

混沌フォルダからの組織知識の抽出手法の提案

斉藤 典明

組織内で長年運用された共有フォルダでは、ファイルやフォルダが大量になり混沌としてしまう場合がある。混沌とした共有フォルダに対して、フォルダ名の命名方法から資料を指定する適切なカテゴリと、ファイルのタイムスタンプから時間情報を抽出した。これらの情報を用いて共有フォルダ内の資料を整理と利用できるカレンダー型のインタフェースの実現により、組織知識の抽出をおこなう手法を提案する。

An Organizational Knowledge Extraction Method from Chaotic Shared Folders

SAITO Noriaki[†]

Shared folders tend to become chaotic while it has been using it for a long term in the organization. To solve such problem, I propose an organizational knowledge extraction method by using calendar type interface which is realized by using extracted category from sub folder's naming rules and timestamps of files. As a result, chaotic shared folders will be improved to well-formed shared folders.

1. はじめに

組織活動を行う上で組織内に蓄積されている知識を活用することは重要なことである。組織における知識とは、教科書やマニュアルに記載されているような定式化されているものよりも、記載されている内容をどのように活用すればよいのかという、実践的な知識である。例えば、ソフトウェアの開発標準があったとして、それをそのまま推進するだけでは良いソフトウェアは開発できない。開発標準で規定されている範囲で、その時々状況に応じて柔軟に対処する必要がある。このような組織内の実践的知識は、組織内でトップダウン的に広まるものではなく、むしろボトムアップ的に広がっている。

組織活動を長期に渡って持続してゆくには、このような実践的な知識を捉えて確実に組織内で継承してゆく必要がある。このような組織の知識の蓄積と継承は、組織メンバーが安定しており、定期的にメンバーが出入りする場合には、その時々組織メンバーから口伝で継承してゆくことも可能である。しかしながら、現在は協働環境が多様化しているため必ずしも口伝では継承できない。そこで、様々な資料が電子ファイルとなっている現在、情報ネットワークを用いることで、組織の知識を蓄積し継承してゆくことで効果を挙げられると考えられる。

組織活動の中で生成される電子ファイルは、電子メールや共有フォルダで共有される。電子メールで共有された資料は配布された各自の所蔵になるが、共有フォルダの場合は後々のメンバーでも参照可能で、再利用性の高い電子ファイルとすることができる。そこで、共有フォルダは多くの組織で定着していることから、共有フォルダを有効活用することで組織の知識の蓄積と継承を支援する方法を検討してきた[1]。検討にあたって、NTT 研究所内のいくつかの研究チームの共有フォルダを確認

した結果、組織内で作成された電子資料が後々まで活用しやすい形で使われている共有フォルダと、蓄積されている資料が属人化してしまっている共有フォルダがあった。前者と後者の場合は共有フォルダの構造に違いがあることがわかった。前者の場合は、第一階層が時系列で管理されており、後者の場合は第一階層が案件名やカテゴリ名で管理されている実態があった。そして、長期的な組織活動では第一階層が時系列で管理の方が合理的な構造であることを突き止めた。

一方で、世の中の共有フォルダの推奨構造を調査した結果、長期的な組織活動の支援目的のものはなかった。そのため、多くの組織においては、長期的な組織活動の中で共有フォルダが混沌としてしまう可能性がある。そこで、本研究では、このような長期間の運用で混沌としてしまう共有フォルダに対して、長期間の運用でも共有フォルダ内資料の活用を容易にできるインタフェースを提案する。

2. 一般的な共有フォルダの整理方法

これまでの検討の中で提案する長期的な組織活動支援を目的としたフォルダ構造は、図 1(A)に示すような第一階層が時系列であり、その下に何らかの分類がある構造である。これに対して、長期的な活動の中で共有フォルダが混沌としてしまう構造は図 1(B)に示すような、第一階層が何らかの分類であり、その下に時系列のフォルダがある場合であった。混沌とした共有フォルダとは、どこにどのようなファイルがあるのかは、ファイルを蓄積した人だけがわかる、蓄積の属人化の状態になることをさす。この場合、ファイルの蓄積は継続的に行われるものの、蓄積資料の活用は限定的になる。具体的には、資料を蓄積した人自身による資料の再利用か、特定資料の受け渡しという利用に限定されがちである。共有フォルダを有効活用すべき自主

