

OSS プロジェクトにおけるユーザ用メーリングリストの実態調査

伊原 彰紀^{1,a)} 松本 健一¹

概要:近年、ソフトウェア関連ビジネスで、オープンソースソフトウェア (OSS) を商用ソフトウェアの一部として利用している。そのため、多くの OSS プロジェクトでは、ユーザが求めるテクニカルサポートを支援するためにユーザ用メーリングリストが準備されている。しかしながら、独立行政法人情報処理推進機構 (IPA) の調査では、OSS プロジェクトからテクニカルサポートを得られていないことを OSS 導入時の最も大きな課題として挙げている。本論文では、大規模 OSS プロジェクトのユーザ用メーリングリストを対象に、ユーザがテクニカルサポートを得られているかを返信数と返信率から分析した。分析の結果、ユーザによってユーザ用メーリングリストにたてられたスレッドの中で返信があったのは 56.5%であり、月曜に立てられたスレッドの返信率が 59.3%に対して、金曜にたてられたスレッドは 54.3%に低下することが分かった。

1. はじめに

今日、多くのソフトウェア関連ビジネスはオープンソースソフトウェアに依存している。例えば、Android OS は今や iOS と並んでスマートフォンの主要な OS となっており、携帯端末の開発企業のみならず、アプリケーション開発企業や個人開発者も Android 向けアプリの開発に取り組んでいる。

OSS を利用する企業は、開発期間の短期化、開発コストの削減、利用できるソフトウェアの多様化を期待しているが、OSS の特殊な開発形態による課題が多く、OSS の導入に不安を抱える企業も多い。独立行政法人情報処理推進機構 (IPA) による第 3 回オープンソースソフトウェア活用ビジネス実態調査 (有効回答数 700 社以上、2009 年度の調査結果) [1] によると、企業が考える OSS のデメリットとして「緊急時の技術的サポートが得にくいこと」を挙げている。現在、社内に OSS のシステムを管理できる人材を確保できていない場合が多く、ソフトウェア開発企業は OSS に関するテクニカルサポートを OSS プロジェクトが提供する FAQ やユーザ用メーリングリストに頼ることになる。

本稿では、企業にとってテクニカルサポートの窓口のように利用されているユーザ用メーリングリストの実態調査を行い、ユーザが容易にサポートを受けるための方法を検討する。ユーザがテクニカルサポートを受けるシステムに

は、メーリングリストの他に OSS プロジェクトが提供する不具合追跡システムがあるが、多くの OSS プロジェクトは不具合追跡システムを議論の場として使用することを避けており、不具合追跡システムに報告する前に一度メーリングリストで既知の課題であるか確認することを推奨している。従って、利用者がテクニカルサポートを受けるには、メーリングリストが主流となっている [3]。

OSS プロジェクトのユーザ用メーリングリストでは、開発者がユーザから直接意見を得られる場であるため開発者も散開しており、OSS プロジェクトが新たな機能拡張のアイデアを得る上で重要なコミュニティを形成している [12]。ユーザ用メーリングリストにスレッドをたてた (メッセージを送った) 場合、返信するのは開発者だけでなく、その他のユーザが返信することもある。ユーザの中には豊富な知識を有する者も参加しており、ユーザ用メーリングリストはユーザ間で教え合うこともある。本稿では、送信者の特徴、及び、送信時期に基づき、スレッドをたてた送信者が開発者や専門知識を持つユーザからテクニカルサポートを受けているか (返信をえられているか) を明らかにする。

続く 2 章では関連研究を述べ本研究の立場を明らかにする。3 章では、本稿で扱う OSS プロジェクトが提供するユーザ用メーリングリストについて説明する。そして、4 章において大規模 OSS プロジェクトとして Apache プロジェクトが提供するユーザ用メーリングリストの使用状況を分析した結果と利用者がテクニカルサポートを受けるための方法について検討する。最後に 5 章においてまとめと

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Japan
^{a)} akinori-i@is.naist.jp

