

共有フォルダから“活動の流れ”を抽出する

Extraction of the Flow of Activities from Shared Folders

斉藤 典明

SAITO Noriaki

日本電信電話（株）NTTセキュアプラットフォーム研究所
NTT Secure Platform Laboratories

概要：現在の組織活動では、様々な資料を組織活動の記録として残す活動が定着している。そして、長期間に渡って組織活動を行うことによって組織活動の記録が膨大になる。その結果、資料の全体像の把握や必要な資料の探索が困難になり、蓄積された知見が十分に活用できないという問題が発生する。この問題を解決するために、資料作成のタイムスタンプに着目して資料の時系列分布を可視化することで活動の流れを抽出し、資料探索の糸口とする手法を提案する。

Abstract: Shared folders are used as the simplest method for the organizational knowledge succession. So, many documents which are generated by organizational activities are accumulated in shared folders. However, shared folders have a big problem of originating in use for a long term. That is, folder structure become chaotic and using of the stored documents becomes troublesome. So, organizational knowledge which were stored by using shared folder becomes impossible to use enough. To solve such problem, I propose a new shared folder interface which offers the beginning of file discovery. The interface can be visualizing the flow of organizational activities by using extracted timestamps from files in the shared folder, and it is described in this paper.

1. はじめに

現在の組織活動では、様々な資料を組織活動の記録として残す活動が定着している。これらの資料は多くの場合、組織内の共有フォルダや、引継ぎ用に作成されたメディアの中にフォルダとして蓄積されている。活動期間が長くなればなるほど蓄積される資料も膨大な量になる。ここに、新しく組織活動に加わったメンバーや活動を引き継いだメンバーは、過去の資料を活用することにより効率的に活動を遂行できると考えられる。しかしながら、共有フォルダや引継ぎ用のフォルダは、資料を整理して蓄積した人物と、資料を活用する人物が違うこと、資料も膨大であることなどの理由から、目的の資料を探し出すには膨大な労力を必要とする状態になりがちである。

目的の資料を探し出すためには、資料に対するキーワードなどで検索する方法があるが、キーワードは特定の資料を探し出す場合には有効であるものの、実施しようとしている活動の全体像を把握するには適切ではない。蓄積資料は組織活動に関わるものであり、断片的な資料の探索だけでなく資料の作られた経緯や背景を含めて探し出す必要がある。そこで、

資料の時系列や時系列の変化から組織活動の特徴を抽出するなど、蓄積資料から「活動の流れ」を抽出することにより目的の資料を取り出す手法を検討した。

2. 蓄積資料に対するニーズ

これまで、研究所における事例を調査してきた。調査対象とした研究所では、10～20人程度の研究グループを最小単位として研究開発活動をおこなっている。調査した複数の研究グループでは全て自発的に共有フォルダを開設し組織活動で使用された資料を蓄積している。特に蓄積期間が10年以上になると、蓄積量は数100GB～数TBになっていた。

2. 1 共有フォルダにおける問題

このような大量に資料を蓄積している共有フォルダから目的の資料を探すことは困難になりつつある。一般的に、資料を探すのが困難になる原因として次のような問題が考えられる。

- (1)情報の分類に関する問題
- (2)利益不均衡に関する問題
- (3)記憶や利用者に関する問題

(4)手がかり不均衡の問題

(5)負荷の問題

「情報の分類に関する問題」は、そもそも資料を共有フォルダのような分類によって蓄積することには限界がある、という問題である。例えば、複数の意味を持つ資料を特定の分類のフォルダに入れることなどである。「利益不均衡に関する問題」は、資料を入れるときと取り出すときで労力と効果にアンバランスがある、という問題である。例えば、取り出すときを考慮して蓄積する際に多大の労力を払うことは稀である。「記憶や利用者に関する問題」は、資料を蓄積する人物と取り出す人物が同一の場合は、蓄積と取り出しのタイムラグによりどこに蓄積したか思い出せない。蓄積と取り出しで別な人物の場合は、そもそもどこに蓄積したのかわからない、という問題である。例えば、昨年資料を取り出したときに、本人であればどこにしまったのかを思い出せない場合であり、第三者であればそもそも資料があるのかもわからない。「手がかり不均衡の問題」は、資料の蓄積時と取り出す時で、同じ資料であっても趣旨がことなるので整理したラベル名や検索キーワードが一致しない問題である。例えば、ファイル名をイベントで蓄積したものの、探すときは議事録というキーワードで探したい場合などである。「負荷の問題」は、資料が大量になることによって、記憶・視覚・操作という観点でユーザの負担が増えるという問題である。例えば、共有フォルダの第一階層が5個しかない場合と50個もある場合では、後者の方が探す際に負荷が高い。