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

的な過去資料の活用には膨大な労力を要してしまう状況である。

研究所内の観察で得られた知見が世の中の知見と比べてどのような差異があるかを確認するために、一般的な共有フォルダの整理方法にはどのようなものがあるかを調査した。学術的にまとめられた文献は見つからなかったため、検索エンジンで得られた結果について分析した。Googleで「共有フォルダ」と「整理方法」または「整理」で検索し、個別事象、製品マニュアル、製品紹介、QAの記事、文章として不十分なもの、共有フォルダを対象としていないものを除外し、一般的化した共有フォルダにおける整理方法を書いた記事10件を抽出した[2-11]。この10件に対して、①整理手法の目的、②構造的な類似性、③長期保存の観点で評価した。なお、英語による検索「shared folder」と「structure」などの検索においても学術的な文献は見つからなかった。

評価結果は、①整理手法の目的は、今回の抽出基準なのでいずれの手法も適合している。調査した中で、異なる整理目的にはディスクの有効活用という目的などがあつた。②構造的類似性は、第一階層が時系列かどうかで判断した。10件のうち1件だけ、第一階層を時系列にするものがあつた。③長期保存の観点は、多くの手法では共有フォルダの長期的な運用については特に述べられていなかった。長期的な運用を考慮している場合は、定期的に不要ファイルを消去することを推奨しているため、長期保存の観点ではないと判断した。その結果、一般的な整理手法では長期保存に適しているものがないことが判明した。

つまり、一般的な多くの整理手法で共有フォルダを運用していると、長期的な運用の間に混沌としてしまう、と結論できる。よって、長期運用により混沌とした共有フォルダを、ファイルの消去ではなく、ファイルを保存する観点で整理できる手法の必要性があると考えられる。



図 1 共有フォルダ構造

表 1 共有フォルダ整理法

出典	概要	①	②	③
[2]	・第一階層はあらかじめ決めた分類を用意 ・第二階層以降のフォルダ名に通番と日付 ・ファイルの受け渡しはフルパスで行う	○	×	—
[3]	・第一階層はプロジェクトごとに構成する ・第二階層は時系列で構成する ・フォルダは細かく分類しない	○	×	—
[4]	・浅い階層にファイルを置かない ・第一階層は重複しないジャンル分け ・フォルダの名称は短くする ・第二階層のフォルダの頭は二桁数字 ・旧バージョンのファイルを入れるフォルダ ・ファイルの頭に日付を入れる ・よく使うフォルダはショートカットを作る	○	×	—
[5]	・第一階層はプロジェクト単位 ・ファイル名に種別・分類・日付・通番	○	×	—
[6]	・2階層目までフォルダ構成を台帳管理 ・定期的にフォルダ構成を棚卸しする ・2階層目まではフォルダのみとする ・フォルダ名の前を2桁連番 ・外部フォルダへはショートカット ・「ごみ箱」フォルダを作成	○	×	×
[7]	・第一階層を年代、第二階層をカテゴリ分類とし、3階層程度に抑える ・ファイル名は日付や案件名など ・同じファイルを2か所に置かない ・ファイルのバージョンを明記	○	○	×
[8]	・保存するファイルは少なくする ・ファイルコピーはしない ・フォルダの分類方法を決めて徹底する ・フォルダ階層はなるべく浅くする ・わかりやすい名前	○	×	—
[9]	・フォルダ階層は3階層以下にする ・フォルダに連番を付け管理する ・空(カラ)のフォルダに目印 ・ファイル名に年月日を入れる	○	×	—
[10]	・ファイル名は「日付+種類」 ・第一階層は用件ごと、第二階層は期間や繰り返しごと ・年1回はバックアップをとる、ファイルの保存期間も決めて古いものは削除	○	×	×
[11]	・組織構成の変更に対応できる構成＝第一階層は組織構成で第二階層は年度 ・一定期間たったフォルダは消去 ・内部統制を意識してアクセス権管理 ・個人フォルダを作らない	○	×	×
図 1-A	・第一階層を年度、第二階層をカテゴリ	○	○	○

- ①整理手法の目的：
 手法は共有フォルダ内の資料を探しやすくする目的か？
 (○は Yes、×は No)
- ②構造的な類似性：
 手法は第一階層が年度単位か？(時間軸か？)
 (○は Yes、×は No)
- ③長期保存の観点：
 手法は長期保存ファイルの活用を意識しているか？
 (○は Yes、×は整理を推奨している、-は記述なし)

