

ファイルの蓄積パターンからの活動手順の抽出

齊藤 典明

定期的に繰り返し実施される活動において、過去に実施された活動の記録を参考に新しい活動の計画を立てることが有効である。しかしながら、過去の記録から活動計画を立てる際に、大量の過去資料を確認するために困難を伴う場合がある。そこで、共有フォルダ内の過去の資料の蓄積状況から、フォルダ構造、ファイルのタイムスタンプ、ファイル量の分析により、活動計画の立案に必要なガントチャートのひな形を自動的に抽出する手法を提案する。

An Extraction Method of Activities Procedure by Using a Shared Folder Structure and a Stored File Attribute

SAITO Noriaki[†]

When we make a new activity plan, a former activities often serves as a reference. Especially, when those are cyclic activities, the effect is large. However, when we make an activity plan by using former records, we faces the raft of trouble by enormous former records. For example, finding a point of activities, a procedure of activity and so on. To solve such problems, I propose a new method which generates a kind of Gantt chart by using number of files, file time stamps and folder structure of the shared folder.

1. はじめに

組織活動では多くの資料が生産される。現在のオフィスワークでは、これらの資料は電子的データであり、使用されたのちに、共有フォルダに蓄積される。蓄積された資料を活用し、次の活動に役立てることで、共有フォルダを組織知識として活用することが可能になる。

一方で組織活動では、長期間活動するにつれてメンバーの交代が必ずおこる。あるいは長期間の活動に限らず、メンバーを交替しつつ活動を継続して実施することがある。このような組織活動では、過去の活動の記録が有益になる場合が多い。しかしながら、過去の記録は多い方が良いが、記録が多くなるにつれて、過去の資料を再利用するための困難さが増加するという特徴もある。

そこで、大量に蓄積された過去の活動資料において、フォルダ構造、ファイルのタイムスタンプ、ファイル量などに基づいて活動の流れを明確にし、活動手順の雛形を生成する手法を考案したので提案する。

2. 組織知識活用における問題

まず、ここで扱う問題について整理する。

2.1 蓄積資料の特徴

現在の組織活動では、様々な資料を記録として残す活動が定着している。議論を進めるにあたって、この蓄積されている資料を「活動記録に関する公式資料」と「現場における資料」とに分類する(表 1)。

前者の「公式資料」は、決裁文書、契約書、開発プロダクト、報告書などであり、文書の種類によっては一定年数

の保存が義務付けられている。これらの資料は、証跡として、過去の活動の検証などに使われる。特に証跡的な意味が強い資料は紙の資料として保存されている。

一方、後者の「現場における資料」は、公式資料の作成にあたって作成した資料や、完成した公式資料のコピーなどを含む日々の活動で発生する様々な資料である。これらは、作成者本人のPCや末端組織の共有フォルダ内に電子データとして保存されている場合などまちまちである。共有フォルダ内に蓄積されている場合は、組織内の多くのメンバーが利活用でき、業務を効率化するために役立つものである。特に、過去の活動を参考に新たな活動をおこなう場合には、蓄積されている過去の資料が有用になる。しかしながら、組織活動が長期間になればなるほど、参考にするべき過去の資料は膨大になる。そのため、資料が多くなるために資料を読み解くための稼働が大きくなるというジレンマが存在する。

表 1 資料の種類

種類	例	保存	再利用
公式資料	決裁文書など	○	×
現場資料	作業ファイル・関連資料・公式資料のコピーなど	△	○

2.2 共有フォルダにおける問題

組織活動が長期間になるにつれて蓄積される資料は膨大になる。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

[†] NTT セキュアプラットフォーム研究所
NTT Secure Platform Lab.

(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、今年の資料を取り出したいときに、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どうかわからない。「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必要な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ことや、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もある。

2.3 組織活動における問題

組織活動では、過去の資料を役立てたほうが効率的な場合がある。しかしながら過去資料が十分に蓄積されていても資料を活用するためにはさらなる問題がある。そこで、過去の資料を活用する「引継ぎ」に着目し、「引継ぎ」という組織活動の中でどのように過去資料を活用できればよいかについて検討を進めた。

まず、2013年1月に実施した「引継ぎ」に関するアンケートについて、再度考察をおこなった[1]。アンケートは約200人の研究所の40人に対して「引継ぎ」における問題点などを記述してもらった。アンケート結果では、「引継ぎ」における問題点の第1位に「暗黙情報の問題」が挙げられており。アンケート回答した40人中33人が指摘していた。今回「暗黙情報の問題」として挙げられた自由記述文をさらに分類し、「資料の把握」「経緯」「背景・意味」「欠落情報」の4つに分類した。「資料の把握」は、「どれが重要かわからない」のような、資料全体を把握することによって判断できそうなもの、「経緯」は、資料の流れを追えば判断できそうなもの、「背景・意味」は、その分野や組織に依存する暗黙知のようなもの、「欠落情報」はそもそも記述され

