ソフトウェア開発プロジェクトの探索的分析を支援する視覚的分析ツール

矢崎 聖也 三末 和男 田中 二郎

本研究では、開発プロジェクトにおけるチケットデータの探索的な分析を支援するための視覚的表現を構築し、それを用いたツールを実装した。視覚的表現には、チケット属性値、イベント属性値、イベントの存在、日時のすべての情報を表現し、また縮小されてもそれらの傾向を視覚的に維持する表現を構築した。そして、数千件から1万件程度の大量のチケットデータを上記の視覚的表現によって提示し、かつ属性を切り替えることで表現間を対比できるツールである「ColorWave」を実装した。ケーススタディでは、視覚表現による概観と切り替えによる対比を通して、実際のチケットデータから多角的な知見を得られたことを示す。活動に関する情報を持つ大規模なプロジェクトは多く存在しており、本研究の成果はそれら活動の分析の支援に役立つことが期待される。

We propose a new visualization approach and a tool that helps explorative analysis of ticket data of software development projects. Our visualization approach presents ticket attribute, event attribute, event existence and time at once and shows trend of them when the visualization zoom out. ColorWave is a tool which uses the visualization approach to show up to 10 thousand tickets and uses attribute switching to contrast those visualizations. In a case study, we show multilateral knowledge extracted from some real ticket data using the general view and switching. There are many projects which have a lot of activity information and which can be analyzed with our proposed tool.

1 はじめに

現在、プロジェクト管理システムによって、多くの ソフトウエア開発プロジェクトにおける活動情報が管 理されている。プロジェクト管理システムが有する活 動情報には、バグや要望の内容や進捗に関する情報が あり、それらは「チケット」と呼ばれる形で管理され ている。プロジェクトの関係者は、バグや要望などに 対応するチケットを作成し、さらに進捗があった時に チケットを改変する。そのとき、バグや要望の概要、 詳細、進捗状況などの内容がチケットの属性値として 記録され、さらに属性値は進捗に応じて改変される。

プロジェクト管理ツールのチケットなどを通して活 動状況を分析するための先行研究や既存のツールもあ るが、それらは既に知られている類の傾向について分 析することを主目的としている。しかしながら、特定 の指標や集計法にこだわらずに、単に活動状況の傾向 を知ることも、プロジェクトの活動状況についての理 解を深め、有意義な活動や評価を行う上で重要なこと である。そこで、プロジェクト管理ツールのチケット 群の概観と、異なる側面からの概観の対比を可能とす ることで、活動状況の探索的分析を助けることが本研 究の目的である。そのために、絞り込みやデータのそ ぎ落としをすることなく、大量のチケットの属性値や 改変を表現するための視覚的表現を実現した。また、 その視覚的表現を実装した分析ツール ColorWave を 開発するとともに、ColorWave を用いたケーススタ ディを実施した。

 $[\]label{thm:constraint} \mbox{Visual support tool for explorative analysis of software development projects.}$

Seiya Yazaki, 筑波大学大学院 システム情報工学研究 科 コンピュータサイエンス専攻, Dept. of Computer Science, University of Tsukuba.

Kazuo Misue, 筑波大学大学院 システム情報工学研究 科 コンピュータサイエンス専攻, Dept. of Computer Science, University of Tsukuba.

Jiro Tanaka, 筑波大学大学院 システム情報工学研究 科 コンピュータサイエンス専攻, Dept. of Computer Science, University of Tsukuba.

2 関連研究

本研究は、ソフトウエア開発プロジェクトの活動情報の可視化・分析と、大量の時間変化するデータの可視化に関連している。

2.1 ソフトウエア開発プロジェクトの視覚的分析 ツール

ソフトウエアの開発プロジェクトの活動に関する情報を可視化し、そこから知見を得ることを支援するための研究は多く行われてきている。

例えば、Storey らの調査[7] によれば、レポジトリから得られるソースコードのファイルや行、およびそれらに対する変更履歴に着目した研究が数多くある。しかし、本研究では、ソフトウエア開発プロジェクトのプロジェクト管理ツールのチケットデータについて着目している。

MDSViews [4] は、CVS(レポジトリ)とBugzilla(プロジェクト管理ツール)から得られる情報に多次元尺度構成法と呼ばれる手法を適用することで、プロジェクト内の要素同士の関連性を網図によって視覚的に表現するものである。これに対して、本研究では、プロジェクト全体の活動を、個々の活動の情報の集合として視覚的に表現する。

