

非機能要件に着目した Request For Proposal (RFP) 評価



齊藤 康廣[†]



門田 暁人[†]



松本 健一[†]

アブストラクト

本論文では、ベンダへの提案依頼書 (Request For Proposal: RFP) の品質を定量的に評価する方法を提案する。評価対象は、提案依頼者となるソフトウェア発注者にとって重要度の高い「保守と運用に関する 55 個の非機能要件 (NFR)」であり、評価の観点はその記述の明確さである。評価結果は、RFP の「総合評価点」と要件毎の評価点を俯瞰するための「レーダーチャート」として示される。地方自治体、図書館、政府機関、大学、病院などが WWW 上に公開していた 5 ドメイン 29 件の RFP を評価対象としたケーススタディによって、記述が不十分な要件を特定したり、基準値との比較を通じて特に改善が必要な特性を明らかにしたりできることなどが確認された。

Evaluation of Request For Proposal (RFP) Focusing on Non-Functional Requirements

Yasuhiro Saito[†], Akito Monden[†], Kenichi Matsumoto[†]

Abstract

This paper proposes a method to evaluate the quality of Request For Proposal (RFP). The proposed method especially focuses on 55 Non-Functional Requirements (NFRs) on maintenance and operation, which are very important for most clients of software development. Evaluation criterion is clarity of descriptions of these NFRs. The evaluation result of the proposed method consists of two parts; overall rating and radar-chart. As a result of a case study of applying the proposed method to 29 RFPs that have been published on the Internet and can be classified into six application domains, we confirmed that the proposed method can identify NFR descriptions which are not so clear and have to be revised in comparison with evaluation criteria.

1. はじめに

提案依頼書 (Request For Proposal, 以後は RFP とする) は、ソフトウェア開発を委託するにあたり、委託元企業 (ユーザ) が、委託先候補の企業 (ベンダ) に対して、開発に関する具体的な技術提案 (技術仕様・技術提案書の作成) を依頼する文書である。RFP には、機能要件、非機能要件、事務要件、システム要件、ライセン

ス事項、開発者資格、契約要件などが記述されている。ユーザは、提示された技術仕様・技術提案書に基づいてベンダを選定し、契約仕様書の作成、契約の締結を経て、

【脚注】

[†] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

ソフトウェアの開発作業が開始されることになる。RFPは、ソフトウェアの委託開発のベースとなる、重要な文書の一つであり、その品質が、ソフトウェア開発の成否を大きく左右することになる [Roth]。

RFPは多様な情報で構成されているが、品質評価の重要な対象の一つとなるのが、「非機能要件 (Non Functional Requirements: NFR)」である。NFRは、開発すべきソフトウェアのアーキテクチャに対する制約条件となり、アーキテクチャの実現可能性に大きく影響する。アーキテクチャは、ソフトウェア品質を決定する主要因の一つとされている [Kazman]。更に、開発開始後のアーキテクチャ変更が容易でないことから、RFPに基づく技術仕様・技術提案書の作成において、アーキテクチャの策定やその実現可能性の評価は、ベンダにとって極めて重要な作業の一つとなっている。NFRが明確に記述されているか否かは、RFP品質を議論する上で重要な観点の一つと言える。

本論文では、ベンダへの提案依頼書 (RFP) 提示に先立ち、RFP作成者であるユーザ自身が、RFPの品質を定量的に評価する方法を提案する。評価対象とするのは、RFPで示されるべき非機能要件 (NFR) であり、評価の観点は、その記述の明確さ、である。RFPに記述すべきNFRを示すガイドラインや報告書、あるいは、NFRを評価するためのメトリクスは、これまでも数多く提案されている [IPA2007] [IPA2010] [JUAS] [JUAS2] [MEXT] [Nikkei] [TRM]。本論文で提案する方法は、それら既存のガイドライン、報告書、メトリクスを基盤として、RFPに記述すべきNFRを、より委託元企業 (ユーザ) の視点で評価する手順を示すものである。具体的には、評価対象を、ユーザにとって重要度の高い「保守と運用に関する55個の非機能要件」に限定した上で、要件記述の明確さを最大5段階で評価するためのメトリクス (評価基準スキーム) を定義し、評価結果は、RFPの「総合評価点」と要件毎の評価点を俯瞰するための「レーダーチャート」として示すものとする。

以降、2章では、関連研究として、NFRに関する代表的なガイドライン、報告書、メトリクスを紹介する。3章では、提案法を示し、4章では、WWW上に公開されていた29件のRFPを対象としたケーススタディの結果を示し、提案法の適用容易性や有用性について議論する。最後に、5章では、まとめと今後の課題について述べる。