るべきものが抜けているものとした。アンケート回答文は自由記述文であるため、どちらとも取れる内容もあり、それぞれの分類の境界は曖昧である。この4つのカテゴリを、蓄積資料を時間かけて読み解けば把握できそうか、そうではないかで、並べてみた(図1)。図を見ると、半数以上の問題点は時間をかけて資料を読み解けば「暗黙情報」について得られると考えられる。よって、大量の蓄積資料について、資料の全体像や時系列を把握させる手法があると、蓄積資料に対する理解が深まると考えた。

次に、業務を引継いだ際の経験的な知見として、大量の資料を引き継ぎ新しい活動計画を立案する際、まず必要となるアクションは、資料の全体像を確認すること、引き継いだ業務においてまず実施すること、いつまでのどのようなことをすれば良いのかという作業順序と作業量の見積もり、などを把握し、ガントチャートなど「活動の流れ」が把握できる方法でまとめてゆく。

以上のことから、大量の過去資料を活用して、新しい活動の計画を立案するには、意に反して過去資料を読み解くのに多大な労力を要するという問題がある。ここでは、この問題を解決するには蓄積資料から「活動の流れ」を把握できる手法があると効果があると考えた。

「引継ぎにおける問題点」のアンケートにおいて
40人中33人が「暗黙情報の問題」を指摘。
「暗黙情報」について細分化を行った。



図1 暗黙情報の内訳

3. 提案アプローチ

次に、問題を解決するためのアプローチについて述べる。ここでは、これまで共有フォルダから蓄積資料を効果的に取り出せる手法として、カレンダー型インタフェースに「活動の流れ」を抽出する方法を検討したので、これについて述べる。

3.1 カレンダー型インタフェース

まず、はじめにカレンダー型インタフェースについて述べる。これまでの研究では、組織内に蓄積される資料の多くはイベントに紐づくものであることが分かっている。また、組織内のイベントはカレンダー形式のスケジューラで管理することが定着している。そこで、この2点に着目し、カレンダー形式のインタフェースで共有フォルダを利用する手法を提案した[2]。

対象とする共有フォルダの領域には、単一の活動に対する資料が蓄積されており、蓄積資料が比較的少ない状況で有効性が確認できた。しかしながら、実組織における共有フォルダ内では、共有フォルダ内には様々な活動の資料が

混在し、共有フォルダの運用期間が長期間になれば蓄積資料も大規模になる。この場合、カレンダー型インタフェースでは資料と日時の紐づけ、資料が大量になることにより探しにくくなることへの改善が必要であることが判明した。

これらの問題を解決するために、(a)既存のファイルの種類を限定してファイルのタイムスタンプを取得すること、(b)フォルダ名の命名方法の特徴から活動項目を抽出する方法、(c)ファイル名ではなくファイルの存在を表示したあとにファイル名を表示する方法を導入した。

提案手法では、対象フォルダに対してクローリングを行い(図 2-①)、ドキュメント系ファイルに限定してファイルの蓄積場所とファイルの最終更新日時のタイムスタンプを取得する。その後、ファイルの蓄積場所であるフォルダ名から活動項目を抽出する。活動項目の単純な抽出方法はフォルダ名をそのまま用いる方法であり、発展的な方法ではフォルダ名の命名規則に着目して代表的な単語のみを抽出する方法などがある。活動項目を抽出したのちに、活動項目と時間軸でカレンダー形式の表示領域(図 3)の中に、活動項目と年月ごとのファイル数、ファイル名、共有フォルダ内のファイルへのハイパーリンクを HTML ファイル内にマッピングしたカレンダー型インタフェースを出力する(図 2-②)。このカレンダー型インタフェースを操作することによって共有フォルダ内のファイルへのアクセスを行う(図 4)。操作方法は、カレンダー形式の Web 画面において、時期(図 4-①)と、フォルダ名から生成したカテゴリ(図 4-②)から、必要な時期の必要なカテゴリのファイルの有無(図 4-③)を確認し、詳細化によりファイル一覧(図 4-④)を取得する。ファイル一覧から必要なファイルをクリックするとハイパーリンクおよびシンボリックリンクにより共有フォルダ内のファイルを取得する(図 4-⑤)。