このような理由から、膨大な資料を持つ共有フォルダ内から資料を探すのが困難になり、蓄積するだけの共有フォルダになるか、ファイルの受け渡し場所としての共有フォルダに陥る場合がある。

2. 2 過去資料の価値

蓄積量が膨大な共有フォルダには、蓄積情報から目的の情報へたどり着くまでの障壁がある。蓄積資料には障壁を乗り越えるだけの価値のあるのかという疑問もある(図1)。

そこで、そもそも過去の資料を活用するメリットについて研究所員にアンケートをおこなった。アンケートは約200人の所員のいる研究所において2014年2月に実施した。所員の2割強になる45人から回答を得た。

アンケート項目は表1の通りである。アンケートの質問の注釈として、対象とする業務は、会社の業務および会社で委嘱を受けている業務とした。後者は学会の委員の活動などが含まれる。過去資料の活用という観点において、過去資料の点検作業は対象

外とした。過去資料については、本人作成、他者の作成を問わないこととし、ひな形・マニュアル・チェックシートの参照そのものは対象外、ひな形やマニュアルを活用して生成された資料を対象とすることとした。また、参照・流用の有用性は、ないよりあった方が良い場合も有用と判定する、という前提条件とした。

資料を探す障壁 過去資料の価値

図1 過去資料の価値

表1 アンケート項目(その1)

	質問項目	回答形式
1	これまでの業務、普段の業務の中で過去資料の参照・流用を必要とする作業はありますか？	Yes/No
2	(質問項目1がYesの場合)これまでの業務、普段の業務の中で、十分に習熟していない作業において、以下のどのタイプの作業で過去資料の参照が有用でしたか？	複数回答可 a.ほぼ毎日、または、逐次発生する作業 b.週1回程度の作業 c.月1回程度の作業 d.四半期の1回程度の作業
3	(質問項目1がYesの場合)これまでの業務、普段の業務の中で、ほぼ習熟した作業において、以下のどのタイプの作業で過去資料の参照が有用ですか？	e.半年に1回程度の作業 f.1年に1回程度の作業 g.2年に1回程度の作業 h.ごくまれに発生する作業 i.その他()
4	(質問項目1がYesの場合)過去資料を参照する理由はなんですか？	複数回答可 a.やり方がわからないから b.やり方を覚えてないから(忘れたから) c.参照や流用した方が早いから d.確認漏れやミスが減るから e.以前の作業の結果や留意事項を知りたいから f.作業そのものが複雑なので過去資料が必要だから g.その他()
5	(質問項目1がNoの場合)過去資料を必要としない理由はなんですか？	(該当者なしだったため略)

回答結果は、質問項目1は全員がYesの回答であった。そのため質問項目2~4について分析する。質問項目2と質問項目3は慣れていない作業に対する過去資料の活用と慣れている作業に対する過去資料の活用に関する質問項目であり、図2に示す。質問項目4の過去資料の利用形態であり、図3に示す。

2つのグラフから読み取れることとして、共有フォルダから過去の資料を取り出す活動は毎日頻繁に

起こっているわけでない。また、慣れている作業でも慣れていない作業でも一定スパン、特に1年以内の資料を取り出すことが多いことが読み取れる。そして、過去資料は主に再利用が目的で活用されることがわかった。

反対に、必要な資料は過去の時系列の中に均等に出現するのではなく、一定の周期をもって出現すると考えられる。言い換えると、過去資料を必要とする活動は、一定の流れがあると考えられる。