2. 関連研究

2.1. ガイドライン・報告書

日本情報システム・ユーザー協会 (JUAS) による「非機能要求仕様定義ガイドライン」[JUAS]には、ソフトウェアライフサイクルを通じて使用することが推奨され

る200個を超える非機能要件が、ISO/IEC9126等に準拠する形で示されている。ただし、ソフトウェア開発終了後の保守や運用に関する非機能要件は、必ずしも網羅されていない。一方、「システム構築のトラブルを回避するためのITシステム契約締結の手順とポイント」[Nikkei]、および、「情報システム調達のための技術参照モデル (TRM)」[TRM]は、ユーザとベンダ間でソフトウェア開発契約を締結する上で重要となる、サービスレベルに関する合意 (Service Level Agreement: SLA) に必要な要件を示すとともに、保守と運用に関する非機能要件も数多く示されている。提案法では、これら3つのガイドラインで示された非機能要件を、評価対象の候補とする。

「システム/ソフトウェア製品の品質要求定義と品質評価のためのメトリクスに関する調査報告書」[MEXT]には、利用者ニーズに応えるソフトウェア品質の確立、および、そのために広く利用可能なメトリクスの選定を目的とする事例調査の結果がまとめられている。報告には、非機能要件の重要度に関するユーザ・ベンダ企業へのアンケート結果が含まれている。提案法では、このアンケート結果を、評価対象とする非機能要件の選定に利用する。

多種多様な非機能要件間の関係を明らかにする研究も行われている。日本情報システム・ユーザー協会 (JUAS) による「ソフトウェア開発管理基準に関する調査報告書」[JUAS2]では、品質目標 (SLA指標)、運用容易性、障害対策、災害対策といった観点で、非機能要件が整理されている。また、情報処理推進機構ソフトウェア・エンジニアリング・センター (IPA-SEC) による「共通フレーム2007」[IPA2007]では、運用と保守のプロセスに関する非機能要件の整理がなされている。提案法では、これら2つの成果に基づき、評価対象とする非機能要件55個を3階層でグループ化している。

2.2. メトリクス

IPA-SECによる「非機能要求グレード」[IPA2010]は、情報システムにおけるセキュリティや性能、業務の手順など、機能以外に関する要件 (非機能要件) を定義すると共に、要件に対する要求レベルを評価し、ユーザ・ベンダ間で合意を形成するための枠組みを与えるものである。要件を階層的にグループ化し、評価基準を要件毎に定義するというアプローチは、提案法と同じであるが、要求レベルの評価はベンダ視点で行われ、ユーザにとって重要な保守に関する要件などについては言及されていない。

RFPや要求仕様書など、ソフトウェア開発の初期に作成される文書の評価に、自然言語処理技術を用いる研究も報告されている。佐藤らは、要求仕様における品質



図1 提案する RFP 評価法の概要

要求の含有率を、形態素解析に基づく重要語句の抽出などにより測定する具体的な方法とツールを提案している [Sato]。評価対象には非機能要件も含まれているが、評価の粒度は、「セキュリティ」、「成熟性」、「運用性」などであり、提案法に比べると大きい。

3. 提案法

3.1. 概要

提案法は、ソフトウェア開発に向けて作成される提案依頼書 (Request For Proposal: RFP) の品質を定量的に評価するものである。品質評価の観点は、「運用と保守に関する非機能要件」に関する記述の有無、および、明確さである。評価結果は、RFP の総合評価点 (100 点満点)、および、要件毎の評価点を俯瞰するためのレーダーチャートとして示される (図 1 参照)。

提案法の主な利用者は、RFP 作成者 (ソフトウェア開発をベンダに依頼するユーザ) である。RFP 作成者は、ベンダに対する RFP の提示に先立ち、非機能要件に関する記述の明確さを提案法により定量的・視覚的に把握する。明確に記述されていない要件があれば、必要な加筆修正を RFP に対して行う。

3.2. 評価対象とする非機能要件

評価対象とするのは、2.1 で示した 3 つのガイドライン [JUAS][Nikkei][TRM] で示されている非機能要件のうち、保守と運用に関する 55 個の非機能要件である。これは、本提案法の主な利用者となる委託元企業 (ユーザ) が、ソフトウェアと最も直接的に関わるのが「保守

と運用」であり、それら要件をベンダに正確に伝えることが RFP 作成の主要な目的のひとつと考えられるからである。また、非機能要件は、セキュリティ対策、冗長化、応答時間といったアーキテクチャの制約条件となる場合が多く、アーキテクチャの実現可能性を評価する上でも役立つ。これとは反対に、ベンダによるソフトウェア開発管理に関する要件、ユーザが自身のために行う開発管理に関する要件 (ベンダに伝える必要性の低い要件) は、評価対象とはしていない。

