp1.pdf
A-2	雨量等防災情報提供システム基本設計 https://www.pref.nagano.lg.jp/gijukan/kensei/nyusatsu/kokyokoji/bukyoku/documents/101014sankousyo.pdf
A-3	山梨県立図書館情報システム基本設計書 http://www.pref.yamanashi.jp/toshokan/documents/system-kihonsekkei/documents/system-kihonsekkeisyo.2.pdf
A-4	年金業務システム 基本設計書 http://www.mhlw.go.jp/sinsei/chotatu/chotatu/shiyousho-an/dl/090327-1d.pdf
A-5	地方自治体等における文字情報基盤適用実証実験 基本設計書 http://mojikiban.ipa.go.jp/wp-content/uploads/2013/06/pdf/PMT_Spec.pdf
A-6	病名検索システム 基本設計書 http://www.dis.h.u-tokyo.ac.jp/byomei/dispack/latest/doc/disPACK-spec-031002.pdf
A-7	消防業務支援システム 基本設計書 http://www.city.yokohama.lg.jp/shobo/koukai/ippan-nyusatsu-pdf/shoubou.kihonnsekkeisyotou.pdf
A-b	被災者支援システム改修 基本設計書 http://www.soumu.go.jp/main_content/000120786.pdf
D-1	来訪者受付システム http://www.openrtm.org/openrtm/sites/default/files/4724/_RH%E8%A9%B3%E7%B4%B0%E8%A8%AD%E8%A8%88%E6%9B%B805.10.pdf
D-a	住宅履歴情報管理システム 詳細設計 ソフトウェアプログラム設計書 http://www.cbl.or.jp/slc/pdf/rireki25.pdf

表 3 対象ドキュメント
Table 3 Documents in the analysis.

ID	種別	ページ数	単語数 p	名詞語数 q	名詞種類	定義あり名詞数 r	名詞種類
R-1	要求定義書	59	22771	14200	1293	11917	1058
R-2	要求定義書	100	72609	46392	2183	34419	1418
R-a	要求定義書	74	28539	15912	1596	13156	1244
R-b	要求定義書	11	4524	1803	413	1398	320
A-1	基本設計書	56	21398	14651	1386	11381	997
A-2	基本設計書	60	13446	7923	1144	6522	861
A-3	基本設計書	67	20711	11538	1805	9756	1364
A-4	基本設計書	71	26563	18702	2036	14506	1510
A-5	基本設計書	80	28743	18346	1660	14262	1119
A-6	基本設計書	118	33066	23925	1367	14353	660
A-7	基本設計書	327	134450	93817	4291	79503	2450
A-b	基本設計書	86	32184	19268	1018	13915	669
D-1	詳細設計書	70	8823	7316	799	2747	389
D-a	詳細設計書	56	28026	18002	1050	13457	381

かにする．具体的には，区切り $v = 300, 600, 1200$ において $D_1(d_1, \dots, d_{v-1})$, $D_2(d_v, \dots, d_{2v-1})$, \dots , $D_o(d_{(o-1)v}, \dots, d_p)$, $(o-1)v \leq p \leq ov-1$ に分割し， D_i において $d_k \in W$ となる d_k の個数 r_i を求め，区間ごとに辞書 W で定義されている単語の割合 r_i/v を求める．個々のドキュメントによって異なるが，1200

語はおおむねテキストのみの場合の 1 ページに，600 語は半ページのテキストや図の入った 1 ページ分程度に該当する．300 語は 1/4 ページのテキストに該当し，図が多くの部分占める 1 ページ分程度に該当する．

- 分析 2：多義語の割合と多義語の抽象度の算出方法による結果の比較

目的 分析1で明らかにした国語辞書で定義された単語数 r のうち、何割の単語が多義語であるかを明らかにする。また、本研究で目指す計算機支援による評価手法では、多義語の抽象度を何らかの方法で求めなければならない。分析2ではその第一歩として、それぞれの語義の抽象度の平均値、中央値、最大値、最小値を用いて、多義語の抽象度を求め、四つの方法による抽象度ごとの分布を比較する。

手順 $N = (n_1, \dots, n_r)$ のうち、多義語であるもののみを選び、 $N' = (n'_1, \dots, n'_s)$ とする。また抽象度が定義されている名詞の単語数 r に対する多義語の割合 s/r を求める。 N' は、 $N = (n_1, \dots, n_r)$ のうち、 $L(n_k) = (l_1, l_2, \dots, l_u)$ の u が $u \geq 2$ となる n'_k からなる。また、 N' のうち、全ての語義で抽象度が同一のものを除いた多義語を $N'' = (n''_1, \dots, n''_t)$ とする。 N'' は N' のうち、 $L(n'_k) = (l_1, \dots, l_{u_k})$ において $l_1 = l_2 = \dots = l_{u_k}$ となるものを除いた多義語である。