Social Health Overview (SHO) [1] は、Bugzilla か ら得られた個々のチケットとその属性を点と色として 表現として見ることで、プロジェクトの活動状態の 健全性を評価するためのツールである。SHO は、プ ロジェクト管理ツールから得られる個々のチケットの 情報を視覚的に表現することで、全体の活動の様子 を視覚的に表現する点と、表現に対して操作を行う ことで多角的に分析する点で本研究と類似している。 しかし、SHO はいくつかの時点での状態それぞれの 視覚的表現を並べることで時間的な推移を表現して いる。これに対して本研究では、一つの視覚的表現の 中で時間的な推移を連続的に表現しているので、それ ぞれのチケットの時間的な連続性や、時間と属性値の 相関性を一望することができる。また、SHO は最大 3500 件程度のチケットを対象としており、最大数万 件のチケットを対象としている本研究とは、規模の面

でも異なっている。

2.2 大量の時間変化するデータの視覚的分析ツール時間変化する大規模なデータ (時系列データ) をまとめて可視化することで、大量のデータに対する理解を助ける研究も多く行われてきており、Aigner らの調査[2] においてそれらの先行研究が示されている。また、Extreme Visualization[6] には、単なる表でない方法でより視覚的に大量のデータを表現することについての先行研究が示されている。

複数の属性値を持つ大量のデータについての傾向を示す研究としては、上記以外にも Chen らによるdendrogram を用いた研究 [3] や Muelder らによる研究 [5] が挙げられる。ガン検査の結果についての大量のデータをクラスタリングして並べて表示している Chen らの研究 [3] は、色の合成を用いながら大量のデータを並べて傾向を示す点が本研究と類似しているが、データの時間位置や長さというものが存在しない。クラスタ上でやりとりされた大量の MPI^{†1}メッセージの傾向を可視化している Muelder らの研究 [5] は、大量のデータの属性値のみならず時間位置や長さも表現し、データの並びを変えることで傾向を見いだす点も似ているが、イベントを表現していない点、データが重なって隠れてしまいうる点、切り替えなどによる対比を行っていない点、が異なる。

このように、時間変化するデータの可視化について、本研究はデータの時間位置、属性値、イベントの全てを同時に表現し、かつ異なる側面から対比することを可能にしている点が特徴的である。

3 対象データ

本研究では、チケットと呼ばれるデータの数千から数万件の集合を対象としている。チケットを見ることで、個々のバグや要望などについての現状とこれまでの進捗を知ることが出来る。bugzilla †2 や redmine †3 などの広く使われているプロジェクト管理ツールは、問題事に関するある時点での状態を「属性」とし

^{†1} Message Passing Interface

^{†2} http://www.bugzilla.org/

^{†3} http://www.redmine.org/

て記録することが一般的である。属性としては、たとえば Assignee (担当者)、Status (状態: NEW, RESOLVED, REJECTED など)、Product (プロジェクトにおけるカテゴリ分け)、Priority (重要度) といったものがあり、 $5\sim20$ 程度の属性が提供されていることが一般的である。進捗は属性値いくつかの変化として表され、本研究ではこういったチケット状態の変化のことをイベントと呼ぶ。

3.1 チケットデータの例

ソフトウエア開発プロジェクトにおけるチケットの 実例を一つ示す^{†4}。このチケットは、表 1 に示す属 性値を初期状態に持っており、表 2 に示すイベント を持っている。イベントそれぞれが、いくつかの属性 値の変化を伴っている。この例では、バグが担当者に 割り当てられてから、解決と再開を繰り返しており、 その過程でいくつかの属性値が変更されている。

表 1 属性の初期値の例

属性名	属性値
Status	NEW
Assignee	mono-bugs@lists.ximian.com
Resolution	(なし)
Severity	Critical
Priority	P5 - None
Product	Mono: Runtime
Found in Version	2.0.x

3.2 チケット群、チケット、イベントの構造 本研究では、チケットを扱うために以下のような概 念を定義した。

まず、チケットに対するある時点での更新を表すイベントというものを定義する。イベントは日時と、いくつかの属性 (イベント属性) を持つ。チケットは、1つ以上のイベントと、いくつかの属性 (チケット属性) を持つ。最初のイベントはチケットが作られた瞬間に対応する。また、チケット属性には、チケットの作られた日時 (始端の日時) や、閉じられた日時 (終端の日時) も含む。また、任意数のチケットの集合で