今後の課題を示す。

2. 関連研究

2.1 メールを対象としたコミュニケーション分析

メールは、人々が意思疎通を図るための主要なコミュニケーションツールとなっている。特に企業では、日々受信する多くのメッセージから重要なメッセージを取捨選択することが業務の効率化に向けて重要であるとされている [5][7]。メールに関する研究は 1988 年初旬から行われており、数々のインタビュー、分析を通して、メッセージ内容の重要性を分類する技術が提案されている [4][6]。例えばスパムメールの排除、また、優先度の高いメッセージ(返信を要するメッセージ)の推薦ために機械学習技術などが使用されている。日々受信するメッセージにはアナウンスやリマインダーなど、返信する必要のないメッセージが多く含まれているが、OSS プロジェクトが管理するユーザ用メーリングリストは開発者からのアナウンスメールを除く、ユーザからのメッセージは質問や不具合報告など返信を要するものが多い。本稿では、ユーザ用メーリングリストにたてたスレッドが開発者や知識を持つユーザから対応される方法を確立することを目的としている。

2.2 OSS 開発を対象としたコミュニケーション分析

OSS 開発における協調作業を理解するために、OSS プロジェクトの参加者がメーリングリストで行ったコミュニケーションをネットワークモデル化し、そのネットワークモデルの構造について研究が行われている [9][10]。Howison らは OSS プロジェクトの開発者同士がメーリングリストのやりとりからメールの送受信関係をネットワークで表現し、そこから開発者間の社会構造を分析している [9]。また、まつ本らは、OSS プロジェクトの開発者用メーリングリストとユーザ用メーリングリストの送受信関係をネットワークモデル化し、分析を行っている。分析の結果、活動が活発な OSS プロジェクトでは、両方のメーリングリストに参加している人物が、2つのメーリングリストに対して偏りなく交流する傾向があることを確認している。多くの従来研究は、開発者間のコミュニケーションを対象とし、特に、プロジェクトの中心人物を評価する研究が多く見られる。その一方で、利用者を支援することを目的とした研究が少なく、OSS において参加者から返信のないメールを対象とする研究はない。

2.3 コミュニケーションの品質分析

OSS 開発において、開発者とユーザとのコミュニケーションの不一致について研究が行われている。OSS の不具合は誰でもプロジェクトに報告することができ、不具合票に修正するための十分な情報が含まれていないことが多いため修正に長い時間がかかっている。Bettenburg ら

は、不具合票の質を向上するために、開発者と報告者を対象としたアンケート調査で、両者が不具合を修正するために有用であると考えた記述に差が生じていることを挙げて [14]。開発者は不具合を修正するために再現方法やテストケースの記述が有用であると答えているが、その一方で、報告者は不具合が発見されたプロダクトの種類や OS の種類が修正するために有用であると考え、両者の考えが一致していないことが分かった。本課題はユーザがたてるスレッドにも同様のことが言えると考えられ、ユーザの質問に回答するための情報がメッセージに含まれていないため、サポートが得られていないと考える。

3. OSS 開発におけるメーリングリスト

ユーザが OSS のサービス内容、サポート状況などを理解する場合、Web ページの次に参照すべきコンテンツはメーリングリストである [3]。メーリングリストはプロジェクトにおける議論の場となっており、大規模なプロジェクトでは、目的に応じて複数のメーリングリストを管理している。例えば、開発者用メーリングリストは、開発に関する話題を取り扱うメーリングリストであり、不具合修正や新規機能の追加についての議論が主として行われる。ユーザ用メーリングリストでは、ユーザが利用するメーリングリストであり、主として導入時のトラブルやバージョン問題など、開発者を交えて議論が行われる。その他に、アナウンス用メーリングリスト、パッチ投稿用メーリングリストなどがある。

OSS プロジェクトが提供するユーザ用メーリングリストでは、返信者は開発者だけでなく、ユーザが返信する場合もある。ユーザの中には OSS に関する豊富な知識を持つ者も参加しており、ユーザ間で助け合い、直面した課題の解決方法を教え合うということがある。しかしながら、IPA の調査 [1] から、ユーザがサポートを十分に受けられていないというのも事実であり、ユーザ用メーリングリストが十分に機能していないことが示唆される。その理由を3つ観点に分類する。