対象ソフトウェアドキュメント $D = (d_1, \dots, d_k, \dots, d_p)$ の d_k に対し、抽象度 $l(d_k)$ を辞書 W を用いて求める。 $l(d_k)$ は、 $L(w) = (l_1, l_2, \dots, l_u)$ を用いて次のように得る。

$$l(d_k) = \begin{cases} 0 & d \notin W \text{ のとき} \\ l_1 & u = 1 \text{ のとき} \\ f(l_1, \dots, l_u) & u \geq 2 \text{ のとき} \end{cases} \quad (1)$$

$f(l_1, \dots, l_u)$ をそれぞれの抽象度の平均値、中央値、最小値、最大値とし、抽象度と出現頻度の分布を調べる。

- 分析3: 同一ソフトウェアの異なる工程のドキュメントの比較

目的 ドキュメントの具体化によりドキュメントに含まれる単語の抽象度も小さくなる(具体的になる)を確認する。同一ソフトウェアの異なる工程 α, β (α の方が β よりも前工程)のドキュメント D_α, D_β を比較し、 D_β の方が D_α よりも具体的な単語を多く含むか確かめる。多義語の抽象度は分析2と同様に平均、中央、最大、最小を用いて求め、結果を比較する。また、出現頻度上位の単語にそのソフトウェアが対象とするドメイン(システム化やソフトウェアにより実現する業務、機能)の単語が含まれているか確認する。なお、ドキュメントにはソフトウェアが対象とするドメインの単語も含まれる。異なるドメインのドキュメントの比較では、具体化によって抽象度が小さくなっ

ているのか、ドメインが異なるために抽象度が小さくなっているのかを判断できない可能性があるため、同一ソフトウェアの異なる工程のドキュメントのみ(R-a/D-a, R-b/A-bの組)とした。

手順 D_α/D_β を R-a/D-a, R-b/A-b とし、抽象度ごとに単語の出現度数を集計し分割表を作成する。帰無仮説を“ D_α と D_β の抽象度ごとの単語の出現度数は同程度である”とし、フィッシャの正確確率検定を実施し、分布に違いがあるかを確かめる。抽象度ごとの単語の出現度数が同程度であるとはいえない場合には、 D_α, D_β の間で、抽象度別の単語の出現割合を比較する。また、4ドキュメントのそれぞれにおいて出現頻度の大きな単語(上位10位)の出現頻度を求め、ドメインの単語が存在するか明らかにする。

5. 分析結果

5.1 分析1

ドキュメント D に出現する単語 (d_1, \dots, d_p) のうち、辞書 W で抽象度が定義されている単語の割合 r/p を求めた。結果を表4に示す。ドキュメント R-b, D-1を除いて抽象度が定義されている単語の割合は全体の単語数に対しておおむね40~50%となった。R-bにおいて名詞の割合が小さい理由は、他のドキュメントと比較してページ数が小さいことと図や目次に含まれる記号の割合が相対的に大きかったためである。D-1において名詞 q の割合が高い理由は、D-1において機能名称、メソッド名称が羅列してある表が多数あったためである。それらの名称が辞書において抽象度を定義

表4 全単語数に対する名詞 (D')、抽象度が定義された名詞 (N) の割合

Table 4 Percentages of Noun(D') and Noun with defined abstraction level(N).

ID	抽象度が定義された名詞 (r/p)	名詞 (q/p)
R-1	0.523	0.624
R-2	0.474	0.639
R-a	0.461	0.558
R-b	0.309	0.399
A-1	0.532	0.685
A-2	0.485	0.589
A-3	0.471	0.557
A-4	0.546	0.704
A-5	0.496	0.641
A-6	0.434	0.724
A-7	0.591	0.698
A-b	0.432	0.599
D-1	0.311	0.829
D-a	0.480	0.642

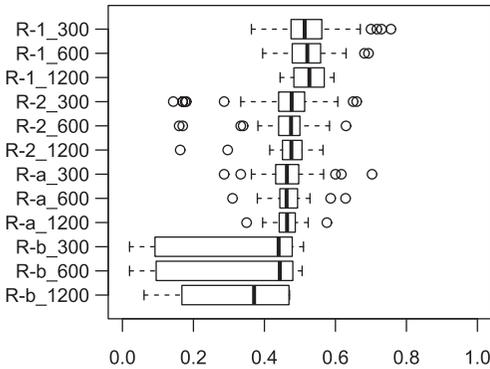


図2 要求定義書における抽象度の定義のある単語の割合
Fig. 2 Percentage of words with defined abstraction level in requirement documents.