55 個の要件のうち 34 個は運用に関する要件、21 個は保守に関する要件である。また、55 個の要件のうち 17 個は、サービスレベルの合意に必要な要件である。残る 38 個は、文献 [MEXT] で実施されたアンケートにおいて、3 分の 1 以上のユーザ企業が、「RFP に実際に記述している」あるいは「記述すべき」と回答した要件である。

3.3. 非機能要件評価シート

非機能要件評価シートは、評価対象とする 55 個の非機能要件それぞれについて、「評価メトリクス (明確さの評価基準スキーム)」と「重要度 (評価における重み)」を与えるものである (図 1 参照)。なお、評価対象とする要件が 55 個と多数にのぼるため、評価結果の俯瞰が難しくなる可能性がある。そこで、類似する要件をグループ化し、17 個の「中項目」として設定し、更には、それら中項目を、ソフトウェア利用者の観点で設定した 7 個の「大項目」に対応付けている。

評価対象とする要件それぞれについての記述内容は次の通りである。

■非機能要件 i

名称：

定義：

■評価メトリクス (評価点 s_i)

明確さ 4 の評価基準

3 の評価基準

2 の評価基準

1 の評価基準

0 の評価基準

■重要度 w_i

提案法では、各要件は最大 5 段階で評価される。評価点の取りうる値は、0 から 4 の整数値である。「明確さ評価基準」は、文字通り、当該要件の明確さを評価するための基準を示すものである。当該要件が（十分に）明確に記述されている場合の評価点は 4、記述がない、もしくは、記述の明確さが著しく低い場合は 0 となる。ただし、要件によっては、記述の明確さに区別はなく記述の有無だけで評価できる要件、記述の明確さについての議論や検討が（現時点では）十分ではなく 5 段階評価が難しい要件、などがある。そうした要件については、明確さ 3 の評価基準、同 2 の評価基準、同 1 の評価基準のいずれか、もしくは、全てを「該当なし (N/A)」とできるものとする。例として、いくつかの非機能要件とその明確さ評価基準を図 2 に示す。図 2 (a) に示す非機能要件「バックアップ方式」では、5 つ全ての評価基準が示されており、5 段階評価が行われる。図 2 (b) に示す非機能要件「システムソフト」では、明確さ 3 と 1 の評価基準評価が「該当なし (N/A)」となっており、3 段階評価となる。図 2 (c) に示す非機能要件「応答時間」では、明確さ 3 から 1 の評価基準が全て「該当なし (N/A)」となっており、2 段階評価となる。

「重要度」は、RFP における当該要件の重要度を相対的に示す数値である。前述の通り、要件の明確さの評価点を取り得る値は、全ての要件において、0 から 4 の整数値である。そこで、RFP の総合評価点 (100 点満点) の算出において、複数の要件の評価点を加算するにあたって、この重要度を重みとして用いる。要件の重要度は、対象ソフトウェアのドメインや利用組織毎に異なり、一律に定めることは出来ない。本論文では、一例として、文献 [MEXT] で実施されたアンケートにおいて、「重要な要件であり、RFP に実際に記述している」あるいは「記述すべき」と回答したユーザ企業数に基づき重要度を決定した。例えば、「バッチ処理正常終了率」の重要度は「オンラインシステム稼働率」の重要度の 6.2 倍となっているが、これは、同アンケートにおいて、上記のように回答したユーザ企業数が 6.2 倍あったことを意味する。同アンケートの対象外の要件については、システム発注・

非機能要件 No.27	
名称	バックアップ方式
定義	データ及びハードウェアに関するバックアップ仕様
評価メトリクス	
明確さ 4 の評価基準	ハードウェア及びソフトウェアのバックアップ構成が系統的に記述されている。
明確さ 3 の評価基準	ハード及びソフトのバックアップについて記述されている。
明確さ 2 の評価基準	バックアップの記述はあるが具体的な方式の記述がない。
明確さ 1 の評価基準	バックアップ方式について提案を要求している。
明確さ 0 の評価基準	バックアップ方式についての記述がない。
重要度	3.6
カテゴリ	大項目：障害対策 中項目：冗長化

(a) 5 段階評価

非機能要件 No.41	
名称	システムソフト
定義	システムで使用する OS 及びユーティリティソフトウェア
評価メトリクス	
明確さ 4 の評価基準	使用するシステムソフトウェアの名称が具体的に記述されている。
明確さ 3 の評価基準	N/A
明確さ 2 の評価基準	使用するシステムソフトウェアの名称が具体的に記述されていない。
明確さ 1 の評価基準	N/A
明確さ 0 の評価基準	システムソフトウェアについての記述がない。
重要度	1.5
カテゴリ	大項目：保守生産性 中項目：保守容易性