送信時間 OSS プロジェクトの開発者、及び、利用者は世界各地に点在するため、コミュニケーションのタイムラグが発生する可能性が高い。また、OSS の開発、利用を昼または夜に行う者、平日または祝日に行う者がいるため、送信時間によってコミュニケーションのタイムラグが発生する可能性が高いと考えられる。

送信者の特徴 ユーザ用メーリングリストの中には長期間 OSS を利用している者、OSS の開発に携わっている者が参加している。このような参加者は OSS を理解した上で質問を明確なメッセージで送ることができるかもしれない。また、OSS プロジェクトへの関わり(例えばメーリングリストの利用頻度)が多いユーザは、専門性の近いユーザ同士でコミュニティを形成し、コ

コミュニティのメンバ間で返信している可能性がある。一方で、OSS プロジェクトへの関わりが少ないユーザは、OSS 活動の経験が少ないため、質問を明確に相手に伝えることが難しいかもしれない。このような場合、開発者が返信しない可能性があると考える。

メッセージ内容 メーリングリストには日々多くのメッセージが送られるため、メーリングリスト読者は全てのメッセージを読んでいくわけではない。そのため、メッセージの内容（OSS プロジェクトで使用される専門用語の有無、文章能力など）で返信するか否かを判断している可能性がある。

本稿では、送信時間と送信者の特徴による返信数と返信率を分析した結果を報告し、メッセージ内容については今後の課題とする。

4. ユーザ用メーリングリスト使用状況

本稿では、Apache プロジェクトにおけるユーザ用メーリングリストを対象に分析を行う。本章では、ユーザ用メーリングリストの基本統計量を示した上で、実際にユーザ用メーリングリストでユーザがサポートを受けているのか否かを送信時間、及び、送信者の特徴に基づいて分析を行う。本稿では、送信者の文才能力のような個人の技量に依存する特徴やメッセージの内容のような定性的な分析は今後の課題とし、送信時間や開発者の経験数（利用者の送信回数）のような個人の技量に依存しない定量的な特徴に着目する。また本稿において、ユーザがユーザ用メーリングリストでサポートを受けているか否かを判定するために、返信の有無を計測する。返信の有無は、ユーザがたてたスレッドに対して、当該ユーザ以外の者が返信をした場合にサポートを受けたと考える。返信をする者は開発者に限らず、ユーザも含まれる。ユーザの中には OSS に関する豊富な知識を有する者がおり、ユーザ間の議論で問題が解決する場合もあるため、本稿では、ユーザが返信した場合もサポートを受けたと考える。

4.1 実験対象データ

本稿で実験対象とする Apache プロジェクトは、2013 年現在、世界一のシェア（Web サーバ市場の半数以上）を誇っている Web サーバ（HTTP Server）ソフトウェアである。パフォーマンスの高さや拡張性の高さに定評があり、欠陥の修正パッチを継続的に多数リリースしているため、非常に高品質なソフトウェアとして広く認知されている。[11]. Apache プロジェクトは、開発者メーリングリスト、ユーザ用メーリングリスト、ドキュメント作成を議論するためのメーリングリストなど、複数のメーリングリストを管理している。ユーザ用メーリングリストは 2001 年 11 月から運用が開始されており、ユーザ用メーリングリストを紹介している Web ページでは、メーリングリストに質

表 1 実験対象データ（Apache プロジェクト）

分析期間	2001 年 11 月 - 2013 年 2 月
メッセージ数	106,059 件
スレッド数	32,081 件
- 開発者によるスレッド数	2,124 件
- 利用者によるスレッド数	29,957 件
- 返信ありスレッド数	16,936 件
- 返信なしスレッド数	13,019 件
利用者数（開発者を除く）	13,931 人

問を投稿するための方法を掲載している*1。

表 1 は、ユーザ用メーリングリストの利用状況である。ユーザ用メーリングリストには 32,081 件のスレッドが立てられている。その中で、開発者がスレッドをたてているのは 2,124 件（6.6%）ユーザがスレッドをたてているのは 29,957 件（93.4%）である。ユーザによってたてられたスレッドのうち、開発者、もしくは、その他の利用者からの返信があったのは 16,936 件（56.5%）であった。約半数のメッセージには返信がないため、OSS の利用者からデメリットとして「緊急時の技術的サポートが得にくいこと」が挙げられたと示唆される。