3. 混沌フォルダへのアプローチ

一般的に述べられている共有フォルダの整理方法では、長期間の運用により混沌としてしまうことがわかった。そこで、長期間の運用により混沌としてしまった共有フォルダを活性化させる方法を検討する。

混沌フォルダの自主的な活用を促進するためには、使いやすさという観点のほかに、利用の定着という利用者や組織の文化や習慣的・規則な要素もある。後者は、長期間の運用により混沌となったことを前提とするため、共有フォルダへの資料の蓄積は定着していることを前提とする。そのため、今回の評価対象外として、共有フォルダ利用の構造的な観点から蓄積資料を活用する際の使いやすさ改善するアプローチをとる。

これまでの検討において、組織内に蓄積された資料を有効活用するための仕組みとして、資料を組織内のイベントと日も付けて探し出す方法をこれまで提案した[12]。その実施例の一つとして共有フォルダをカレンダー型のインタフェースで利用する方法がある。今回、カレンダー型インタフェースを活用することとした。

3.1 カレンダー型インタフェースの概要

組織知識の継承にあたってこれまでの検討では、組織内の継承すべき資料は組織のイベントに基づいたものが中心であり、組織のイベントはスケジューラの中で管理されている。そこで、これまでの検討ではカレンダー型のインタフェースを有するスケジューラと共有フォルダを連動する方式を提案した。ここでの検討は、ある限られた活動においてカレンダー型インタフェースで資料を探せるようにすることによって、活動の流れが把握でき、ユーザの負担を軽減できる効果があることを確認した(図2)。

図2は、ある活動で活用された資料を複数年表示のカレンダー型インタフェースで表示した例である。事例とした活動は例年の活動がほぼ繰り返されているが、毎年、少しずつ差異がある。このような活動で用いられた資料を時系列となるカレンダー型インタフェースでアクセスできるようにした。これにより、いつごろ、どんな資料があったか、また複数年の資料を比較することによりそれぞれの年の活動の特徴が把握できる効果がある。

先の検討では1つの活動範囲に対して実施し有効性を

検証した。一般に共有フォルダには複数の活動範囲が混在して資料が蓄積されている。このような状況で、共有フォルダ内のファイルを単純にカレンダー型インタフェースで表示するとこの活動の流れは埋もれてしまいわからなくなる(図3)。図3は、年間100個の資料を数年分表示した例である。共有フォルダ内の資料をカレンダー型インタフェースで効果的に利用するためには、複数の活動範囲を分離して提示する必要がある。

このような大量表示に伴う問題点を解決するために、2つの方針を実施した。一つ目は、カレンダー型インタフェースの初期画面ではファイル名は表示せず、ファイルの所在のみを表示する方法とした(図4)。該当の年月にファイルがある場合はファイル数と「+」の記号を表示し、実際のファイルへのアクセスは「+」をクリックして該当の領域のファイルリストを表示しアクセスする方式とした。二つ目は、活動範囲をカテゴリとして、カテゴリに分けてカレンダー型インタフェースで表示することとした(図5)。



図 2. カレンダー型インタフェースの表示例



図 3. 大量表示例



図 4 ファイル表示の改善

カテゴリわけ		2012											
		01	02	03	04	05	06	07	08	09	10	11	12
○○研究会		3+	4+	2+	3+			1+	3+	4+			
××編集委員		1+	1+	2+	2+				2+				1+
		2013											
		01	02	03	04	05	06	07	08	09	10	11	12
○○研究会		3+	4+	2+	3+			1+	3+	4+			
××編集委員		1+	1+	2+	2+				2+				1+
		2014											
		01	02	03	04	05	06	07	08	09	10	11	12
○○研究会		3+	4+	2+	3+			1+	3+	4+			
××編集委員		1+	1+	2+	2+				2+				1+