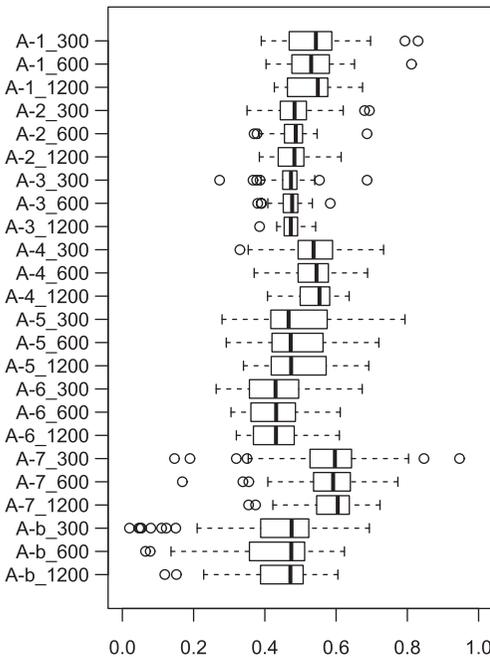


図3 基本設計書における抽象度の定義のある単語の割合
Fig. 3 Percentage of words with defined abstraction level in architectural design documents.

されていない単語であったため、D-1において抽象度が定義された名詞 r の割合が小さくなっていった。

v を 300, 600, 1200 としたときの $D_i(d_{iv}, \dots, d_{(i+1)v-1})$ における r_i/v を求めた結果を図2, 3, 4に示す。図中の横軸は抽象度が定義されている単語の割合 (r_i/v) である。縦軸はドキュメントと v を変化したものであり、例えば“R-1_300”はドキュメント R-1 において $v = 300$ とした場合の抽象度が定義され

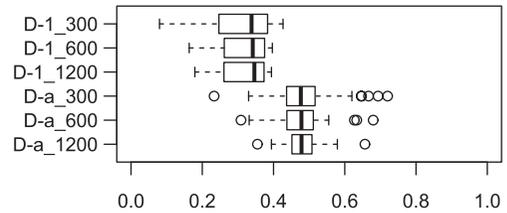


図4 詳細設計書における抽象度の定義のある単語の割合
Fig. 4 Percentage of words with defined abstraction level in detailed design documents.

た単語の割合を表している。

図2, 3, 4から分かるように、いずれのドキュメント、いずれの v においても r_i/v の値がゼロとなったものはなかった。また、いずれのドキュメント、いずれの v においても中央値は0.3より大きかった。また、R-bを除く全てのドキュメントで第一四分位数が0.2よりも大きかった。 r_i/v の値が小さい部分の多くは用語定義やシステムに特化した名称や記号が集中していた。

辞書に抽象度の定義のなかった単語（名詞）として以下があった。

- ソフトウェア・システムに固有の単語
“アラート”, “ログイン”, “帳票” 等
- 複合語の一部: “ブレ”, “セミ” 等
- 数字, 記号: “iii”, “N” 等
- アルファベットの略称, 英単語: “RDBMS”, “Roaming” 等

5.2 分析 2

ドキュメント D に出現し抽象度が定義されている単語のうち、多義語であるものの割合 s/r を求めた。また、多義語のうち、全ての語彙において抽象度が同一のものがあつたため、それを除いた割合 t/r を求めた。結果を表5に示す。表中の各行は一つのドキュメントに対応する。

“平均”, “中央”, “最大”, “最小” の列は f を平均値, 中央値, 最大値, 最小値としたときに単語の出現割合の大きな順に上位6位までの抽象度を記入したものである。ただし、出現割合は5%以上のものを対象とした。平均値では抽象度を整数とするため小数第1位を四捨五入した。例えば、表中のドキュメント R-1 において多義語を平均値とした場合に最も出現頻度が大きい抽象度は8であり、7, 6, 9, 5が続く。5位の抽象度5よりも出現割合の小さい抽象度の単語は出現割合が5%未満であるため、6位は空欄となっている。

表 5 多義語の抽象度の順位
Table 5 Abstraction levels ranking in words with multiple meanings.