このインタフェースにより、共有フォルダ内の資料を活動項目と時系列に従って容易に取り出すことが可能になる。

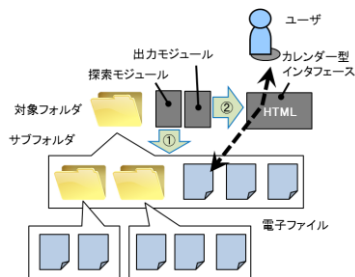


図 2. アプローチ概要

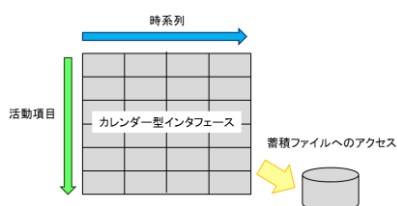


図 3 カレンダー型インタフェース

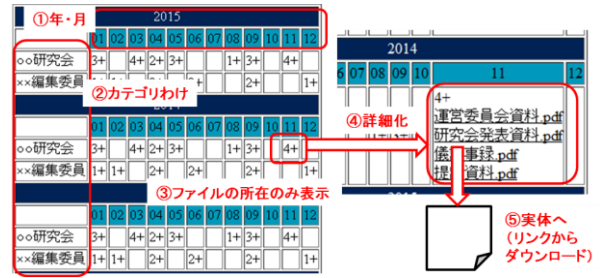


図 4 操作イメージ

3.2 カレンダー型インタフェースによるアプローチ

次に、検討の前提となる活動や共有フォルダの使い方を整理する。組織活動には様々なタイプの活動があるが、資料を引き継ぐという観点から、引継ぎの発生する場合わけとして、以下の2種類を考えた。

- (i) 役職として引き継いだ場合
- (ii) タスクを完遂することが目標の活動

前者の役職として引き継いだ場合の例は、部長、課長のような役職のほか、学会の委員のような組織構成員の活動も含まれる。ここでは、組織運営に必要な様々な活動をこなしてゆく。また、様々な活動が含まれており蓄積した資料の分類も複雑になる。

後者の場合は、特定の目的を達成するためのプロジェクトのような組織活動を想定している。後者は、活動の範囲を明確にして引き継ぐため、蓄積資料はすべて同じ活動として分類することができる。

検討を進めるにあたって、前者の場合は、様々な活動が混じっており、「活動の流れ」を明確するという目標に対して、事前に「特定の活動」を抽出する技術が必要になる。この技術はまだ確立できていないことから、ここではまず、後者の場合を対象とし、その中で「活動の流れ」を明確にする手法を確立することとした。

この後者の場合の共有フォルダのモデルケースを図 5 に示す。ここでは、活動範囲が特定された共有フォルダがあり、第一階層に時系列のフォルダ名、第二階層以下に活動上のサブタスクとなるフォルダ名、第三階層以下に資料となるファイルが蓄積されているものとする。ファイルの蓄積者は蓄積ファイルを表現する適切な名称をフォルダ名として付けていると仮定する。ファイルの蓄積者により自由に命名したフォルダ名は、これまでの検討の中においてもファイル探索のために有用であることが分かっているため、これらをあてにした[3]。反対に、機械的に付与されるようなフォルダ名の場合は検討の対象外とする。

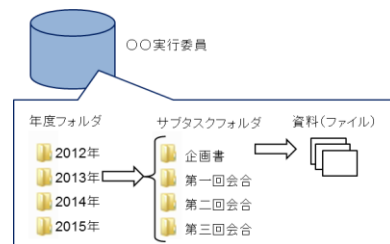


図 5 共有フォルダ構造 (例)

3.3 カレンダー型インタフェースの改良

次に、カレンダー型インタフェースにおいて「活動の流れ」を抽出する方法について述べる。

図5のような比較的管理の良い共有フォルダであっても、OSのフォルダ機能を使ってファイルを探すには複数のフォルダを行き来する必要がある。そのためファイルを探すためのユーザ負担がある。このようなユーザ負担を軽減するためにカレンダー型インタフェースを導入すると図6のようなインタフェースでファイルを探すことができる。階層的なフォルダ名と、ファイルの作成月の領域(図6-A)に、ファイルの存在が表示される(図6-B)。必要な時期に必要なカテゴリのファイルについて詳細化(図6-C)し、必要なファイルをダウンロード(図6-D)して使用する。このインタフェースにおいて、一定のスパんで繰り返される活動の場合は、活動のスパンを重ねたカレンダー表示をすることにより、複数回の活動を比較することもできる。これにより、過去の資料を探すことが容易になる[2]。