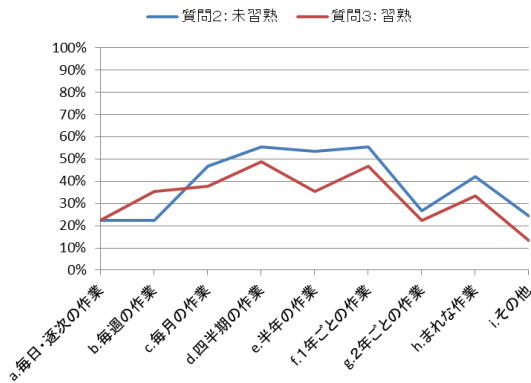


図 2 過去資料の利用頻度

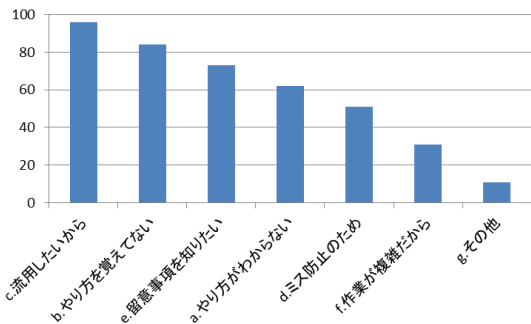


図 3 過去資料の利用目的

2. 3 共有フォルダからの資料活用動向

さらに、表1のアンケートとほぼ同じメンバーに、共有フォルダからの資料の活用方法について2015年2月に追加のアンケートをおこない、37名から回答を得た。アンケート項目を表2に示す。Q1の結果は、約9割の33名がYesであった。この「自発的に共有フォルダから過去資料を探すことがある」と回答した33名について、「自分が作成した資料の活用の有無」(Q2)、「他人が作成した資料の活用の有無」(Q3)、「他人の資料活用の契機」(Q4)について質問をおこなった。Q1~Q4までの回答結果を図4に示す。

グラフから、「自分が作成した資料」、「他人が作成した資料」とともに活用する必要があることがわかる。わずかながら、自分の資料よりも他人が作成した資料を探す目的が多い。また、他人が作成した資

料は1/3程度は自発的に見つけられているが、2/3は過去のメールや問い合わせなどにより共有フォルダ内の場所を特定してもらうことによって取り出せている実態がつかめた。

表 2 アンケート項目(その2)

質問項目	回答形式
Q.1 (報告作業などで強制的に使われるときは除いて) 自発的に共有フォルダから過去資料を取り出すことはありますか?	Yes/No
Q.2 Q.1がYesの場合 自分が作成したファイルを活用しますか?	Yes/No
Q.3 Q.1がYesの場合 他人が作成したファイルを活用しますか?	Yes/No
Q.4 Q.3がYesの場合 他人のファイルの所在はどのようにして知りましたか? (複数回答可)	1.共有フォルダ内を探して見つけた 2.過去にメールなどでお知らせがあった 3.必要になったとき作成者に問い合わせして入手した 4.その他、具体的に記述してください。()

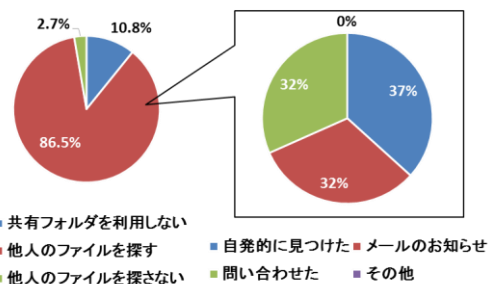
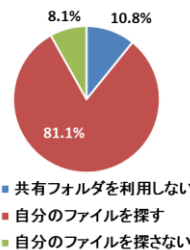


図 4 共有フォルダからの資料の活用

2. 4 活用動向からの仮説

以上を踏まえて、共有フォルダの資料を活用する上での仮説を述べる。

過去の資料は、作業を効率化するために活用されることが多い。そしてそれは、自分自身が作成した

もの、他人が作成したもの、慣れた作業、慣れない作業を問わない。また、過去資料を必要とする作業は、一定の作業の流れがある。必要な資料の約6割は現在の共有フォルダの機能だけでは探せていない。よって、過去資料を必要とする作業を対象にファイルを容易に取り出せるインタフェースがあると有益になる、と考えられる。

3. ファイル探索インタフェース

これまで、過去資料を大量に蓄積した共有フォルダから、目的の資料を効率的に探し出すことができる、組織知識継承基盤を検討している。本検討では、蓄積ファイルの活用状況の調査結果を踏まえて、より効果的に蓄積ファイルを活用する方法を提案する。