ID	多義語の割合 (s/r)	抽象度が複数の 語の割合 (t/r)	平均						中央						最大						最小					
			1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
R-1	0.599	0.590	8	7	6	9	5	8	7	9	6	5	8	9	7	6	10	5	6	8	7	9				
R-2	0.674	0.670	6	8	9	7	5	6	8	9	7	5	9	8	7	6	10	6	5	8	7	9				
R-a	0.685	0.675	6	8	7	9	5	6	8	7	9	5	8	7	9	6	10	6	5	8	7	9	4			
R-b	0.577	0.549	6	7	8	9	5	6	7	9	8	5	8	9	7	6	10	6	5	8	7	9	4			
A-1	0.578	0.569	8	6	9	7	5	6	9	8	7	5	9	8	6	7	10	6	8	5	9	7	4			
A-2	0.628	0.612	6	8	9	7	5	6	9	8	7	5	9	8	7	6	10	5	6	5	7	9	8	4		
A-3	0.637	0.623	8	6	7	9	5	8	6	9	7	5	9	8	7	6	10	6	8	7	9	5	4			
A-4	0.644	0.635	8	6	9	7	5	8	6	9	7	5	8	9	7	6	10	6	8	5	7	9	4			
A-5	0.557	0.550	8	6	7	9	5	8	6	7	9	5	8	9	6	7	10	6	8	7	5	9				
A-6	0.632	0.628	8	7	6	9	5	8	6	9	7	5	8	9	7	6	10	6	7	8	5	4	9			
A-7	0.635	0.632	8	9	6	7	5	8	9	6	7	5	9	8	6	7	10	6	8	7	5	9	4			
A-b	0.628	0.623	6	8	9	7	5	4	6	8	9	7	5	4	5	6	4	5	6	5	9	8	7	4		
D-1	0.650	0.634	7	6	8	9	5	6	7	8	9	5	8	7	9	6			6	5	7	8	9			
D-a	0.652	0.649	8	6	7	9	4	8	6	7	9	4	9	8	7	6	4			6	8	5	7	9	4	

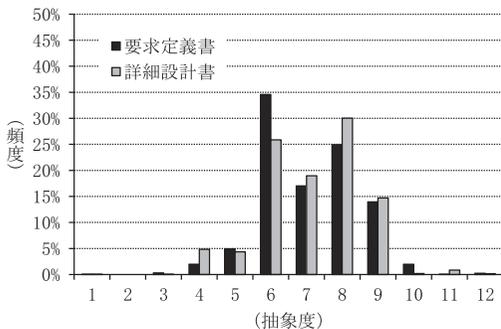


図 5 住宅情報管理システム R-a/D-a での抽象度の分布 (多義語：平均値)
Fig. 5 Distribution of abstraction level in documents R-a and D-a (Average).

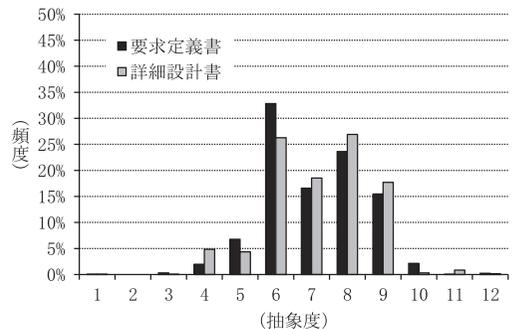


図 6 住宅情報管理システム R-a/D-a (多義語：中央値)
Fig. 6 Distribution of abstraction level in documents R-a and D-a (Mean).

平均値, 中央値, 最小値, 最大値の全てにおいて順位が同一のものはなかった. 全体的な傾向として, 平均値, 中央値, 最大値において上位 4 位に抽象度 6, 7, 8, 9 以外の抽象度はなかった. 最小値では, 5, 6, 7, 8, 9 以外の抽象度はなかった. 平均値と中央値では出現頻度の順位が同一となるドキュメントがあり, R-1, R-2, R-a, A-4, A-5, A-6, A-7, A-b, D-a であった.

5.3 分析 3

ドキュメント R-a/D-a, R-b/A-b の組において, 分析 2 で述べた四つの抽象度の求め方を用いて, 抽象度ごとに単語の出現度数を集計し分割表を作成した. 帰無仮説を“ D_α と D_β の抽象度ごとの単語の出現度数は同程度である”とし, フィッシャの正確確率検定を実施したところ, R-a/D-a, R-b/A-b それぞれにおいて平均値, 中央値, 最大値において有意水準 5%で帰無仮説が棄却された.

2 組のドキュメントの抽象度ごとの割合を図 5 から図 10 に示す. 横軸は抽象度を, 縦軸はその抽象度の単語の v に対する出現割合を表す. なお, 抽象度は日本語彙大系において抽象度として定義されている段数をそのまま使っている. グラフでは右側にあるほど具体的な単語である. 多義語の抽象度を平均値, 中央値, 最大値とした場合において, 両ドキュメントとも抽象度 9 の単語の出現割合が D_α よりも D_β の方が大きかった. また, R-a/D-a の組の最大値を除いて, 抽象度 8 の単語の出現割合が D_α よりも D_β の方が大きかった. R-a/D-a の組では抽象度 6 の単語の割合が, R-b/A-b の組で抽象度 7 の単語の割合が D_β よりも D_α の方が大きかった.

