

ソフトウェア要求仕様書分析と対応する 既存仕様回答パラグラフ自動抽出システムの開発

岡田伊策^{1,2} 齋藤稔¹ 笈田佳彰^{1,2} 稗方和夫² 中村覚²

Isaac Okada^{1,2}, Minoru Saito¹, Yoshiaki Oida^{1,2}, Kazuo Hiekata², and Satoru Nakamura²

¹ 富士通株式会社

¹ FUJITSU LIMITED.

² 東京大学

² The University of Tokyo

アブストラクト: 世界貿易機関 (WTO) 「政府調達に関する協定を改正する議定書」(2012 年) に対応して、日本の公共調達の仕組みは、精緻かつ詳細に規定された。例えば情報システムに係る調達では、調達仕様に精緻に対応した技術仕様回答書が必要となった。技術評価点が総合評価方式の中で重視され、入札者側は、調達仕様内容を網羅的に理解し、それに精緻に対応する技術仕様回答技術文書を作成することが重要になった。このため入札提案の熟練者に作業が集中するようになったが、熟練者数は不十分で極度に多忙である。熟練者の養成には、長い期間と高いコストがかかる。熟練者を模して、調達仕様から要件を自動抽出、過去の仕様回答実績群から当該要求と類似要求に対応するパラグラフも自動抽出して類似度の高いものを再利用、技術仕様回答書の草稿を短時間に作成するシステムを開発し、初期評価実験として機能動作確認した。

1. 目的

「情報システムに係る政府調達」の仕組みは、最近著しく緻密化・詳細化してきた。^[1] 例えば、図 1 に示す政府機関の調達仕様書例では、調達仕様書本編が 36 ページ、見出し項目が 111 項目あり、別添資料が 148 ページある。大量かつ多岐に渡った内容である。

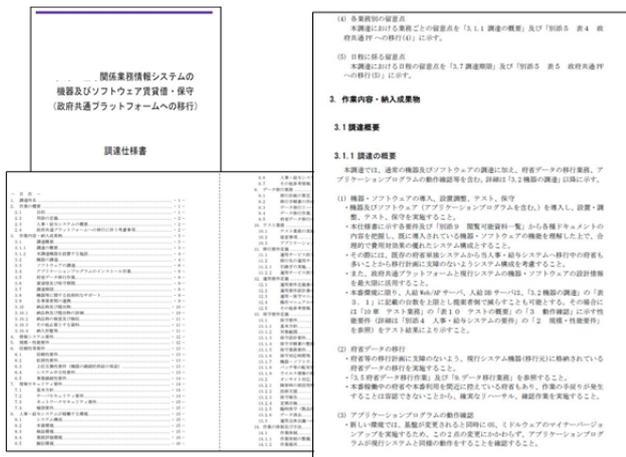


図 1 政府機関の調達仕様書例

これに対応して、入札側の熟練者の経験に基づく調達仕様書 (情報システムソフトウェアへの要求仕様書: 以下 RFP) 分析と、技術仕様回答書 (以下、技術提案書) 作成時の効率的な情報検索・再利用行

動様式を解析・手順化して、非熟練者でも RFP 解析と技術提案書草稿作成できるように支援するシステムを開発した。具体的には、RFP の要件に基づいて、既存の提案実績 DB から、過去に熟練者が作成した該当・類似のパラグラフを探索・提示・再利用して、短時間に草稿を作成できるようにすることを目的とした。

2. 背景と課題

世界貿易機関 (WTO) 「政府調達に関する協定」(Agreement on Government Procurement: 略称 GPA) は、2012 年 3 月 30 日 「政府調達に関する協定を改正する議定書」(「改正議定書」) を採択した。^[2] 日本もその条約締結国である。これを受けて日本では政府を中心に、例えば「情報システムに係る政府調達」および公共調達において、一層の公開競争の厳密化・詳細規定化・精緻化を決定した。^[3]