図 5. カテゴリわけによる改善

3.2 混沌フォルダへの適用

長期間の運用により大量の資料を蓄積した共有フォルダをカレンダー型インタフェースで表示するには、活動範囲となるカテゴリと、資料の時期を特定する必要がある。しかしながら、資料とイベントとの紐付けが失われているだけでなく、対象となる資料が大量である。そのためカテゴリや時期との紐付けを再現するために労力を使うことは現実的ではない。そこで、同じフォルダのファイルは同じ活動によって生成されていると考えられるためフォルダ名をカテゴリとし、ファイルの最終更新日時をタイムスタンプを時期とし、これらを共有フォルダへのクローリングにより取得する方式を実現した(図6)。

具体化にあたって、実際の混沌とした共有フォルダを例題に実施した。例題の共有フォルダは第一階層が整理番号付きでカテゴリ分類されている、第二階層以下は必要に応じて自由に命名してフォルダを作成している。組織メンバーのそれぞれの利用者が必要に応じてサブフォルダを作成しているため、フォルダ名もサブフォルダの深さもまちまちである。共有フォルダは資料の蓄積場所や資料の受け渡し場所としては組織内で利用が定着している。利用期間は10年以上であるが、組織メンバーは10年を超えるものは2名程度である。組織メンバーは利用期間を通して10人から20人の間で推移してきている。組織内は、共有フォルダの利用よりも所属期間が短いメンバーが中心であるため、各自が蓄積した資料の所在はわかるが、他のメンバーによる資料の蓄積は把握できない、蓄積資料の属人化が起きている。2015年1月時点での共有フォルダ内の資料の蓄積状況を表2に示す。このような共有フォルダに対して単純にクローリングした結果が表3の上段になる。

タイムスタンプを確認すると、組織の存続期間に対して明らかに不自然な日付のファイルが出力された。内容を確認したところ、インターネットなどからプログラムのソースコードをダウンロードして展開したものであった。このような、共有フォルダ内には組織活動の資料とは言えないものも多数あることがわかった。ほかに、プログラムが出力するような隠しファイルなどもある。そこで、組織知識

の共有と継承ということが主目的であることから、組織活動で生成された資料に絞ってクローリングをする方式とした。その結果、例題の共有フォルダでのクローリング結果は、ファイル数は半数以下になり、ファイルのタイムスタンプについても妥当な期間となった(表3下段)。

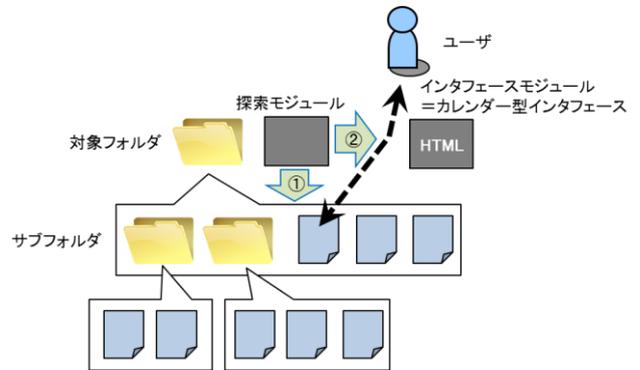


図 6. アプローチ概要

表 2. 例題とした共有フォルダ

データ量	ファイル数	フォルダ数	第一階層フォルダ数	フォルダ階層
102 (GB)	46, 973 個	8, 885 個	49 個	16 階層

表 3. 共有フォルダへのクローリング結果

対象	ファイル数	開始年	最大階層
全てを対象	46, 973 個	1980 年～	16 階層
ドキュメントのみ	21, 341 個	1997 年～	14 階層

表 4. クローリング対象ファイル

種別	概要	対象
ドキュメントファイル	組織活動で生成される資料のファイル	○
データファイル	画像ファイルなど	△
プログラムファイル	exe ファイル、スクリプトファイルなど	×
隠しファイル等	機械的に出力されるファイルや、プログラムが使用するファイル (ini ファイルなど)	×
ダウンロードファイル	外部で作成されたファイルをネットワーク経由などで取得した場合	△
展開ファイル	アーカイブ形式で取得し、その後展開したファイル (ソフトウェアのソースコードなど)	×

○ : 対象とする

× : 対象としない

△ : 条件によっては対象とする (今回の例では対象外)