そこで, 多義語の抽象度を平均値, 中央値, 最大値とした場合に出現割合が上位 4 位となる抽象度 6, 7, 8, 9 を対象として, 抽象度 6, 7 の単語を抽象的な単語, 抽象度 8, 9 の単語を具体的な単語と考え, それ

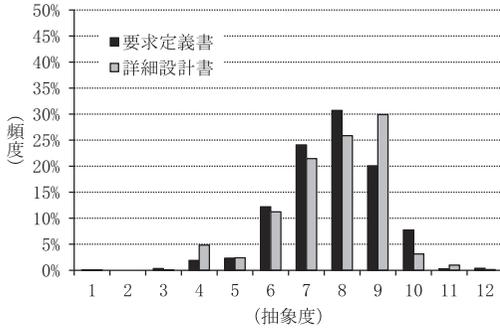


図7 住宅情報管理システム R-a/D-a (多義語：最大値)
Fig.7 Distribution of abstraction level in documents R-a and D-a (Maximum value).

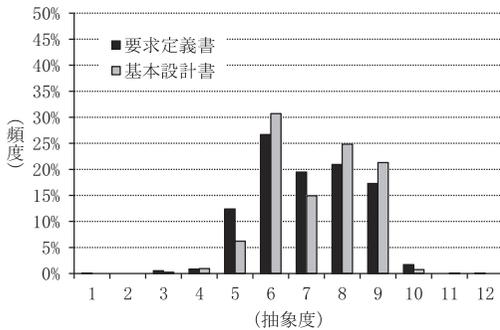


図8 災害被災者支援システム R-b/A-b (多義語：平均値)
Fig.8 Distribution of abstraction level in documents R-b and A-b (Average).

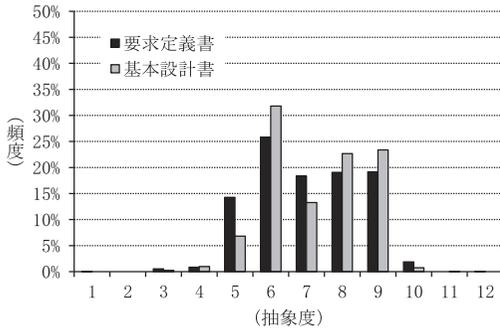


図9 災害被災者支援システム R-b/A-b (多義語：中央値)
Fig.9 Distribution of abstraction level in documents R-b and A-b (Mean).

それぞれの出現割合を計数した。R-a/D-a での結果を表6に、R-b/A-b での結果を表7に示す。表の“6, 7の合計”は $l(w) = 6$ となる単語の数と $l(w) = 7$ となる単語の数の和の r に対する割合である。“8, 9の合計”は $l(w) = 8$ となる単語の数と $l(w) = 9$ となる単語の数の和の r に対する割合である。表に示すように両方の

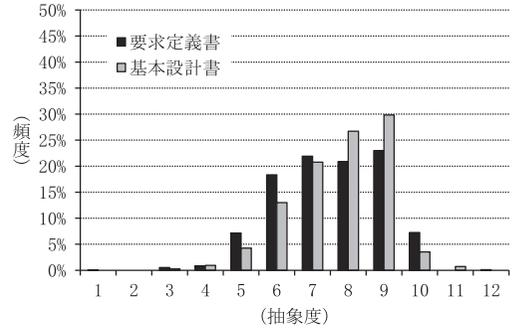


図10 災害被災者支援システム R-b/A-b (多義語：最大値)
Fig.10 Distribution of abstraction level in documents R-b and A-b (Maximum value).

表6 R-a, D-a の抽象度ごとの出現割合
Table 6 Frequency of abstraction level in document R-a and D-a.

抽象度	平均値		中央値		最大値	
	R-a	D-a	R-a	D-a	R-a	D-a
6, 7 の合計	0.516	0.448	0.494	0.448	0.401	0.433
8, 9 の合計	0.389	0.447	0.391	0.446	0.270	0.299

表7 R-b, A-b の抽象度ごとの出現割合
Table 7 Frequency of abstraction level in document R-b and A-b.