表 2 イベントと属性値の変更履歴の例 (部分)

イベント日時	属性名	新しい属性値
2009/03/26	Assignee	gonzalo@novell.com
06:14:04	Severity	Normal
2009/03/27	Status	RESOLVED
20:08:07	Resolution	FIXED
2009/03/30	Status	REOPENED
08:09:39	Resolution	(なし)
08:10:20	Status	NEEDINFO
2009/03/30	Status	RESOLVED
17:08:53	Resolution	FIXED
2009/04/01	Status	REOPENED
10:05:53	Resolution	FIXED
10:06:43	Status	NEEDINFO
15:15:05	Status	REOPENED
2009/04/01	Proprity	P2 - High
15:15:38	Severity	Major

あるチケット群というものも定義する。これはいわゆるプロジェクトなど一つに対応するものである。

たとえば、表 1 のチケット例ならば、チケットの作成日時、終了日時、それに Product 属性値 ("Mono: Runtime") などがチケット属性とみなせる。そして、Status 属性値 ("NEW", "REOPENED" など) やAssignee 属性値 (担当者のメールアドレス) などはイベント属性であり、イベントの都度に更新されているとみなせる。

制約として、あるチケット群内のチケットやイベントの持つことのできる属性の種類は一定である。なお、表2の例ではイベント属性の差分のみを示しているが、たとえ値に変化がなくとも全てのイベント属性値をイベントが保持していると見なすものとする。また、チケット属性やイベント属性それぞれの値は不変である。チケットの状態変化は、新しいイベントが作成されることによってのみ表される。

このように捉えることで、既存のプロジェクト管理 ツールで扱われているチケットのメタデータを網羅し つつも、イベント属性とチケット属性と時間という分析に使える軸を提供する。活動は必然的に時間を伴う ものなので、チケットの時間による変化を表すことは 重要である。

^{†4} https://bugzilla.novell.com/show_bug.cgi?id=489019 の一部を引用

4 分析ツールの設計方針

本研究では、プロジェクトのメンバーや観察者が、 対象プロジェクトにおける活動の傾向や法則性を見出 すことが活動分析であると捉えている。そして、探索 的な活動分析のために、以下にあげる設計方針を考 えた。

4.1 属性値や日時をあまさず表現すること

属性、時間情報それにイベントも持つチケットを概観する場合、事前にデータに対して何らかの検索や集計を施すことが必要とされてきた。たとえば、いくつかのプロジェクト管理ツールでは、さまざまな集計条件や描画方法を指定することで棒グラフや円グラフを生成する機能が提供されている。探索的分析では、どのような傾向があるのかについて事前に分からないので、検索条件や集計方法を見いだすまでに多数の試行錯誤を行う必要が生じる。

また、既存のツールや既存研究は特定の統計手段、指標や観点を前提としている。たとえば、活動情報を概観したり対比したりすることが可能な先行研究 Ellis ら [1] は、プロジェクトの健全性評価に着目した上で指標などの前提知識を持ち込んでいる。そのため、特定の指標によって健全性評価をするためには有用だが、どのような傾向があるのかそれ自体を探索するためには適さない。

探索的分析のためには、まずチケットやイベントの 属性値や日時を情報の損失を抑えつつ表現すること が重要であると考える。チケットやイベントの属性値 や日時を眺められるようにすることで傾向や法則性 に気づけるようにすることが、活動状況の探索的分析 の端緒となると考えられる。

4.2 複数の面からの俯瞰を可能とすること

表やあるいは折れ線グラフや円グラフなどの手法を用いて属性と時間情報とイベントも持つチケットを表現した場合、チケットの持つ情報の限られた側面のみについての俯瞰が得られる。そのため、一部の情報を元に、活動状況を概観したり、概観するために試行錯誤することとなる。また、チケットの持つ時間情報

や属性およびイベントを、大量の図表や文字列の羅列で表現する方法は、数千件以上の大規模なデータの全体像を俯瞰するには適さない。そこで、活動分析の過程での自由度を確保するために、属性や時間情報、イベントの情報をまとめて俯瞰できるようにすることを考える。

4.3 対比を容易に行えるようにすること

クエリやパラメーターを与え、チケットデータの持つ情報の一部を表示するモデルでは、さまざまなクエリやパラメータを与え、出力された多数の別々の表現から相関性を見いだすことで対比することになる。傾向を見いだすためにこのような試行錯誤をするためには、クエリやパラメーターの再入力と結果の表示を何回も反復しつつも、同時に多数の結果それぞれの間にある相関性を見いだす能力が必要とされる。そこで、多角的な分析や対比のために、対比を容易に行えるようにすることを考える。