具体的には、情報システムに係る総合評価落札方式の徹底 (競争評価の客観化・公開化) し、入札説明書 (総合評価基準を含む) に準拠して、RFP に精緻に対応した技術提案書 (図 2) の作成が必要となった。^[4]

調達者による評価は、技術点の配分が高まる傾向にあり、受注競争に勝てる文書作成には、多大な時間と手間がかかっている。技術評価に勝たんがために技術提案書作成は、必然的に既存の技術パラグラフに精通した熟練者に集中している。熟練者はその数

が不十分で極度に多忙である。しかし、熟練者の養成には時間がかかり、多忙な熟練者による教育指導など、非効率所以の高い養成コストがかかっている。



図 2 技術提案書例

RFP の記述には、より良い提案を引き出すために、敢えて冗長であったり、逆に曖昧に表現していることもあり、それらを精緻に分析して、回答すべき要素を網羅的に抽出して、必要十分な技術提案書を作成する必要がある。毎回技術提案文章を書き起こすのではなく、経験と熟練によって、過去の実績ある技術文章を検索・再利用すれば、少なくとも草稿作成は短縮できる。

また、情報システム調達における技術提案書では、企業紹介／納入実績／セキュリティリスク対策／知財権対策など、いずれの技術提案書でも必要となる共通記載事項がある。これらは提案毎の差別化ではなく、提案者側の技術力証明であるので内容は一定となる。本来、再利用可能なはずだが、結果的に属人的に作られた過去の当該内容を再利用掲載している。非効率であると同時に、同じ組織が作成・提出する技術提案書でも、回答者により内容が揺らぐリスクがある。

3. 解決アプローチ

解決施策として、熟練者の作業を(1) RFP を読解・分析するプロセスと(2) 既存の提案実績を検索・再利用するプロセスに分けて、それぞれの手順を解析、それを模すことにした。(図 3)

そもそも、RFP に各々特徴があるため、既存の実績ある技術提案書を、「章」などの比較的大きい単位で再利用しようとしても、RFP の特徴に精緻に適合することを意図しているため、章の中に再利用のできる部分とできない部分が混在して利便性が悪い。逆に文章を一文ずつ吟味して再利用すると、細かすぎて、組み立てに手間と時間がかかる。

そこで、熟練者が「より細かいが文意が整って流

用しやすい」としている「パラグラフ (節、段落)」単位で流用できる方策をとることにした。

(1) RFP の自動パラグラフ分解・名詞抽出・ベクトル化、目次抽出

①今回対象とする RFP を解析、熟練者の RFP 解析視点を模して、まずパラグラフに分解、各パラグラフの TF 値から名詞を抽出してベクトル化する。また、②当該 RFP の目次を抽出して、技術提案書の目次とした。

このように、非熟練者が技術提案に必要な提案要素を見落とさないように、網羅的な抽出を機械支援する。

(2) 既存提案実績文書群のパラグラフ分解・名詞抽出・ベクトル化

一方、熟練者が作成した既存の提案実績文書群の文書もパラグラフ毎に分解、分解した各パラグラフの TF 値から名詞を抽出してベクトル化する。

(3) RFP の各パラグラフと、類似する既存提案実績のパラグラフを抽出

RFP の各パラグラフ名詞ベクトルと、既存提案書のパラグラフ名詞ベクトルの文書間の類似度計算で最も一般的なコサイン類似度を用いて算出してスコアの高いものを提示する。

過去の受注できた提案実績を参照・再利用すれば、新規作成時の試行錯誤による無駄の発生に比して、より熟練者の知見を効率的 (短時間/低コスト/高品質) に再利用できると想定した。

(4) 目次順に抽出したパラグラフを並べて技術提案書の草稿作成

(1) ②で作成した目次順に、(3) で提示した 5 件の既存提案実績パラグラフを並べて、技術提案書の草稿が整う。後は、修整して技術提案書が完成する。

RFP を自動パラグラフ分割した各パラグラフと、それらに対応した熟練者の実績ある類似のパラグラフを複数候補抽出して、非熟練者に相似ポイント/相違ポイントを熟読させれば、非熟練者にとっても具体的な題材ある勉強になり、非熟練者を熟練者並に養成する効率を向上できる。漠然とした流用ではなく、構成要素を識別・理解して再利用させることにより、より良質な実績の反復再利用増加を誘導する