3.1 ファイル探索インタフェースの概要

まず、これまで検討してきた組織知識継承基盤の概要について述べる。共有フォルダ内の蓄積資料は、長期間の組織活動により、膨大かつ分類整理が混沌としてしまい、資料の所在わからなくなる傾向にある。また、組織活動で生成される資料の多くは組織活動のイベントと紐づくことが多いこともわかった[1]。多くの場合、イベントの確認では、カレンダー形式のインタフェースを用いることが多い。そこで、蓄積ファイルをカレンダー形式のインタフェースで利用できる方式の組織知識継承基盤を考案した[2]。

カレンダー形式のインタフェースから蓄積ファイルへアクセスするには、蓄積ファイルの時系列情報が必要になる。そこで、すでに蓄積されている資料から時系列情報を取り出す方法の一つとしてファイルのタイムスタンプに着目した。クローラーにより共有フォルダ内を巡回し、最終更新日時のタイムスタンプを収集する(図5)。また、タイムスタンプの取得対象ファイルは、蓄積ファイルの全てではなく、組織知識共有に必要なものに絞った。ここでは、主にMS-Office系文書、PDF文書を対象とした。対象を絞ることによりOSの機能で作成されるファイルなどが除外されるため組織活動で生成される資料に絞ることができる。また、タイムスタンプの収集と同時に、フォルダ名やファイルの蓄積場所の情報も収集する。

収集された情報はカレンダー型インタフェース内で再整理する。カレンダー型インタフェースでは、時系列と活動項目で指定される領域にファイルの存在とファイルへのリンクが表示される(図6)。ここで、活動項目は、共有フォルダ内のフォルダ名とする方法、フォルダ名から新たに生成する方法[3]など

幾つかの手法がある。

次に、このようにして生成された、カレンダー型インタフェースによるファイル探索の手順について述べる(図7)。はじめに、年・月の時系列(図7-①)と、フォルダ名から抽出された活動項目(図7-②)で指定される領域に、ファイルの有無が表示(図7-③)される。資料探索者は、必要な資料の項目と時期を頼りに資料を探す。指定した領域のファイル数をクリックすると該当領域のファイル名が表示(図7-④)され、ファイル名をクリックするとハイパーリンクにより共有フォルダ内の該当の資料が参照可能になる(図7-⑤)。この手法により、無意味に共有フォルダを探索する稼働が削減される効果がある。

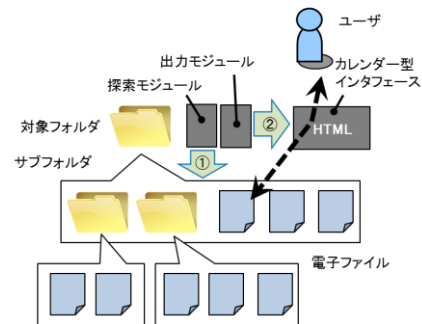


図5 組織知識継承基盤

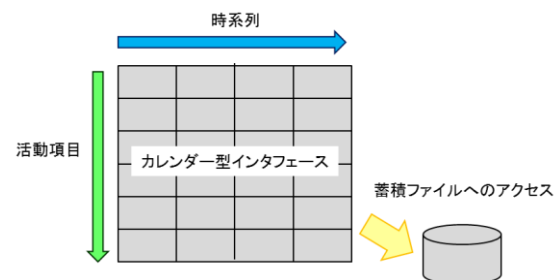


図6 カレンダー型インタフェース



図7 操作イメージ

3.2 適用事例

カレンダー型インタフェースを、共有フォルダ相当の実データに適用した。適用したデータは、毎年実施されるあるイベントの実行委員の引継ぎデータである。実行委員は、毎回、別々のメンバーで組織

される。毎回の実行委員のメンバーは、作成した資料をまとめてフォルダに蓄積し、次年度の実行委員に引き継いでゆく。引き継いだ実行委員は、それまでの実行委員が作成した資料を頼りに担当年の活動をおこなう(図 8)。その結果、2013 年の実行委員は 2007 年からの引継ぎ資料として、ファイル数 6,176 個、フォルダ数 548 個、データ量 4.2GB を引き継いだ。その結果、現在のフォルダ構造では、資料を引き継いだ実行委員は、過去資料は大量にあるものの、どこから資料を読み解いてよいのかわからない状態に陥る。