本稿では、メッセージが返信されない理由を、送信時間や送信者の特徴（送信経験数）に基づき定量的に分析する。

4.2 時間的特徴と返信率の関係

OSS 開発者の多くはボランティア開発者であるため、一般的な仕事終了時刻前後である夕方から深夜にかけて送信数が増加する [15]。しかしながら、これまでユーザを対象に分析された研究はなく、ユーザはいつ活動し、また、送信したメッセージに対してサポートが得られているか否かを分析された研究は無い。本稿では、ユーザがユーザ用メーリングリストにたてたスレッドに対して、返信の有無を分析する。

図 1 は、送信曜日別の返信数（上図）と返信率（下図）である。両図では、横軸が曜日となっている。縦軸はそれぞれ返信数と返信率である。週末よりも平日にメッセージが送信されていることが多いことが分かった。Web サーバソフトウェアとして利用される Apache は多くの企業で利用されており、業務中のトラブルや課題についてメッセージを送信していることが示唆される。一方、返信率は週末（土曜、日曜）が平日よりも返信率が高く、平日は週末に向けて返信率が低下している。金曜の返信率は、土曜の返信率に比べて約 5%低い。本分析では数千のスレッドを対象としているため、土曜よりも金曜の方が返信されたスレッドが数十件少ない。金曜に送信したメッセージが土曜に比べて返信されない理由の一つは、当該スレッドを確認するのが次の営業日（月曜）が多いと考えられ、メーリングリ

*1 How To Ask Questions The Smart Way:
<http://www.catb.org/esr/faqs/smart-questions.html>

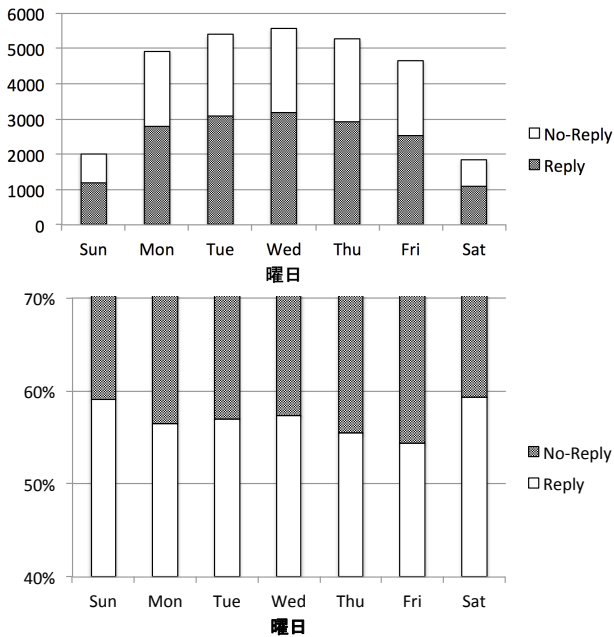


図 1 送信曜日帯別の返信数と返信率

ストに埋もれて気づかれないことが多いと考えられる。

図 2 は、送信時間帯別の返信数（上図）と返信率（下図）である。両図では、横軸が送信時間帯となっている。縦軸はそれぞれ返信数と返信率である。従来研究で開発者用メーリングリストでは夕方から深夜にかけて送信数が増加していた一方、ユーザ用メーリングリストでは、日中に送信する機会が多く、図 1 と同様に、企業が日中の業務で Apache の課題に直面し、ユーザ用メーリングリストで問い合わせていることが示唆される。返信率は、7 時前後と 23 時前後の返信率が高く、返信率の低い午前 4 時頃に比べて約 10% 高い。早朝に送信されたメールは業務開始時間に多くの利用者、及び、開発者が午前中に確認し、返信される可能性が高いと考えられる。一方、23 時前後のメッセージは、深夜にかけて活動する開発者からの返信があったのではないかと考えられる。

図 3 は、送信時期別の返信数（上図）と返信率（下図）である。両図では、横軸が送信年月となっている。縦軸はそれぞれ返信数と返信率である。ユーザ用メーリングリストの利用は、Apache プロジェクト発足当初から年々減少している。また、返信率も年々低下していることから、ユーザがサポートを得ることが困難になっており、ユーザ用メーリングリストの利用が減っていることが考えられる。