5 視覚的表現

本研究で開発した視覚的表現手法は、チケットそれぞれの視覚的表現をチケットの数だけ並べることで、チケット群の傾向を視覚的に表現するものである。これは、チケット属性値、イベント属性値、イベントの存在、そして時間の全てを大量のチケットについて一画面にて可視化するための視覚的表現である。個々のチケットの視覚的表現でイベント属性値、イベントの存在、時間を表現し、チケットの並び順でチケットの属性値を表すことで、画面全体にそれらの情報を収めるようにした。また、個々のチケットの視覚的表現が極小に圧縮されたとしても、色の傾向としてデータの傾向が維持されるようにすることで、大量のチケットを対象にできるようにした。

5.1 視覚的表現へのマッピング

まず、一つのチケットを 図 1 のようにして表現 する。

イベントが点に対応し、その間が線分で結ばれている。イベントの点と、次の点までの間の線分の色は、 そのイベントの特定のイベント属性の値に対応する

図 1 チケット一つの描画例 (横軸が時間に対応)

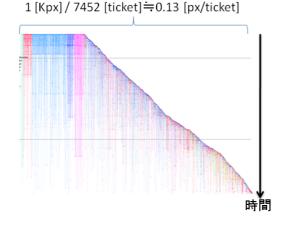


図 2 大量のチケットの表現 (縦軸が時間に対応)。千ピクセル程度の幅で 7452 チケットを表現した。

色相の色で塗られている。点より線分の彩度を低くすることで、たとえそのイベントの前後で値が変化していなくても、点が識別されやすくなっている。

また、点の位置はイベントの日時に対応しており、 チケットの表現の端の座標はチケットの生存期間^{†5}に 対応する。

次に、上記の表現手法を用いたチケットを並べることで、大量のチケットを表現する。縦軸を時間にとり、図1の表現を90度回転した表現を、チケット属性についてユーザーによって指定されたソート順によって横に並べる。例えば、ケーススタディにて用いているチケットの実データ7452件を表示すると、図2のようになる。また、この図の一部を拡大し、拡大縮小による劣化を排除すると、図3のようになっている。

画面幅よりもチケット数が多い場合、1 ピクセルに 多くのチケットの色が合成されて描画されうる。そ のため、近傍チケット間で値が大きく異なったとして も、分散の存在を読み取れないという問題がある。

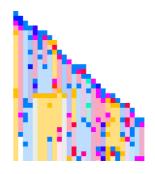


図3 視覚的表現の一部を拡大した図(縦軸が時間に対応)

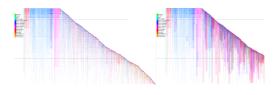


図 4 アニメーション過程

そこで、埋もれている値の存在をアニメーションによって強調することを可能にした(図4)。あるピクセルについて、そのピクセルの色から一定以上離れた色相となる色が、左右に閾値以上存在していた場合、これを埋もれている色であると見なす。この機能では、埋もれている色を、左右に広げて描画する。また、広げる幅は、本来の幅の1倍から、任意の幅の間でスムーズにアニメーションする。

5.2 表現される情報

この表現から、色の傾向や点の集まりを見出すことができる。色や点の位置から、ユーザは属性値やイベントについて以下の傾向が読み取れる。

- チケットの存在密度: その時点とチケット属性値におけるチケットの存在密度によって、描画結果の彩度が変わっている。そのため、画面上の濃淡からチケットの存在密度の傾向を読み取れる。
- イベント属性値の分布: それぞれの時点とチケット属性値におけるイベント属性値によって、描画 結果の色相が変わっているため、色相からイベント属性値の傾向を読み取れる。また、類同の要 因により、同じイベント属性値を有する時点・チケット属性値の領域が固まって見える。

^{†5} チケットの始端日時から終端日時までの区間

- イベントのグループ: イベントの時点・チケット 属性値に対応する位置に点が打たれているため、 近接の要因によって、時点とチケット属性値の近 いイベントが自然に固まって読み取れる。
- 埋もれている値の有無: 埋もれている値を強調 するアニメーションによって、時点・チケット属 性値それぞれにおいて埋もれている値があるか どうかを読み取れる。これによって、真に値が均 ーな時点・チケット属性値とそうでない部分を識 別できる。