抽象度	平均値		中央値		最大値	
	R-b	A-b	R-b	A-b	R-b	A-b
6, 7 の合計	0.505	0.412	0.491	0.394	0.410	0.371
8, 9 の合計	0.359	0.428	0.363	0.425	0.282	0.300

組において、多義語の抽象度を平均値、中央値とした場合に、要求定義書の方が設計書よりも抽象的な単語（抽象度 6, 7）の割合が大きく、設計書の方がより具体的な単語（抽象度 8, 9）の割合が大きかった。

表8, 9に2組のソフトウェアで出現頻度が大きかった単語を示す。表中の“割合”は辞書に抽象度の定義のあった単語数 r に対する各単語の出現頻度の割合である。単一の単語が占める割合は最大で7%程度であった。10位の単語で最大2%程度であり、1種類の単語がドキュメントのほとんどを占めているということとはなかった。

表8, 9から分かるようにソフトウェアが対象とする業務やドメインの単語が D_α, D_β に多く出現していた。R-a/D-a (住宅情報管理システム) では、“情報”、“住宅”といった単語が、R-b/A-b (被災者支援システム) では、“情報”、“防災”といった単語が含まれていた。表8の D_β (詳細設計書) には“ID”、“取得”、“テーブル”、“更新”、“画面”といった実装に必

表 8 R-a, D-a の出現頻度上位の単語
Table 8 Top 10 frequent words in document R-a and D-a.

順位	R-a			D-a		
	単語	抽象度	出現割合	単語	抽象度	出現割合
1	情報	6, 6, 7	0.0754	情報	6, 6, 7	0.0662
2	住宅	5, 5, 6, 8	0.0620	住宅	5, 5, 6, 8	0.0428
3	者	5, 6, 8	0.0349	ID	4	0.0389
4	履歴	7, 7	0.0258	図面	9, 9	0.0302
5	システム	4, 5, 5, 6, 6	0.0194	取得	7	0.0245
6	所有	8	0.0188	テーブル	9	0.0210
7	こと	6	0.0187	履歴	7, 7	0.0198
8	等	5, 5, 7	0.0182	更新	8	0.0194
9	もの	4, 5, 7, 7	0.0115	画面	6, 7, 8	0.0172
10	機関	5, 8	0.0111	登録	9, 9	0.0169

表 9 R-b, A-b の出現頻度上位の単語
Table 9 Top 10 frequent words in document R-b and A-b.

順位	R-b			A-b		
	単語	抽象度	出現割合	単語	抽象度	出現割合
1	情報	6, 6, 7	0.0416	情報	6, 6, 7	0.0401
2	者	5, 6, 8	0.0287	者	5, 6, 8	0.0372
3	防災	9	0.0258	情	9, 9	0.0284
4	ユニット	5	0.0244	防災	9	0.0215
5	検索	8	0.0230	報	6, 7, 9, 9	0.0202
6	システム	4, 5, 5, 6, 6	0.0222	共有	8	0.0172
7	避難	9, 9	0.0215	B	4	0.0170
8	要件	6, 8	0.0215	システム	4, 5, 5, 6, 6	0.0167
9	共有	8	0.0208	ユニット	5	0.0167
10	団体	5, 6, 6	0.0201	被災	7	0.0152

要となる具体的な単語が上位に含まれていた。

多義語の抽象度を平均値, 中央値, 最大値により求めた場合に出現割合が上位 4 位となる抽象度 6, 7, 8, 9 において抽象的な単語 (抽象度 6, 7) として, R-a では“情報”, “履歴”, “こと”が, R-b では“情報”, “団体”があった。同様に R-b での具体的な (抽象度 8, 9) 単語として, D-a では“図面”, “テーブル”, “更新”, “登録”が, A-b では“情”, “防災”, “共有”があった。

6. 考 察

6.1 分析結果

分析 1 では, どのドキュメントにおいても 30%以上の語が辞書に定義されていることが分かった。また, 全てのドキュメントと v において, r_i/v がゼロになっている区間はなかった。分析の対象とした 14 件のドキュメントにおいて, 本研究で目指す評価はドキュメントの特定箇所に偏った評価とはならないことを示す結果が得られた。なお, 抽象度が定義されている名詞 r を分子とし, MeCab により名詞と判断された単語数 q を分母とすることもできたが, D_i の区切り v と

揃えるため単語数 p を用いた。 D_i の区切りを MeCab により名詞と判断された単語数 q と揃えなかった理由は, ドキュメントの特定箇所に MeCab により名詞と判断された単語が偏っている可能性を排除するためである。