4.3 送信者の特徴と返信率の関係

OSS プロジェクトの参加者の多くは一時的な参加であり、長期間参加する者は少ない [16]。OSS プロジェクトのように不特定多数の者が参加するコミュニティでは、積極的に活動する者、一時的に活動してプロジェクトを離脱する者が存在する。送信経験が多いほど、返信されるメッ

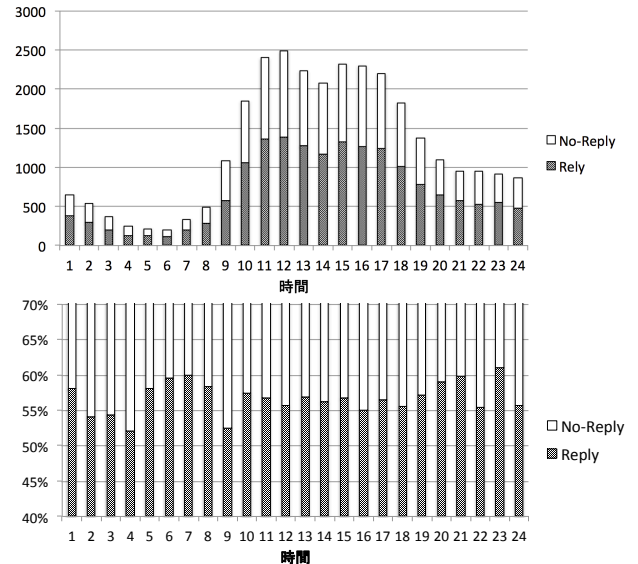


図 2 送信時間帯別の返信数と返信率

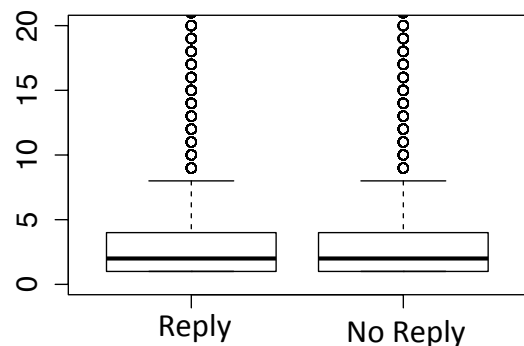


図 4 返信される送信者と返信されない送信者の経験数

ッセージを送っていると考えられる。本稿では、プロジェクトにおける活動量と返信率の関係、送信者と返信率の関係を分析する。

まず、活動量と返信率の関係を分析する。図 4 は、返信される/返信されないスレッドをたてた送信者の経験数の分布を示す箱ひげ図である。経験数とは、ユーザが過去にたてたスレッド数である。両箱の中央の線が中央値を示しており、共に経験数が 2 回であった。また、30 回以上投稿している送信者のスレッドのうち、49.0%(258/566) は返信があり、たとえ初めての投稿（経験数が 0）の送信者がスレッドをたてた場合でも、56.9%(6042 件/7964 件) のスレッドに返信がある。従って、返信されるか否かは送信者の経験数に依存しないことが分かった。

次に送信者と返信率の関係を分析する。図 5 は、送信者別の返信率である。横軸は送信者で送信数が多い順に並んでおり、縦軸は返信率である。返信率はスレッドをたてた回数に依存していない。送信回数が多いにもかかわらず返信率が高いユーザもいれば、返信率が低いユーザも存在す