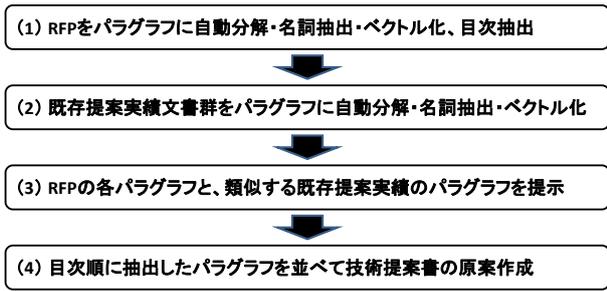


図 3 解決施策フロー

4. 関連研究

先行研究として、森本^[5]は、文書化知識の再利用を3つのプロセス「収集・蓄積プロセス」、「検索プロセス」「抽出プロセス」に分類した。各プロセスの課題を解決するアプローチとして「絞り込み検索機能」「対話誘導機能」「リンク型閲覧ビュー機能」を開発し効果検証した。

OASIS「Darwin Information Typing Architecture (DITA) Version 1.2」^[6][6]では、文書を4つのタイプ（「トピック」）のモジュールに分類。それぞれの記述様式を規定した。（図4）

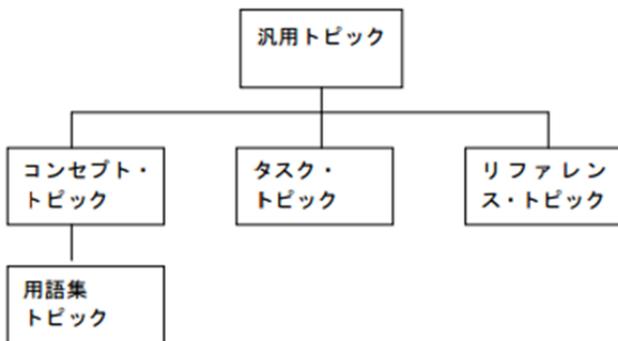


図 4 DITA 1.2 4つのタイプのモジュール

※OASIS : Organization for the Advancement of Structured Information Standards,（構造化情報標準促進協会）e-ビジネス標準の開発、統合および採用を推進する非営利国際コンソーシアム。<http://dita.xml.org/>

山岡、秋吉^[7]は、提案要求者の特性を表す指標と提案文書の部分的内容との対応関係を分析することにより抽出した提案項目の選別規則に基づいて既存文書を部分的に再利用することにより、顧客特性にあった提案書を効率よく作成する手法を提案した。