分析 2 では, 語義によって抽象度が異なる多義語 N'' の割合が抽象度が定義されている名詞 N の 50%を超えることが分かった。文脈を解釈して適した語義を選択する方法を自動化するのは容易ではないため, 多義語の扱いは今後の重要な課題の一つである。現段階では, 平均値, 中央値, 最小値, 最大値といった方法をとったり, ソフトウェア開発の文脈では, 明らかに使われない語義がある場合には, 辞書の抽象度を再定義したりする方法が考えられる。

分析 3 では, 2 組のドキュメントにおいて, 多義語を平均値, 中央値, 最大値とした場合に抽象度ごとの単語の出現頻度が異なることが分かった。また, 多義語の抽象度を平均値, 中央値とした場合に, 抽象度 6 と 7 の単語の割合が設計書よりも要求定義書の方が大きかった。逆に抽象度 8 と 9 の単語の割合は要求定義書よりも設計書の方が大きかった。つまり, 要求定義

書の内容を具体化した設計書では、より具体的な単語の出現割合が大きく、抽象的な単語の出現割合が小さかった。基本設計書や詳細設計書の作成後や作成途中で抽象度 6, 7 及び, 8, 9 の単語の出現割合を要求定義書と比較することにより、目視なしで設計書の具体化度合いを評価できる可能性がある。

6.2 妥当性

同一ソフトウェアの異なるフェーズのドキュメントが公開されているものがほとんどみつからず、分析 3 の対象ドキュメントは 2 組だけであった。より多数の同一ソフトウェアの異なるフェーズのドキュメントにより評価し、分析 3 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一つである。分析 3 で対象としたドキュメント D_α から D_β が順調に具体化されたかを公開情報だけから推測することは難しいが、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、信頼できるものであると考える。

辞書による抽象度の定義とソフトウェア開発における抽象度が乖離しているものには抽象度の再定義が必要となる可能性がある。今回使用した辞書では“ID”のようなソフトウェア開発においては具体的といえる単語の抽象度が 4 と定義されていた。そういった辞書による抽象度とソフトウェア開発における具体化度合いに乖離がある単語の出現割合が大きくなると分析 3 のような出現割合の比較が難しい可能性がある。

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 や Word から閲覧、編集アプリケーションソフトウェアの機能を使って、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。

7. むすび

事前に多くのルールを準備することなくソフトウェアドキュメントの具体化度合いを網羅的に評価することを目指して、単語の抽象度が定義された国語辞書を用いてドキュメントに出現する単語の抽象度を計測する上で課題となる点を実際のドキュメントと辞書を用

いて分析した。課題は (1) 辞書において抽象度が定義されている単語によってドキュメントを網羅的に評価できるか、(2) 語義を複数もつ単語の抽象度をどのように求めるか、(3) ドキュメントの具体化により含まれる単語の抽象度の分布に変化があるか、である。

課題 (1), (2) を検討するために、インターネット上で公開されている 14 件のドキュメントと単語の抽象度が定義されている日本語語彙大系を用いて実証的に分析した。日本語語彙大系には収録されている名詞に対して 1 から 12 までの抽象度 (段数) が定義されている。抽象度 1 が最も抽象的である。14 件のドキュメントでは、出現単語のうち、30%以上の単語の抽象度が日本語語彙大系で定義されており、14 件のうち 12 件のドキュメントでは 43%を超えていた。また、抽象度が定義されている単語がドキュメントの特定箇所に偏って存在しておらず、本研究で目指す評価手法がドキュメント全体を網羅的に評価できる可能性が示された。また、複数の語義をもち、それぞれの語義で抽象度が異なる単語の抽象度を、平均値、中央値、最大値、最小値として比較した。

課題 (3) を検討するために、14 件のドキュメントのうち、同一のソフトウェアの要求定義書、設計書の組 2 件を対象として、抽象度ごとの単語の出現頻度を比較した。多義語は課題 (2) の検討のときと同様に、平均値、中央値、最大値、最小値の四つの方法で求めた。帰無仮説を“要求定義書、設計書に含まれる単語の抽象度ごとの出現頻度に違いはない”とし、フィッシャの正確確率検定を実施したところ、平均値、中央値、最大値において有意水準 5%で帰無仮説が棄却された。多義語の抽象度を平均値、中央値として、出現割合が上位 4 位となる抽象度 6, 7, 8, 9 において、2 組のドキュメントで、抽象的な単語 (抽象度 6 または 7) の出現割合が設計書よりも要求定義書の方が大きかった。また、具体的な単語 (抽象度 8 または 9) の出現割合は要求定義書よりも設計書の方が大きかった。