3.3 カテゴリ抽出手法

共有フォルダへのクローリングにより、共有フォルダ内

のファイルとファイルのパスおよびフォルダ構造、最終更新日時のタイムスタンプのリストを生成することができる。このリストを、単純にカレンダー型インタフェースで表示する場合、活動単位を示すカテゴリが大量になるという問題がある。ファイルが大量になると活動単位のリストが多くなり、かつ文字数も多くなり判別が難しくなりユーザの視覚の負荷は高くなる。例題の共有フォルダでは、ファイル階層は最大 14 階層あるので、最大 14 階層のフォルダ名がすべてつながったものになる(図 7)。

そこで、本検討では、この活動単位であるカテゴリのリストを効果的に絞り込む戦略を検討した。カテゴリを絞り込む方法として、まずは共有フォルダにおける各サブフォルダの命名方法に着目した。例題とした共有フォルダのうち対象となる資料の蓄積されているフォルダ名 4,193 個と、表 1 に示した一般的なフォルダ整理方法を調査し、フォルダ名について、いくつかの分類ができることがわかった。フォルダ名には時系列をあらわす単語、分類を表す単語、内容・イベントを表す単語、注釈となる単語に分類できる(表 5)。

このうち、カレンダー型インタフェースによる表示を前提にすると、そもそもカレンダー型表示が時系列であることからフォルダ名としては時系列の単語は不要となるため除外できる。次に、情報の利用と蓄積では情報整理の観点の異なり、情報を蓄積する際の規則となる整理方法については情報活用の観点ではノイズとなるため除外できる。注釈についても情報の内容についての言及ではないので除外できる。

このようにして、当初のフォルダ階層から起こしたカテゴリ数よりも少ないカテゴリ数でファイルへのアクセスができるようになる。しかしながら、この状態においても共有フォルダへアクセスするためのカテゴリ数としてはまだ多い。例題の共有フォルダにおいては 390 個であった。そこで、抽出されたキーワードが共有フォルダ内のファイルを指定できる量に基づいて抽出することとした。ひとつのキーワードが非常に多くの電子ファイルを指定している場合は、キーワードとしては十分にブレイクダウンしていないと判断し、反対に指定できる電子ファイルがあまりにも少ない場合は、個々のファイルを指定するのと同様であるため、ファイルのグルーピングとしては妥当でないと判断した。例題の共有フォルダにおいては 1 つのキーワードが 20% 以上のファイルを指定する場合と、1% 未満のファイルしか指定しない場合を除外した。その結果、60 個のキーワードが抽出された。

これまでの手法で、当初のフォルダ階層から起こしたカテゴリ数よりも少ないカテゴリ数でファイルへのアクセスが可能になる。しかしながら、カテゴリ数が少ないと、すべてのファイルを指定できているとは限らない。そこで、どのカテゴリにも属さないファイルについては分類上「その他」に入れることとした。多くのファイルが「その他」

に分類されることは好ましくないため、抽出されたキーワードについて全体のバランスをみて調整する必要はある。また、機械的に抽出されたキーワードであるため、意味不明語や記号、不適切な単語が含まれている場合があるため、これらは修正する必要がある。

例題の共有フォルダについては、抽出されたキーワードの修正は 4 箇所、「その他」に分類されたファイルは全体の 12% であったため大きなキーワードの修正は必要なかった。以上の方法で実施した表示例を図 9 に示す。



図 7. カレンダー型インタフェースでの単純表示例

表 5. フォルダ名の分類方法

命名方法	具体例
時系列	年、日付、時期、順番 など
分類方法	組織名、ロケーション名、担当者名、整理方法、整理上の記号 など
内容・イベント	資料内容の性質や使われた活動を名称にしたもの
注釈	資料の扱い方など (期限など)

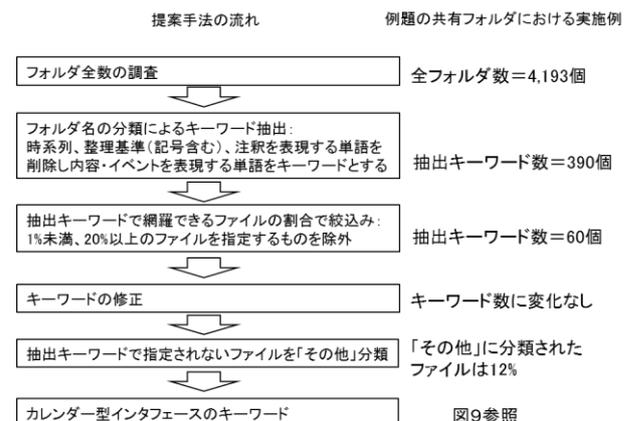


図 8. 提案手法の流れと実施結果

図 9.提案手法による表示例

4. 考察

混沌フォルダに蓄積されている資料を有効活用するために、共有フォルダ内にどのような資料が入っているかを提示し、利活用を容易にできるインタフェースを実現した。従来形式の共有フォルダを利用した場合、カレンダー型インタフェースを単純適用した場合のそれぞれと比較して提案方式の効果について考察する。なお、比較にあたっては個々の表示方法ではなく構造的な違いについて議論するものとする。