t

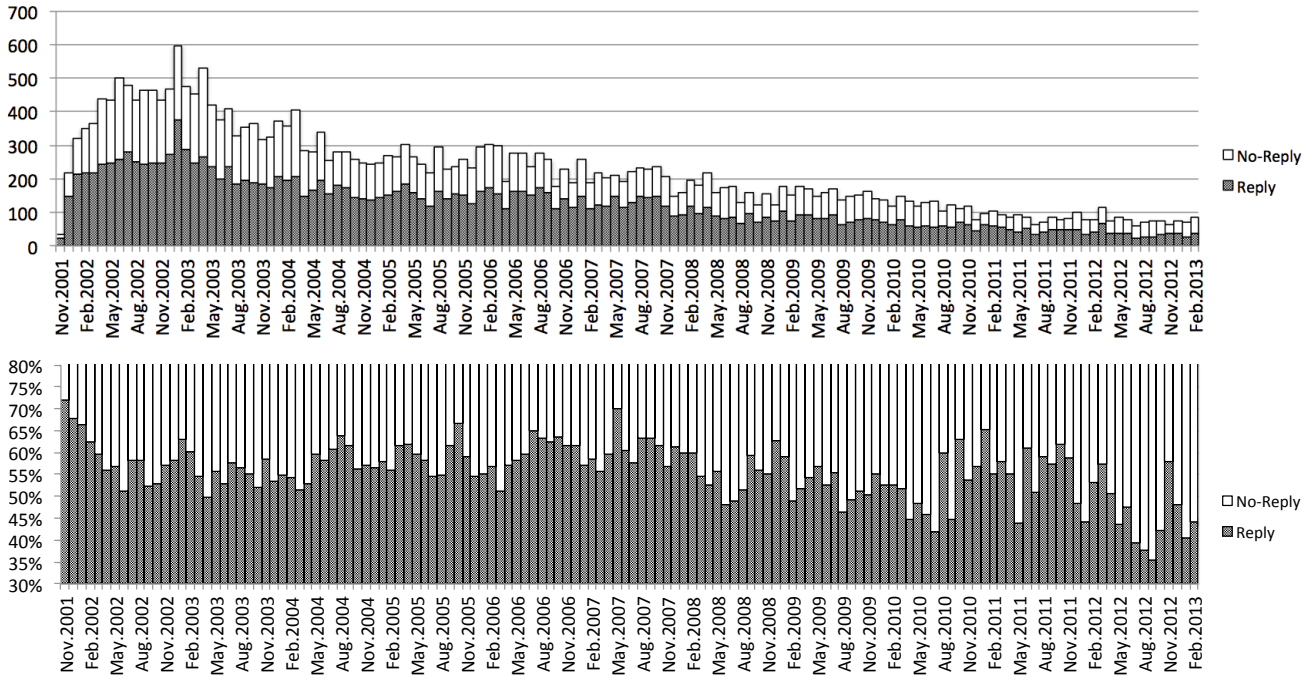


図 3 送信時期別の返信数と返信率

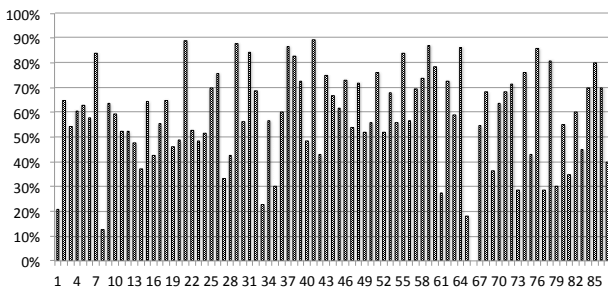


図 5 送信者別の返信率

る。従って、返信される可能性は利用者に依存していることが考えられる。本結論は送信者の文章能力に依存している可能性もあるが、個人の定性的な特徴については今後の課題とする。

5. おわりに

本稿では OSS プロジェクトのユーザ用メーリングリストの実態調査を通して、利用者がテクニカルサポートを得られているか否かを分析した。Apache プロジェクトの分析から得られた知見を以下に示す。

- 平日にたてられたスレッドは金曜日に近づくにつれて返信率が低くなり、週末（土曜、日曜）たてられたスレッドは平日よりも返信率が高い。
- 午前 7 時前後、及び、午後 23 時前後にたてられたメッセージは返信率が高い。
- スレッドをたてた経験数は返信率に依存せず、スレッドをたてた利用者によって返信率が異なる。

スレッドの送信時間による返信数と返信率の分析結果から、ユーザが返信される可能性の高い送信日が明らかとなった。しかしながら、返信されるまでの時間、返信内容がスレッドをたてた送信者を満足させるものであったか否かは明らかにできていない。また、スレッドの送信者の特徴による返信数と返信率の分析結果から、スレッドをたてた経験数は返信率に依存せず、スレッドをたてたユーザによって返信率が異なることが明らかとなった。返信されることは、ユーザ用メーリングリストを利用するユーザの満足度を上げる重要な要因の一つであるが、ユーザの要求が満たされる回答が得られていることが重要である。従って、今後は以下の 2 点を課題とする。