(b) 3 段階評価

非機能要件 No.9	
名称	応答時間
定義	システムとしての応答時間 (画面操作時のデータ更新、通信時間など)
評価メトリクス	
明確さ 4 の評価基準	応答時間が目標時間として (数値で) 記述されている。
明確さ 3 の評価基準	N/A
明確さ 2 の評価基準	N/A
明確さ 1 の評価基準	N/A
明確さ 0 の評価基準	応答時間の目標時間が記述されていない。
重要度	1.3
カテゴリ	大項目：システム運用評価 中項目：稼働品質性能

(c) 2 段階評価

図 2 明確さ評価基準の例

開発に長年携わってきたエキスパートの意見に基づき重要度を決定した。その上で、評価対象とする 55 個の非機能要件全体で、重要度 (重み) の合計が 100 となるよう正規化を行った。その結果、重要度が最も高い要件は「バッチ処理正常終了率」で重要度は 6.2、最も低い要件は「オンラインシステム稼働率」、「アクセス監査」など 18 個の要件で重要度は 1.0 となった。

3.4. 評価結果

「非機能要件評価シート」に基づく評価結果は、RFP の「総合評価点」と要件毎の評価点を俯瞰するための「レーダーチャート」に大別される。総合評価点 S は、

評価対象とする 55 個の非機能要件それぞれに対する評価点を、その重要度で重み付けした加重和である。

$$S = \sum w_i s_i / 4 \quad (i=1, \dots, 55)$$

ここで、 s_i は、要件 i の評価点、 w_i は要件 i の重要度である。55 個の非機能要件全てが明確に記述されている場合、総合評価点 S の値は 100 となり、記述に明確さがなく、あるいは、記述そのものがないほど、要件の重要度に応じて減点されていることになる。

レーダーチャートは、要件間での評価点の比較などが容易に行える表現形式である。ただし、提案法では、評価対象とする非機能要件が 55 個と多数にのぼるため、それら全ての評価値をレーダーチャートで表現することは現実的ではない。そこで、「非機能要件評価シート」において設定した「大項目」および「中項目」を単位としてレーダーチャートを作成する（図 1 参照）。「大項目レーダーチャート」では、大項目それぞれに属する要件の評価点の平均値を示す。「中項目レーダーチャート」でも、同じく、中項目それぞれに属する要件の評価点の平均値を示す。平均値が取り得る値は、いずれも、0～4 であり、要件が明確に記述されているほど高い値となる。

4. ケーススタディ

4.1. 概要

提案法の適用容易性や有用性を評価するために行ったケーススタディの結果について述べる。ケーススタディでは、地方自治体、図書館、政府機関、大学、病院などが、ベンダ候補企業向けの入札情報として WWW 上に公開していた 29 件の RFP を評価対象とした。RFP の評価は、各 RFP の作成者ではなく、システム発注・開発に 10 年以上携わってきたエキスパート 1 名が、対象 RFP 全てに対して行った。

RFP の評価に要した時間は、RFP 1 件あたり最大 1 時間程度であった。評価者は、対象 RFP で表されるシステムやそのドメインに関する知識を十分に有していたわけではなかった。しかし、対象 RFP を熟読することで、非機能要件 55 項目それぞれの評価点を支障なく決定することが出来た。RFP 作成者自身であれば、より短い時間で評価が可能であることは容易に推察される。

また、提案方法は RFP のみに基づいて実施可能であり、対象 RFP を公開している団体や RFP 作成者に対してインタビューを行ったり、追加資料を求めたりする必要のないことも確認された。このことは、(RFP 作成者自身を含む) 複数人で RFP を評価し、デルファイ法などにより、より客観性・妥当性の高い結果を得ることが、比較的容易であることを意味する。

4.2. 総合評価点

図 3 (a) は、29 件の RFP の総合評価点の分布を、RFP が表す情報システムの 5 つのドメイン毎に示した箱ひげ図である。5 つのドメインとそれぞれの RFP 件数は次のとおりである。

地方自治体 6 件
 図書館 8 件
 政府機関 5 件
 大学 5 件
 病院 5 件

箱ひげ図は、データ分布の様相を視覚的にとらえやすく表すために工夫された図である。箱の中に引かれた横線がその分布の中央値を、箱の下辺と上辺がそれぞれ第一四分位数、第三四分位数を、更に、上下にのびたヒゲの先端が、それぞれ最大値と最小値を表す。なお、外れ値がある場合は、箱やひげとは別に、○印で表される。