いずれの研究も、本研究が提案するような要求仕様と既存の実績を分解して、パラグラフ単位の類似度を判定したものではない。

また稗方、岡田、笈田^[8]は、スライド素材の分

割・各スライドへのメタデータ自動付与による既存素材の再利用を効率化する文書管理システムを開発したが、あくまでも素材検索者による検索を前提にしている。

本論文が提案する、RFP という与件を自動的に解析して自動的に適合する過去の実績パラグラフ素材を提示するものではない。

5. 開発したシステム

具体的な実装システムとして「ソフトウェア要求仕様書分析と対応する既存仕様回答パラグラフ自動抽出システム」を開発した。（図5）

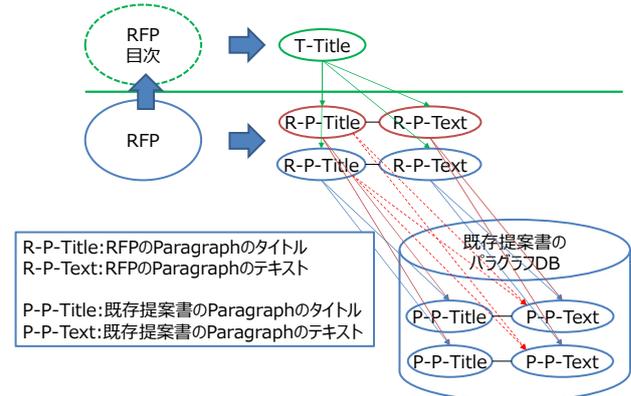


図 5 処理イメージ

- (1) RFP のパラグラフ自動分割と各パラグラフのパラグラフベクトル化



図 6 開始画面 (RFP の指定)

RFP を読み込んで（図6）、Apache POI でパラグラフに自動分割。これを形態素解析して、パラグラフのタイトルと本体をそれぞれベクトル化する。

$$V_{RFP \text{ パラグラフタイトル}} = \langle \text{名詞}, \text{tf 値} \rangle \quad (\text{式 1})$$

$$V_{RFP \text{ パラグラフ本体}} = \langle \text{名詞}, \text{tf 値} \rangle \quad (\text{式 2})$$

No	Word	TF
1	品質	59
2	保証	54
3	仕様	32
4	作業	15
5	システム	15
6	審査	13
7	製造	12
8	実施	11
9	開発	11
10	手順	11
11	ため	10
12	体制	10
13	管理	9

図 7 TF 値確認画面

- (2) 既存実績提案群をパラグラフに自動分割、各パラグラフのタイトルと本体をベクトル化 (図 7)

$$V_{\text{既存パラグラフタイトル}} = \langle \text{名詞, tf 値} \rangle \quad (\text{式 3})$$

$$V_{\text{既存パラグラフ本体}} = \langle \text{名詞, tf 値} \rangle \quad (\text{式 4})$$

- (3) RFP のパラグラフに対応した既存各パラグラフを、コサイン類似度を用いて抽出
- $$\text{Sim}(V_{\text{RFPパラグラフタイトル}}, V_{\text{既存パラグラフタイトル}}) = A \quad (\text{式 5})$$

$$\text{Sim}(V_{\text{RFPパラグラフ本体}}, V_{\text{既存パラグラフ本体}}) = B \quad (\text{式 6})$$

$$A + B \quad (\text{式 7})$$

- (4) RFP 各パラグラフとコサイン類似度が高い既存パラグラフを、当該 RFP に対して回答適合度が高いパラグラフとして提示。(図 8) (図 9)

単語	Relevance	候補文
【提案の要旨】	1.9487	【候補文確認】
品質保証組織と各工程での品質保証のための作業概要	2.0000	【候補文確認】
品質保証のための作業手順-ルール	2.0000	【候補文確認】
審査への報告	2.0000	【候補文確認】
【提案の目的】	1.9535	【候補文確認】
品質保証組織と各工程での品質保証のための作業概要	2.0000	【候補文確認】
開発体制は独立した品質保証組織	2.0000	【候補文確認】

図 8 照合結果表示

No	resourceID	Relevance	提案者の文章	文章
1	355	1.9487	代官業連合会に対する「品質保証組織」に関する提案について行われます。	代官業連合会に対する「品質保証組織」に関する提案について行われます。
2	530	1.9522	テスト実施におけるその他の提案事項	本項では、テスト工程における品質向上するための提案を行います。
3	416	1.9565	テスト実施におけるその他の提案事項	弊社は、総合システムにおける品質向上の取り組みとして品質向上と性能向上を目的とし、品質向上と性能向上を目的としたシステム全体の品質向上を行います。
4	43	1.0000	システム運用管理機能で収集するイベントログ情報を「表 158」のシステムイベントログ情報に付します。	システム運用管理機能で収集するイベントログ情報を「表 158」のシステムイベントログ情報に付します。
5	228	1.0000	性能向上とその評価(信頼性)の向上を図ります。	性能向上とその評価(信頼性)の向上を図ります。