4.1 共有フォルダとの差異

第一階層が時系列になっていない共有フォルダは、長期間の利用により混沌としてします。このような共有フォルダから該当の資料を探そうとすると、情報提供時の整理体系と情報活用時の整理体系の違いから目的のファイルが探しにくい。提案方式では、情報の蓄積は組織内で定着した共有フォルダで行い、情報の活用は、情報時期と内容によって取り出せるインタフェースからおこなう。そのため、情報活用に適した構造を作り出すことが可能になった。

4.2 単純適用カレンダー型インタフェースとの差異

共有フォルダのクローリングによって得られたフォルダの階層構造とタイムスタンプの情報を用いて単純にカレンダー型インタフェースで表示すると、カテゴリの数が膨大になるだけでなく、個々のカテゴリ名も膨大になる。これに対して提案手法では、カテゴリ名を効果の高い単語に限定した。少ない単語で多くのファイルを指定することができるようになった。例題の共有フォルダでは、5,075 個のカテゴリ分類になるのに対して、提案方式では 61 個のカテゴリに分類される。カテゴリ名称についても、単純適用ではフォルダの階層構造すべての文字列になることに対して、提案方式は 1 単語程度の短いものになっている。これらの

ことから可読性が大幅に向上する。

5. おわりに

組織内で長年運用された共有フォルダでは、ファイルやフォルダが大量になり混沌としてしまう場合がある。混沌とした共有フォルダに対して、フォルダ名の命名方法から資料を指定する適切なカテゴリと、ファイルのタイムスタンプから時間情報を抽出した。これらの情報を用いて共有フォルダ内の資料を整理と利用できるカレンダー型のインタフェースの実現と、実際の混沌フォルダを例題に動作確認をおこなった。その結果、組織の知識として共有フォルダに蓄積されている資料を的確に把握できるようになった。これにより組織知識の活用が活発になることが期待できる。

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)もう探さない!迷わない!ファイル&フォルダ管理法(オンライン),入手先 <http://webnaut.jp/direction/568.html>(参照 2015-2-4).
- 3)3つの観点でファイルを整理する管理法(オンライン),入手先 <http://bizmakoto.jp/bizid/articles/0805/21/news023.html>(参照 2015-2-4).
- 4)乱雑なパソコンのフォルダを綺麗に整理する7つのファイル管理術(オンライン),<http://bamka.info/folder-management>(参照 2015-2-4).
- 5)共有ファイル管理の悩み(オンライン),入手先 <http://www.ys-consulting.com.tw/news/22417.html>(参照 2015-2-4).
- 6)荒れ果てた共有フォルダの構成を整理して綺麗に保つ6つの工夫(オンライン),入手先 <http://jmatsuzaki.com/archives/3248>(参照 2015-2-4).
- 7)仕事整理術、メール・ファイル編(オンライン),入手先 <http://wol.nikkeibp.co.jp/article/special/20101202/109456/>(参照 2015-2-4).
- 8)共有フォルダを汚くしないためにどうしたらよいか(オンライン),入手先 <http://ofsilvers.hatenablog.com/entry/cleanup-rules>(参照 2015-2-4).
- 9)情報整理のヒント(オンライン),入手先 <http://selfde.biz/e064.html>(参照 2015-2-4).
- 10) 3つのルールを覚えるだけ!これであなたもファイル整理の達人(オンライン),入手先 <http://azby.fmworld.net/usage/closeup/20120404/>(参照 2015-2-4).
- 11) Windows ファイルサーバーの最適な運用ルールとフォルダ構成(オンライン),入手先 <http://www.se-support.com/server/fileservers-folder.html>(参照 2015-2-4).
- 12) 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).