図 3 (a) より、ドメインによって総合評価点に大きな違いのあることが分かる。また、総合評価点が 60 点以上となったのは、政府情報システムと病院情報システムのそれぞれで 1 件のみである。提案法では、評価対象とする 55 個の非機能要件全てが RFP において明確に記述されているべき、という立場で評価が行われている。総合評価点は、満点となる 100 点にできるだけ近いことが望まれる。しかし、大半の RFP は総合評価点が 100 点からほど遠く、非機能要件がまだまだ明確には記述されていない、ということになる。特に、図書館情報システムでは、総合評価点の中央値が 10 点未満であり、RFP に改善の余地が大きく残されていると言える。

4.3. レーダーチャート

大項目と中項目の評価結果となるレーダーチャートを図 3 (b)(c) にそれぞれ示す。同図では、5 つのドメインそれぞれにおける評価点の平均が示されている。図 3 (b) を見ると、5 つのドメイン全てにおいて大項目「運用開始の準備」の評価点が 0、「災害対策」が 0.5 以下、「システム運用の評価」が 1.0 以下と極めて低いことが分かる。評価点が 0 となった「運用開始の準備」は、図 1 に示すとおり、3 つの非機能要件「運用移行許容障害発生率」、「テスト密度」、「テストカバレッジ」で構成されている。評価点が 0 ということは、これらが全て RFP に一切記述されていなかったことになる。必要がないから記述されていなかったとも考えられるが、「非機能要件を十分に提示している」とするユーザ企業が 22.6%に過ぎないとの調査結果 [JUAS2] もあることから、ここでは、「必要だが記述されていなかった」との立場をとる。今回のケーススタディにおけるユーザは、地方自治体、政府機関、大学、病院等であり、情報システム部門を持た

ず、テストに関する知識や経験が不足していた可能性がある。その結果、テストに関連する要件が記述されず、評価点が0となったと推察する。

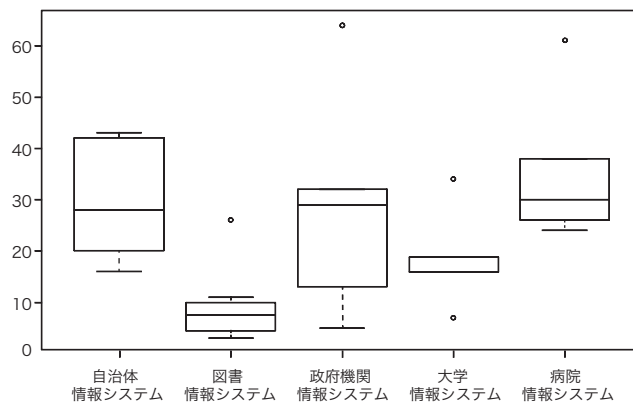
評価点が1.0以下となった「システム運用評価」は、同じく図1に示すとおり、3つの中項目「運用容易性」、「稼働率目標」、「稼働品質性能」で構成されている。図3(c)によれば、このうち、「稼働品質性能」の評価点がどの分野においても低いことが分かる。「稼働品質性能」は11個の非機能要件で構成されており、更に詳細な評価・分析が可能であるが、ここでは省略する。詳しくは、文献[Saito]を参照されたい。

4.4. ベンチマーキング

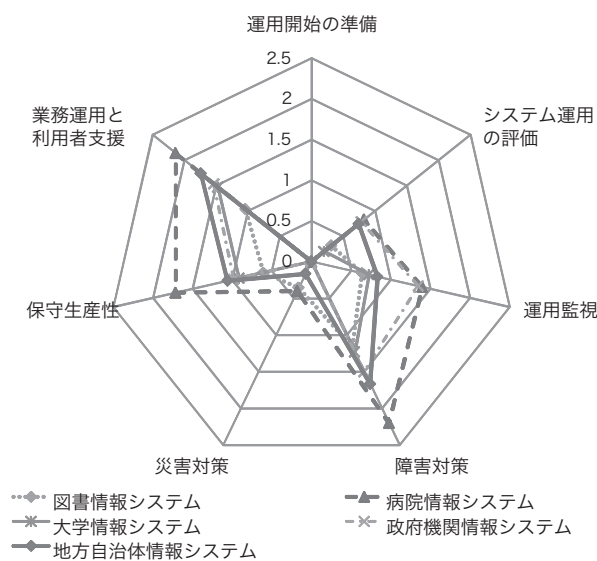
ケーススタディ結果のひとつとして、提案法におけるベンチマーキングについて述べる。先にも示した通り、提案法では、評価対象とする55個の非機能要件全てがRFPにおいて明確に記述されているべき、という立場で、いわゆる減点法により評価が行われる。RFP作成者の目標は、総合評価点が100点、レーダーチャートで示される全ての項目の評価点が4点、となるRFPを作成することと言える。