しかしながら、新たな活動計画を立てる際には、再利用できる資料を探し出すだけでなく、作業全体のどこにクリティカルパスがあるのかなど「活動の流れ」を理解する必要がある。そこで、資料を探すだけでなく、ガントチャート型のインタフェースで蓄積資料を取り出せる手法を導入した。このような方法により「資料の利用」と「活動の流れ」を可視化が実現される。ここで「活動の流れ」の可視化の手順は次のとおりである：

- (1) サブタスクを表すフォルダ名を一連の活動におけるサブタスク名とする(図5における第二階層)。
- (2) ドキュメント系ファイルに限定して、各サブフォルダ内ファイルのタイムスタンプとファイル数を計数する。
- (3) サブタスク内の最新のファイルのタイムスタンプをサブタスクの日付とし、サブタスクを日付順に作業工程として並べる。
- (4) 前の作業工程の翌日から、作業工程の日付までを作業期間と見積もる。
- (5) サブタスク内のファイル数を作業規模と見積もる。

以上のようにして、対象とする共有フォルダ内のファイルから「活動の流れ」をガントチャート型で表示したインタフェースを図7に示す。

ガントチャート型インタフェースでは、カレンダー型インタフェースと同様に時系列とカテゴリの領域(図7-A)が表示される。なお、時系列は日付単位であるが、図7では画面の都合で時系列の目盛は省略してある。ファイルの有無は、ファイル作成の時期とファイル量が矢印の大きさ・長さで表示され、カテゴリ名横の記号(図7-B)をクリックすると該当のファイルの一覧が表示される(図7-C)。ファイルのリストから必要なファイルをダウンロードして使用できる(図7-D)。このようにして、各作業工程から対象のファイルを参照することも容易にできる。また、

本方式は Web 形式のガントチャートであり編集できないことや、参考程度のガントチャートであることから、活動計画として編集するためのガントチャートは別途作成する必要がある。そこで、表示している Web 形式のガントチャートを参考に Excel ベースでガントチャートを作成できるように、表示しているガントチャートを CSV 形式で出力できるようにした。画面上のダウンロードボタンをクリックすると(図7-E)、CSV 形式のガントチャートのひな形をダウンロードできる(図7-F)。このガントチャートのひな形を使うことで効率的に作業計画を設計できるようになる。



図6 カレンダー型インタフェース画面イメージ

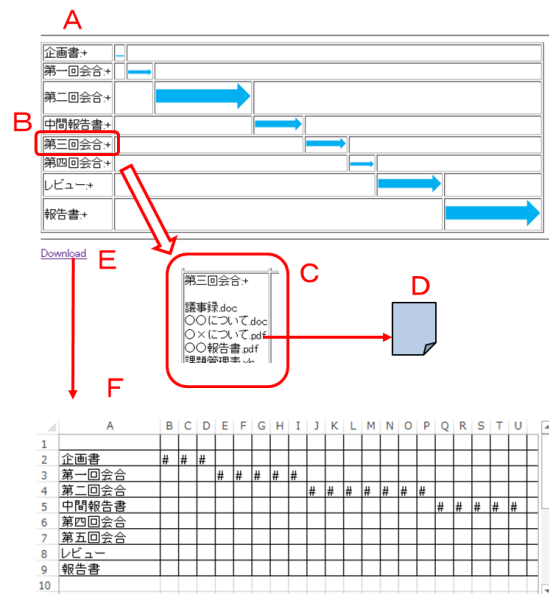


図7 ガントチャート生成イメージ

3.4 実データでの実施例

実際の実施例での画面イメージを Appendix に示す。ここの例は、過去資料をメディアに蓄積し、長年引き継がれてきたデータとして情報処理学会シンポジウムの実行委員会のデータである(図8-A)。データ量は6年分で4.3GB、6,206ファイル、549フォルダである。担当年のシンポジウム開催の計画を立てるために、過去資料を確認するには多くの時間と労力を必要とする。特に OS のフォルダの機能

によって確認するには上位のフォルダから順番に開き、中を確認してゆく必要がある(図 8-B)。フォルダ名から順番などを類推することは可能であるが資料全体を確認すること、複数回の実施記録の違いを見つけること、はじめに実施すること、作業の順番を見つけること、クリティカルパスを見つけること、など様々なポイントを確認するには多くの作業を伴う。