図 9 再利用候補確認画面

- (5) 提示結果を、RFP の目次の順に並べて、技術提案書草稿とする

図 10 技術提案書草稿

6. 初期評価実験

初期評価実験は、既存の「A 機関からの RFP」の目次をベクトル化したものと、それに対応した「A 機関向け技術提案書」をパラグラフに自動分割し、各パラグラフのタイトルと本体をベクトル化した

RFP の目次の各タイトルベクトルを入力として、既存各パラグラフタイトルとのコサイン類似度と、既存パラグラフ本体とのコサイン類似度を算出。その和の大きいものを適合するパラグラフとした。

その結果を、元の熟練者が人手で作成した技術提案書と比較して、システムが目次ベクトルを元に同様のパラグラフを抽出できたかどうか、相違点・類似点を目視確認した。(図 10)

7. 実験結果と分析

RFP にあわせた、提案書体裁のパラグラフ出力は出来た。

しかし「既存仕様回答パラグラフ自動抽出システム」が提示した技術提案書は、元の提案書の約半分程度のデータしか一致せず、元の提案書の半分程度の再現率となった。当初目標とする70%以上の再現性にはいたらなかった。

8. 今後の課題と予定

現状、結果が半分程度に低迷していることに対して、2つの対策を講じている。

- (1) RFP から抽出した目次もベクトル化して、RFP のパラグラフや既存パラグラフのとも照合して、適合度の計算ロジックを豊富にして、適合精度を向上する。
- (2) TF 値を算出するための形態素解析が汎用的な簡単な名詞しか抽出しないため、情報システム固有な複合名詞などの業務用語や、逆にノイズになるストップワードを登録して、文書ベクトルをより精緻にして、適合度の向上を図る。

また、現場実データを使用した実験を繰り返して、チューニングや使用するアルゴリズムの改善を実施していく。

9. まとめ

RFP を解析し、RFP の目次にあわせて、既存の熟練者が作成した技術提案書のパラグラフから適切な素材を抽出する仕組みは開発できた。

今後は、テストデータを豊富に用意して、適合率を70%程度まで向上することを目指す。

10. 参考文献

- [1] [調達関係省庁申合せ]: 「情報システムの調達に係る総合評価落札方式の標準ガイドライン」, e-Gov, http://www.e-gov.go.jp/doc/pdf/03_guideline.pdf, 2013年(平成25年)7月19日, 2015年10月18日確認
- [2] 外務省: 「WTO 政府調達協定」, <http://www.mofa.go.jp/mofaj/gaiko/wto/chotatu.html>, 平成27年8月28日, 2015年10月18日確認
- [3] 「情報システムに係る政府調達の改善」, e-Gov, <http://www.e-gov.go.jp/doc/improve/>, 2015年10月18日確認
- [4] 経済産業省: 「総合評価落札方式 総合評価落札方式ガイドブックー調査、広報、研究開発ー」: <http://www.meti.go.jp/information/downloadfiles/c60815a-3j.pdf>, 2015年10月18日確認
- [5] 森本 由起子(株式会社日立製作所システム開発研究

所), 「文書化知識の再利用支援技法に関する研究」大阪府立大学博士学位論文, 2007,

- [6] OASIS 「 Darwin Information Typing Architecture (DITA) Version 1.2 」, <http://docs.oasis-open.org/dita/v1.2/os/spec/DITA1.2-spec.pdf>, 2010年, 2015年10月18日確認
- [7] 山岡 孝行, 秋吉 政徳, (三菱電機株式会社先端技術総合研究所), 「文書内容の分節化と再利用による個別提案文書構成手法」, 情報処理学会研究報告「情報システムと社会環境」, 2003-03-14,31号, 51-57
- [8] 岡田 伊策, 齋藤 稔, 松岡 伸治, 笈田 佳彰, 大和裕幸, 稗方 和夫, 「プロジェクト提案のための文書情報管理システムの開発と実用化」, IE3-3, 2013年度人工知能学会全国大会(第27回) JSAI2013, 2013年6月