ただし、100点満点のRFPを作成することが、(現時点において)現実的であるかどうかについては議論の余地がある。提案法では、既存のガイドライン、および、RFP作成者となるユーザ企業へのアンケート結果に基づいて、評価対象となる非機能要件を選定し、記述の明確さの評価基準や重要度等を要件毎に定めている。しかし、それら要件を明確に記述することの容易性については考慮されていない。限られた工数・期間の下では、明確に記述されにくい要件が存在する可能性もある。目標としての100点満点とは別に、標準値あるいは基準値を設定し、個々のRFP評価点との比較を行うベンチマーキングも必要であると考えられる。

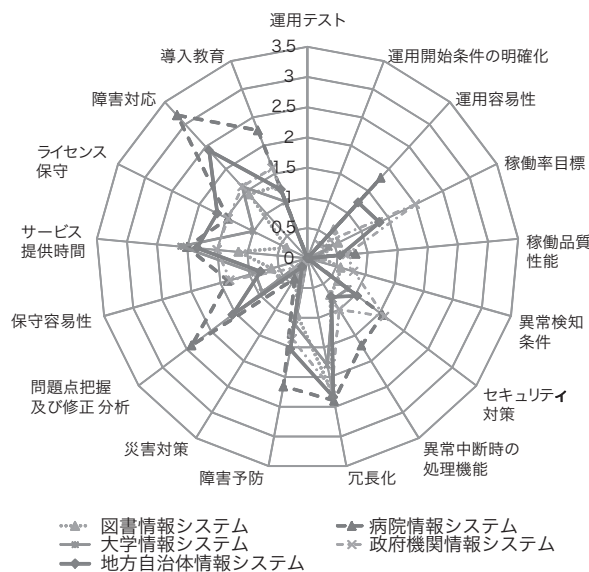
ここでは、一例として、評価対象とした29個のRFPのうち、総合評価点が高かった3個のRFP(RFPトップ3)における平均評価点を、各要件に対する評価点の基準値とした。なお、基準値の設定においては、特異点、あるいは、例外的と思われる値(評価点)は除外する必要がある。特に、著しく高い評価点は、目指すべき高い目標として基準値に組み入れるべきとされる一方で、特異点、あるいは、例外的として基準値設定から除外すべき場合もある。基準値設定に用いた3個のRFPのうち2個の総合評価点はおおよそ60点で、他のRFPに比べれば著しく高い値となっている。ただし、100点満点中の60点であり、要件によっては、他のRFPよりも平均評価点が低くなる場合もあることから、現時点では、特異点、あるいは、例外的とは見なさず基準値設定に用いた。図4は、



(a) 総合評価点



(b) レーダーチャート：大項目



(c) レーダーチャート：中項目

図3 ケーススタディ結果

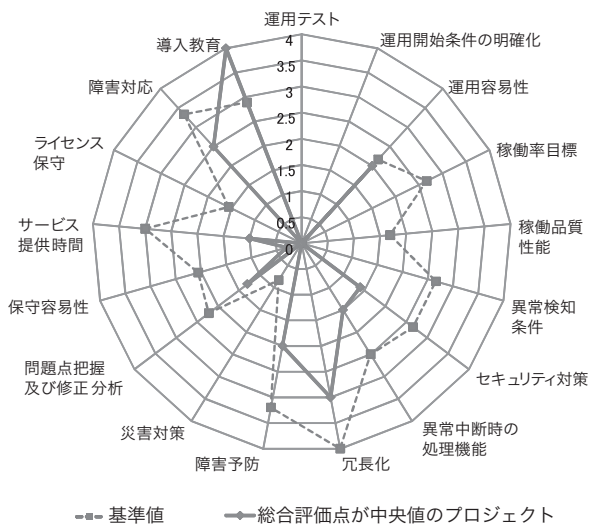


図4 ケーススタディ結果：
基準値（RFP トップ3）との比較

総合評価点が中央値であったRFP（RFP Mと呼ぶこととする）における評価点を基準値と比較した結果である。一般論で言えば、RFP Mの評価値と基準値の差が大きい要件ほど、記述の明確さに改善の余地があることになる。同図より、要件「稼働率目標」、「異常検知条件」、「サービス提供時間」などが該当する。

個別の要件について、もう少し詳しく見ていくと、例えば、要件「導入教育」の評価点は、RFP Mでは4点、基準値、すなわち、RFP トップ3の平均では2.89点となっている。評価点が満点の4点であることから、RFP Mにおいて同要件が相対的にも絶対的にも極めて明確に記述されていることが分かる。