5.3 視覚的表現の圧縮

視覚表現を縮小することで限られた画面領域へ収めたとしても、色や点の傾向は可能な限り保持されるため、上記の情報を読み取ることができる。それによって、時間情報を損なわないようにしつつも一つのチケットの持つ情報を狭い幅で表現し、それらの集合としてチケット群を表現することで、ユーザに傾向を見せている。

6 分析ツール ColorWave

上記の視覚的表現を利用し、チケットデータの探索的な分析を支援するためのツールである ColorWave を実装した。ColorWave では、表現を切り替え対比するために、以下の操作を可能としている。

6.1 着目するチケット属性の切り替え

視覚的表現におけるチケット属性 (横軸のソートに利用) をどれにするのかを選択できるようにした。個々のチケット内のイベント属性 (色相に対応) は維持したまま、横軸 (並び順) が変わる。 チケットの横幅は一定であるので、たとえば始端日時 (チケット属性) をチケットのソート順に用いたならば、縦軸が日時なので、チケットの上端の並びの傾きが [チケット数 / 日時] と対応する。また、複数のチケット属性に着目し、それらによって安定ソートすることもできる。たとえば、まず Product チケット属性ごとに横に並べ、Product 属性が等しいチケット同士は始端日時 (チケット属性) で順序を決める、ということもできる。

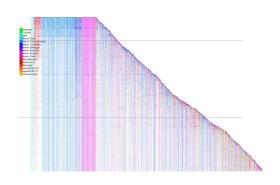


図 5 (Mono) 色: Product 属性、始端日時順

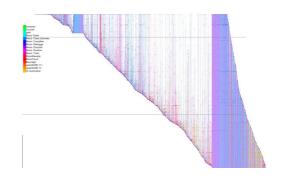


図 6 (Mono) 色: Product 属性、終端日時順

6.2 着目するイベント属性の切り替え

着目するイベント属性 (色相に対応) をどれにする のかを選択できるようにした。チケット属性 (横軸に 対応) を維持したまま、色相が変わる。

7 ケーススタディ

Mono project $^{\dagger 6}$ のチケットデータ 7452 件を、bugzilla $^{\dagger 7}$ から取得し、これについてケーススタディを行った。また、プロジェクト管理ツールである Redmine $^{\dagger 8}$ それ自体の開発についてのチケットデータ $^{\dagger 9}$ 2253 個を取得し、これについてもケーススタディを行った。

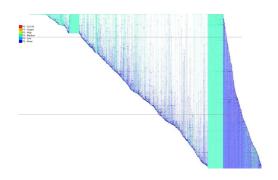


図7 (Mono) 色: Priority 属性、終端日時順

7.1 チケットの処理状況の推移

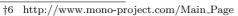
始端時刻(図5)や終端時刻(図6)(チケット属性)を並び順にとることで、チケットがほぼ線形に増えており、処理もほぼ線形なペースであったことが分かる。また、下端まで伸びているチケット群の存在から、放置されているチケットの存在も見て取れた。

また、たとえば Product 属性を色相にしてみると、一定幅の同じような色合いの縦帯があることから特定 Product のチケットがいくつかまとまって出現・処理される傾向があること、近年になって出現した Product (黄) の存在とその活動量の程度、といった点を見て取ることができる。

7.2 作業タイミングの傾向

イベントの存在を表す点は、どのチケット属性を選ぶかによって異なる位置に打たれる。そこで、始端時刻 (チケット属性) や終端時刻 (チケット属性)、あるいは一般的なチケット属性 (Classification など) を選ぶことで、それぞれに応じた作業タイミングの傾向を見出すことができた。

たとえば、色相を Product チケット属性値に対応させる(図6)と、全体的に特定の縦位置に横並びが見えたり、さらには赤い点ばかりのの横並びが見える。このことから、Product について横断的な作業の存在や、赤い Product についての集中的な作業の存



^{†7} http://www.mono-project.com/Bugs

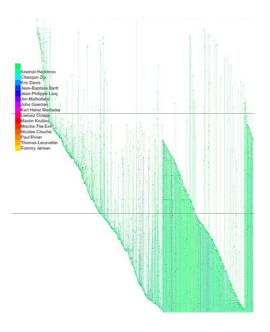


図8 (Redmine) 色: Assignee 属性、終端日時順

在を認識できる。また、Product について横断的な作業については、Priority 属性を色相にしてみると(図7)、シアンの Priority のチケットもあるにも関わらず青い Priority が支配的であると見てとれる。これらの Product について横断的な点の並びは、チケット属性を始端で並べても終端で並べても同じく見出せることから、mono project においては特定 Priorityのチケットをまとめて処理していることがあり、かつそれはチケットの完了とも相関していないと分かる。