このような引継ぎ資料に対してカレンダー型インタフェースを適用した結果を Appendix Fig.1 に示す。ここで、横軸は、各実行委員の担当したイベントを足掛け 3 年長とし、縦軸は、各実行委員のフォルダでまとめた。これにより、2013 年の実行委員はいつ、どのような資料を作成すればよいのかは、過去資料の該当の時期を確認すればよく、資料の探索が容易になる。

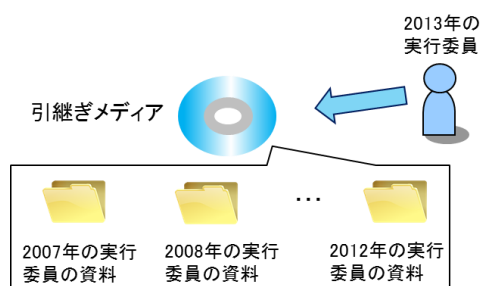


図 8 引継ぎ資料

4 ファイル探索方法と知識の抽出

カレンダー型インタフェースで、蓄積資料をタイムスタンプに基づいて時系列表示することによって、必要なタイミングの資料を探しやすくなることがわかった。しかしながら、どのくらいの稼働をかけたらいいのか、引き継いだ活動の全体像を容易に把握するには至らない。

そこで、新たな提案手法として、該当活動における作成ファイル数の分布の可視化を行った。ファイル数の分布の可視化にあたっては、全体のファイルの、各月の割合を表示する方法と、月ごとのファイル数の最大値に対する割合で表示する方法が考えられる。Appendix Fig.2 は後者で可視化した例である。これにより、作成ファイル数の規模から、最も稼働のかかる時期だけでなく、活動の少ない時期・多い時期などが明確になる。

例えば、事例データは、毎回 7 月にイベントを実

施している。作成資料の規模を見ると、その 1 ヶ月前には多大な準備をしている。3 月にも比較的稼働がかかっていることもわかる。これらのことは、Fig.1 のような数値だけでも把握できないわけではないが、Fig.2 のように可視化することにより、より具体的な把握が可能になる。また、同様の活動をした複数の実行委員の資料を並列して見渡すことにより全体的な傾向を視覚的に把握することもできる。

これに加えて、サブフォルダ名を併記すると、活動内容がより具体的になる(Appendix Fig.3)。ここではファイル名を併記すると文字分量が多くなり混沌としてしまうが、サブフォルダ名程度であれば、適度に資料内容が適度に要約されており活動の全体像を把握するには適している。

以上の方法により、過去の蓄積資料を参考にして活動をおこなうようなシーンでは、提案手法を用いると、同じフェーズで過去の資料を探しやすくなるため、資料探索の糸口となる。また、活動期間中のどの時期に注力すればよいか、どのような順番で活動をおこなえばよいかなど、活動の流れが把握しやすくなる。その結果、蓄積した資料から、活動項目や活動の流れが抽出できることにより、蓄積資料から必要な資料の探索が容易になり、蓄積資料を組織の知識として活用しやすくなる。

5. まとめ

現在の組織活動では、様々な資料を組織活動の記録として残す活動が定着している。そして、長期間に渡って組織活動を行うことによって組織活動の記録が膨大になる。その結果、資料の全体像の把握や必要な資料の探索が困難になり、蓄積された知見が十分に活用できないという問題が発生する。この問題を解決するために、資料作成のタイムスタンプに着目して資料の時系列分布を可視化することで活動の流れを抽出し、資料探索の糸口とする手法を提案した。

参考文献

- [1] 齊藤典明, 金井 敦, 組織知識継承を実現する死蔵されない共有フォルダ構成法, 情報処理学会論文誌, Vol.54 No.1 pp.295-308, (2013).
- [2] 齊藤典明, 金井 敦, 業務の引継ぎを容易にするスケジューラ連動型組織知識継承基盤, 情報処理学会論文誌, Vol.55 No.1 pp.127-142 (2014).
- [3] 齊藤典明, 共有フォルダ活用による組織知識の長期継承, 人工知能学会 第 25 回知識・技術・技能の伝承支援研究会 SIG-KST-025-03, (2015).