また、要件「運用容易性」に注目してみると、RFP Mの評価点は2点、基準値も2.17点とほぼ同じである。RFP Mの評価点だけで判断すると、同要件は必ずしも明確に記述されていない、ということになる。しかし、RFP トップ3と同程度には明確に記述されており、現時点では、改善の余地はそれほどないかもしれない。一方、RFP Mにおいて、評価点が同じ2点となっている要件「障害予防」について見てみると、基準値は3.20点となっており、より明確に記述する余地が残されていることが分かる。こうした違いは、RFP Mの評価点だけを比べても分からない。他にも、要件「冗長化」について言えば、RFP Mの評価点は3点と要件「運用容易性」よりも高い評価となっているが、基準値は4点であり、要件「運用容易性」よりも既に明確に記述されているが、更に明確に記述する余地が残されていることが分かる。

なお、RFP Mにおいて評価点が0点となっているのは、要件「運用テスト」、「運用開始条件の明確化」、「稼働率目標」、「稼働品質性能」、「異常検知条件」、「災害対策」、

「ライセンス保守」の7要件である。このうち、要件「運用テスト」、「運用開始条件の明確化」については、基準値も0点となっているが、いずれもユーザ企業に対するアンケート [MEXT] において重要であるとの回答数が多い要件である。特に、高い信頼性が要求されるドメインでの委託ソフトウェア開発においては、ベンダがシステム開発の完了を確認し、ユーザが運用を開始する条件としてRFPに記述されるべきで要件ある。一方、残りの5つの要件については、より明確に記述する余地があり、RFP Mにおける記述の不明確さには、個別の原因や理由があると考えらるべきである。

4.5. 評価者間の評価点のばらつき

評価者間のばらつきを確認するため、評価者を2名追加し、エキスパートとの間で評価結果を比較する実験を行った。追加した評価者のうち1名は、ソフトウェア工学を専門とする、業務経験のない大学教員（以降、教員と記す）である。もう1名は、エンタープライズ系のソフトウェアエンジニアとして20年以上の経験を有する者（以降、エンブラ系S Eと記す）である。29件のRFPのうち、各ドメイン（地方自治体、図書館、政府機関、大学、病院）から各1件をランダムに選択し評価対象とした。

実験の結果、まず、各要件に対する評価点の評価者間での差（絶対値の平均）は、1非機能要件あたり、エキスパートと教員の間で0.367、エキスパートとエンブラ系S Eとの間で0.585となり、1未満（5段階評価における1段階未満）となった。評価点に有意差（フリードマン検定、有意水準5%）が認められたのは、病院情報システムのRFPに対するエキスパートの評価点とエンブラ系S Eの評価点のみであった。そのケースにおいて、評価点の差が特に大きかった要件は、「スループット」「最大負荷スループット」「最大停止時間」「ターンアラウンド時間」「保証期間」の5つであった。これらはいずれも、要件に関する数値情報が記述されていれば4点、されていなければ0点となる要件で、エキスパートによる評価はいずれも0点、逆に、エンブラ系S Eによる評価はいずれも4点であった。実際には、これら5つの要件に関する数値情報はRFPには記述されておらず、エンブラ系S Eによる評価は妥当でないことがわかった。エンブラ系S Eに追加インタビューしたところ、「数値情報は示されていなかったが、要件に関する記述は見られたので4点と評価した。数値情報の有無を厳密に評価に反映しなかったのは少し寛大なのでは、と指摘されてもいたしかたない。」との回答が得られた。このことから、数値情報の有無が評価に直結する要件については、そのことを評価者に徹底することが必要であり、また、徹底することで、評価者間で評価のばらつきを小さく抑えること

が期待される。

次に、総合評価点（100点満点、重み付き）については、表1に示す結果となった。エンブラ系SEによる病院情報システムに対する評価点を除くと、教員およびエンブラ系SEによる評価点とエキスパートによる評価点との差は、最大でも6.09に留まった。

表1. 各評価者の各RFPに対する総合評価点

評価者 ドメイン	エキスパート	教員	エンブラ系SE
地方自治体	2.71	5.99	6.88
図書館	27.31	32.79	33.40
政府機関	18.91	15.16	16.91
大学	5.06	5.88	5.34
病院	42.75	44.08	64.90

以上より、実務経験のない大学教員であってもエキスパートと有意差のない評価を行えること、また、数値情報の記述が求められる非機能要件については、具体的な数値が記述されていないければ評価点は0とすべきことを徹底することで、評価のばらつきを抑えられる可能性があることが分かった。本結果の信頼性を増すため、より多くの評価者を被験者として評価実験を行うことが今後の課題となる。