7.3 チケットの経緯による差

たとえば、処理済みのチケットの大半は Assignee が青色であるのに対し、そうでないチケットは最初が緑 (未記入) であり、途中から青色に変わっていることが見て取れた(図8)。このように、イベント属性値の時間による変化とその他の情報の相関性を見ることで、チケットの経緯がどのような差に繋がっているのかを見出すことができた。

7.4 プロジェクトの体制の変化

たとえば、 図 9 を見ると、ある時点から始端の傾きが緩やかになっており、また同時に担当者の種類が

 $[\]dagger 8$ http://www.redmine.org/

^{†9} http://www.redmine.org/projects/redmine/issues

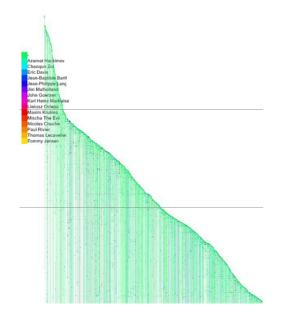


図 9 (Redmine) 色: Assignee 属性、始端日時順

増えていることが分かる。さらに、イベント属性だけを切り替えると、やはり同じタイミングから属性値が切り分けられるようになっていることが分かる。途中から多人数体制にシフトしチケット管理がきちんと行われるようになった、この redmine の場合のように、体制や運用の変化を見出すこともできた。

7.5 議論

このように、mono project や redmine の内情について、特に事前の仮定を置かずとも、データを眺めることによる気づきが得られた。また、イベント属性やチケット属性を切り替えることによって複数の側面からの俯瞰が出来、それらの対比から上記のような情報を見出せた。これによって、両プロジェクトの内情についてのさまざまな種類の傾向を見出し活動分析することができた。

8 まとめ

チケット属性値、イベント属性値、チケットやイベントの存在を、圧縮可能な視覚的表現を用いて提示し、切り替えによって表現間を比較可能にする探索的分析ツールである ColorWave を開発した。ColorWave の視覚的表現がそれらの情報全てを一望することを

可能にし、視覚的表現を対比することで探索的分析を 行うことが、チケットの処理状況の推移、作業タイミ ングの傾向、チケットの経緯による差、プロジェクト 体制の変化、といった情報を読み取ることを可能にし た。これによって、プロジェクトの活動状況の探索的 分析が可能になった。

謝辞

研究および本論文の執筆にあたり、筑波大学システム情報工学研究科インタラクティブプログラミング研究室のメンバーの方々から、多くの有意義な意見を頂いており、深く感謝しております。

参考文献

- [1] Jason B. Ellis, Shahtab Wahid, Catalina Danis, and Wendy A. Kellogg: Task and social visualization in software development: evaluation of a prototype. In CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 577 586, New York, NY, USA, 2007. ACM.
- [2] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann, and Christian Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 14, No. 1 (1985), pp. 47 60.
- [3] Jin Chen, Alan M. MacEachren, and Donna J. Peuquet. Constructing overview + detail dendrogram-matrix views. *IEEE Transactions on Visualization and Computer Graphics*, Vol 15, No. 6 (2009) pp. 889–896.
- [4] Michael Fischer and Harald Gall. Mds-views: Visualizing problem report data of large scale software using multidimensional scaling. In Proceedings of the International Workshop on Evolution of Large-scale Industrial Software Applications (ELISA), September 2003.
- [5] Chris Muelder, Francois Gygi, and Kwan-Liu Ma. Visual analysis of inter-process communication for large-scale parallel computing. *IEEE Transac*tions on Visualization and Computer Graphics, Vol 15, No. 6 (2009) pp. 1129 1136.
- [6] Ben Shneiderman. Extreme visualization: squeezing a billion records into a million pixels. In SIG-MOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 3 12, New York, NY, USA, 2008. ACM.
- [7] Margaret-Anne D. Storey, Davor Čubranič, and Daniel M. German. On the use of visualization to support awareness of human activities in software development: a survey and a framework. In Soft-Vis '05: Proceedings of the 2005 ACM symposium on Software visualization, pp. 193 202, New York, NY, USA, 2005. ACM.