- 送信メッセージの書き方、送信メッセージの内容による返信率の分析
- 返信時間と返信内容の品質の分析

以上の課題を分析し、さらに他の OSS プロジェクトと比較することで、効果的な議論に発展するメッセージ交換方法を確立したい。

謝辞 本研究の一部は、文部科学省科学研究補助費（若手 B：課題番号 25730045）による助成を受けた。

参考文献

- [1] 独立行政法人情報処理推進機構.: 第 3 回オープンソースソフトウェア活用ビジネス実態調査, 2010.
- [2] Laura A. Dabbish, Robert E. Kraut, Susan Fussell, and Sara Kiesler.: *Understanding Email Use: Predicting Action on a Message*, In Proceedings of the SIGCHI

- Conference on Human Factors in Computing Systems (CHI'05). pp.691–700, 2005.
- [3] Karl Fogel.: *Producing open source software: how to run successful free software project.*, O'Reilly Media, Sebastopol, CA, 2005.
- [4] Mary Summer.: *The Impact of Electronic mail on Managerial and Organizational Communications.*, In Proceedings of the conference on Office information systems (COCS'88), pp. 96–109, 1988.
- [5] Gary Boone.: *Concept features in Re:Agent, an intelligent Email agent.*, In Proceedings of the 2nd international conference on Autonomous agents (AGENTS'98), pp. 141–148, 1998.
- [6] Mehran Sahami, Susan Dumais, David Heckerman, Eric Horvitz.: *A Bayesian Approach to Filtering Junk Email*, In Proceedings of the AAAI workshop on Learning for Text Categorization, 1998.
- [7] Laura Dabbish, Gina Venolia, JJ Cadiz.: *Marked for deletion: an analysis of email data.*, In Proceedings of the Conference on Extended Abstracts on Human Factors in Computing Systems (CHI EA'03). pp. 924-925, 2003.
- [8] Liaquat Hossain, Andr? Wu, Kenneth K S Chung.: *Actor centrality correlates to project based coordination*, In Proceedings of the 20th anniversary conference on Computer supported cooperative work (CSCW'06), pp. 363–372, 2006.
- [9] James Howison, Kevin Crowston, Keisuke Inoue.: *Social Dynamics of Free and Open Source Team Communications*, In Proceedings of the 2nd International Conference on Open Source Systems (OSS' 06), pp.319-330, 2006.
- [10] まつ本真佑, 亀井靖高, 大平雅雄, 松本健一.: *OSS コミュニティにおけるオープンコラボレーションの理解* 情報社会学会誌, volume 3, number 2, pages 29-42, 2009.
- [11] Netcraft Ltd.: *Netcraft Web Server Survey*, available from <http://www.netcraft.com/Survey/i> (accessed 2008-08-29).
- [12] Eric S. Raymond.: *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly and Associates, 1999.
- [13] Jacek Sliwinski, Thomas Zimmermann, Andreas Zeller.: *When do changes induce fixes?*, In Proceedings of the international workshop on Mining software repositories (MSR'05), pp. 1–5, 2005.
- [14] Nicolas Bettenburg, Sascha Just, Adrian Schr?ter, Cathrin Weiss, Rahul Premraj, and Thomas Zimmermann.: *What makes a good bug report?*, In Proceedings of the 16th International Symposium on Foundations of software engineering (FSE'08), pp.308–318, 2008.
- [15] 亀井靖高, 大平雅雄, 伊原彰紀, 小山貴和子, まつ本真佑, 松本健一, 鵜林尚靖.: *グローバル環境下における OSS 開発者の情報交換に対する時差の影響*, 情報社会学会誌, volume 6, number 2, pages 13-30, 2011.
- [16] Christian Bird, Alex Gourley, Prem Devanbu, Anand Swaminathan, Greta Hsu.: *Open Borders? Immigration in Open Source Projects*, In proceedings of the 4th International Workshop on Mining Software Repositories (MSR'07), pp.6-13, 2007.