5. まとめ

本論文では、ベンダへの提案依頼書（RFP）提示に先立ち、RFP作成者であるユーザ自身が、RFPの品質を定量的に評価する方法を提案した。評価対象は、ユーザにとって重要度の高い「保守と運用に関する55個の非機能要件（NFR）」であり、評価の観点、その記述の明確さ、である。記述の明確さは、最大5段階で評価され、その結果は、RFPの「総合評価点」と要件毎の評価点を俯瞰するための「レーダーチャート」として示される。地方自治体、図書館、政府機関、大学、病院などがWWW上に公開していた6ドメイン29件のRFPを評価対象としたケーススタディによって、記述が不十分な要件を特定したり、基準値との比較を通じて特に改善が必要な特性を明らかにしたりできることなどが確認された。加えて、ドメインや要件によって評価点やそのばらつきに比較的大きな差があることが、総合評価点の比較やレーダーチャートによる俯瞰により明確となり、提案法に基づくRFPベンチマーキングの可能性についても議論を行った。なお、評価はRFPのみに基づいて実施可能であり、評価に必要な時間も、RFP1件あたり最大1時間程度であった。

提案法は、RFPを対象としたものであり、ベンダへの提示に先だってユーザのみが利用するものと位置づけられている。ただし、RFPに基づいて作成される技術仕様・

技術提案書や契約仕様書へと適用範囲を拡げることは比較的容易である。その場合、技術仕様・技術提案書の作成においてベンダが提案法を利用する、また、契約仕様書の作成に向けた技術協議において、ユーザとベンダの双方が提案法を利用し、非機能要件に関する合意形成を効率よく行う、といったことも考えられる。

また、関連研究においても少し紹介したが、ソフトウェア開発で作成される文書の評価に、自然言語処理技術を用いる研究が盛んに行われている。提案法においても、例えば、非機能要件記述に含まれる典型的な語句や表現を自然言語処理技術で抽出し、非機能要件の文例集を作成することが考えられる。文例集があれば、RFPや対象ドメインに関する知識が十分でない者でも、提案法による評価が可能に、あるいは、より容易になる。評価者による評価結果のばらつきが減れば、評価法に基づくRFPベンチマーキングの信頼性や有用性も高まる。テキストマイニングや機械学習といった技術と組み合わせることで、RFP評価の自動化にも道を開くことになる。

謝辞

本研究は、独立行政法人情報処理推進機構 技術本部 ソフトウェア高信頼化センター（SEC: Software Reliability Enhancement Center）が実施した「2012年度ソフトウェア工学分野の先導的研究支援事業」の支援を受けたものです。

【参考文献】

- [IPA2007] 情報処理推進機構ソフトウェア・エンジニアリング・センター：“共通フレーム2007”，オーム社（2007）。
- [IPA2010] 情報処理推進機構 ソフトウェア・エンジニアリング・センター：“非機能要求の見える化と確認の手段を実現する「非機能要件グレード」”（2010）。
- [JUAS] 日本情報システム・ユーザー協会編：“非機能要求仕様定義ガイドライン”（2008）。
- [JUAS2] 日本情報システム・ユーザー協会：“ソフトウェア開発管理基準に関する調査報告書（ソフトウェアメトリクス調査）”（2012）。
- [Kazman] Rick Kazman, Mark Klein, Mario Barbacci, Tom Longstaff, Howard Lipson, Jeromy Carriere: “The Architecture Tradeoff Analysis method,” Technical Report, CMU/SEI-98-TR-008, ESC-TR-98-008, Carnegie Mellon University, Software Engineering Institute (1998)。
- [MEXT] 経済産業省ソフトウェアメトリクス高度化プロジェクトプロダクト品質メトリクスWG：“システム/ソフトウェア製品の品質要求定義と品質評価のためのメトリクスに関する調査報告書”（2011）。
- [Nikkei] 日経ソリューションビジネス編：“システム構築のトラブルを回避するためのITシステム契約締結の手順とポイント”，日経BP社（2008）。
- [Roth] Bud Porter-Roth（著），渡部洋子（訳）：“RFP入門—初めての提案依頼書”，日経BP（2004）。
- [Saito] 齊藤康廣，門田暁人，松本健一：“ソフトウェア委託開発プロジェクトの超上流工程における非機能要件評価に関する研究”，奈良先端科学技術大学院大学テクニカルレポート，NAIST-IS-TR2013001（2013）。
- [Sato] 佐藤知徳，鈴木俊一，北澤直幸，長田晃，海谷治彦，海尻賢二：“ソフトウェア要求仕様における品質要求の含有率測定ツールの設計”，電子情報通信学会技術研究報告（知能ソフトウェア工学KBSE2007-57），Vol.107, No.540, pp.19-24（2008）。
- [TRM] 経済産業省 商務情報政策局 情報処理振興課，情報処理推進機構：“情報システム調達のための技術参照モデル（TRM）平成22年度版（2011）。