これに対して、カレンダー型インタフェースであれば、フォルダ横断的にファイルの存在を時系列で確認することができ(図 8-C)、確認するためのユーザ操作を軽減できる。カレンダー型インタフェースで複数年を単純表示した結果を Appendix A に示す。紙面の都合で各年のフォルダ階層まで開いた図は省略する。このようにカレンダー型インタフェースにより、ファイルの全体像や時系列が把握できるため、過去資料の活用を効率化できる。しかしながら、ファイルの順番はわかるもの「活動の流れ」まではわからないために、逐次フォルダ名やファイル名を確認する必要があり、過去資料から次の計画を立てるまでには困難さが依然と残っている、という課題があることもわかっている。

これに対して、これまでのフォルダ名の検討の中で：

- ・サブフォルダ内のファイルのタイムスタンプの分布は広くない。(言い換えると、タスクの実施時期ごとにフォルダが作られている。)

- ・フォルダ名はファイルの内容を的確に表していることが期待できる。

ということがわかっている。

これらのことから、フォルダ名をファイルグループの要約キーワードとして活用できると考えられる。また、日付や記号などの情報を除外したフォルダ名はタスク名として活用できる。さらに、フォルダごとのファイル量は該当のタスクの作業量として近似することができると考えられる。そこで、カレンダー型インタフェースのファイルの分布に、タスク名としてフォルダ名を併記したものを Appendix B に示す。

図では、フォルダ名に付属していた日付の情報を削除し、蓄積ファイルを表す単語を抽出した後にカレンダー型インタフェースに併記した。これによりファイルを確認しなくてもおおよその活動の流れが把握できるようになる。この図をガントチャート型に加工した例が Appendix C である。

図は、紙面の都合で日付のスケールは省略してある。Web の画面上で確認する場合は、正確な日数のスケールよりも日数のスケールを省略するほうが、見通しが良くなる。Appendix C におけるタスク名を時系列に並べ、作業工程とした。次に、前工程となるタスクの最新タイムスタンプの翌日から、タスクの最新タイムスタンプまでを作業期間として矢印で示した。また、タスク名に対応するフォルダ内に含まれるファイル数を作業量として矢印の大きさで示した。このようにして、作業の順番や各作業の期間、各作業の作業量を視覚的に把握できるようになる。

なお、機械的に抽出したガントチャートであるので、そのまま使えるとは限らないため、表示されたガントチャートを参考に、CSV 形式のファイルを取り出し、的確な WBS の作成や作業計画を作成することを前提とする。特に、今回の方式では、複数のタスクを並列して実施するガントチャートは描けないため、より実用的な活動計画を作成するには手動による修正が必要である。

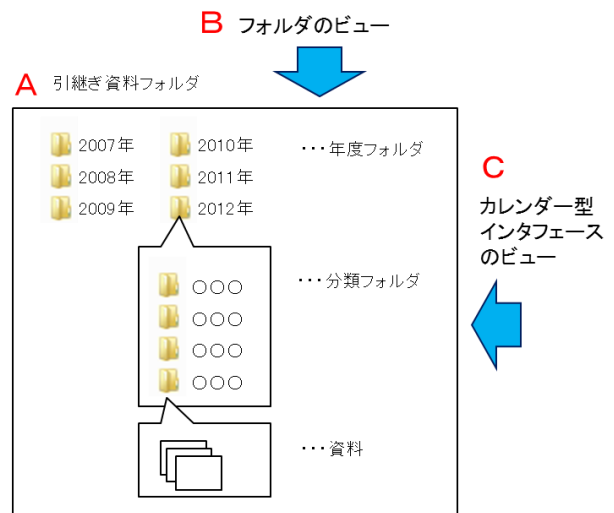


図 8 共有フォルダの構造

4. おわりに

定期的に繰り返し実施される活動において、過去に実施された活動の記録を参考に新しい活動の計画を立てることが有効である。しかしながら、過去の記録から活動計画を立てる際に、大量の過去資料を確認するために困難を伴う場合がある。そこで、共有フォルダ内の過去の資料の蓄積状況から、フォルダ構造、ファイルのタイムスタンプ、ファイル量の分析により、活動計画の立案に必要なガントチャートのひな形を自動的に抽出する手法を提案した。

提案手法について、実データで試用した結果、定性的に妥当なガントチャートが出力されたため、良好な結果が得られたと判断した。

今後の課題として、生成されるガントチャートの妥当性や、計画立案における作業効率の効果について評価を行う。また、並列作業などを考慮した「活動の流れ」を抽出できる手法についても検討を行う。

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.5 No.1 pp.127-142, (2014).
- 3) 齊藤典明: 混沌フォルダからの組織知識の抽出手法の提案, 情報処理学会研究報告, Vol.2015-GN-94 No.12, (2015).