課題 (1) の検討の結果から、事前に多くのルールを準備することなく国語辞書を用いてソフトウェアドキュメントを網羅的に評価できることが期待できる。課題 (3) の検討の結果から、ドキュメントの具体化度合いは抽象度ごとの単語の割合に現れることが示された。既存の国語辞書を用いてドキュメントの評価ができれば、これまで難しかった、準備コストを小さくしつつ、目視を要しない網羅的なドキュメント評価が期待できる。例えば、委託開発の検収の手間や定例の進

捗会議の手間を大きく削減できたり、ドキュメント作成者がセルフチェックとして自身で作成したドキュメントを評価することにより、工程終盤のドキュメントの大幅な修正の予防が期待される。

課題 (2) の多義語の抽象度の決定方法は、平均値、中央値において効果が期待できるが、より適した方法を検討することは今後の課題である。また、公開されているドキュメントでは同一ソフトウェアの異なるフェーズのドキュメントが少なく、課題 (3) の検討には2組のドキュメントしか対象とできなかった。抽象度ごとの単語分布にはソフトウェアが対象とするドメインで頻出する用語の抽象度の影響を受けやすいことが予想されるため、より多くのドキュメントで課題 (3) を検討することが必要である。

謝辞 立命館大学大西淳先生の抽象度辞書に関するご助言に感謝する。本研究は文部科学省科学研究補助費 (基盤研究 B:課題番号 23300009) による助成を受けた。

文 献

- [1] M.E. Fagan, "Design and code inspections to reduce errors in program development," IBM Systems Journal, vol.15, no.3, pp.182-211, 1976.
- [2] F.Shull, I. Rus, and V. Basili, "How perspective-based reading can improve requirements inspections," Computer, vol.33, no.7, pp.73-79, 2000.
- [3] S. Kusumoto, T. Kikuno, K. Matsumoto, and K. Torii, "Experimental evaluation of time allocation procedure for technical reviews," Journal of Systems and Software, vol.35, no.2, pp.119-126, 1996.
- [4] N. Nagappan, L. Williams, J. Hudspohl, W. Snipes, and M. Vouk, "Preliminary results on using static analysis tools for software inspection," 15th International Symposium on Software Reliability Engineering, pp.429-439, 2004.
- [5] T. Thelin, H. Petersson, P. Runeson, and C. Wohlin, "Applying sampling to improve software inspections," Journal of Software and Systems, vol.73, no.2, pp.257-269, 2004.
- [6] F. Lehner, "Quality Control in software documentation based on measurement of text comprehension and text comprehensibility," Information Processing and Management, vol.29, no.5, pp.551-568, 1993.
- [7] M. Wilson, H. Rosenberg, and E. Hyatt, "Automated analysis of requirement specifications," Proc. Nineteenth International Conference on Software Engineering, pp.161-171, 1997.
- [8] 大杉直樹, 並川 顕, 小橋哲朗, 重木昭信, 木谷 強, 山本修一郎, "記述漏れと曖昧な表記の防止を目的とした要件定義書の第三者スコアリングに向けた試み," ソフトウェア品質シンポジウム 2009, 2009.
- [9] H. Yang, A. Willis, A. Roeck, and B. Nuseibeh, "Automatic detection of nocuous coordination ambiguities in natural language requirements," Proc. International Conference on Automated Software Engineering, pp.53-62, 2010.
- [10] R.A. Fleisch, "A new readability yardstick," Journal for Applied Psychology, vol.32, pp.221-223, 1948.
- [11] F. Chantree, B. Nuseibeh, A. Roeck, and A. Willis, "Identifying nocuous ambiguities in natural language requirements," Proc. International Conference on Requirements Engineering, pp.59-68, 2006.
- [12] 日本語語彙大系 CD-ROM 版, 岩波書店, 1999.
- [13] G. Lami, "QuARS: A tool for analyzing requirement," Carnegie Mellon University Technical Report, CMU/SEI-2005-TR-014, 2005.

(平成 26 年 3 月 28 日受付, 8 月 4 日再受付)



森崎 修司 (正員)

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助教。2013 年 4 月静岡大学准教授。2013 年 10 月名古屋大学准教授。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)。情報処理学会, IEEE 各会員。



遠藤 充

2010 年静岡大学情報学部情報科学科入学。2014 年静岡大学情報学部情報科学科卒業。ソフトウェアレビューに興味をもつ。



杉山 岳弘 (正員)

1997 年静岡大・工・情報知識工学科卒。1994 年同大学院修士課程修了。1997 年同大学院電子科学研究科博士後期課程修了。現在, 静岡大学大学院情報学研究科准教授。博士 (工学)。情報システム設計, コミュニケーション分析, 映像メディア処理, 知識コンテンツに関する研究に従事。情報処理学会会員。