



Innovative R&D by NTT

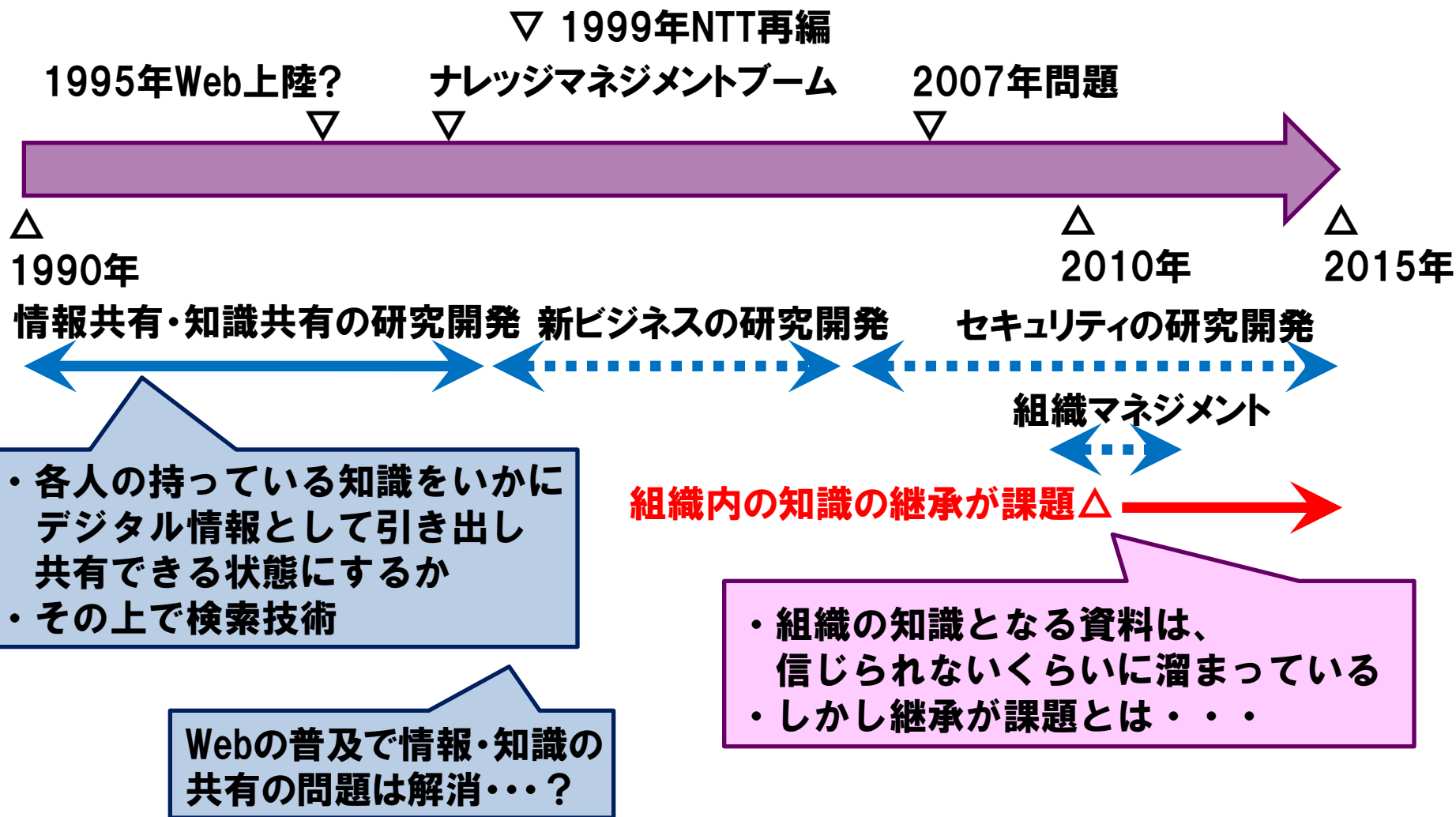
共有フォルダ活用による組織知識の長期継承

2015.7.29

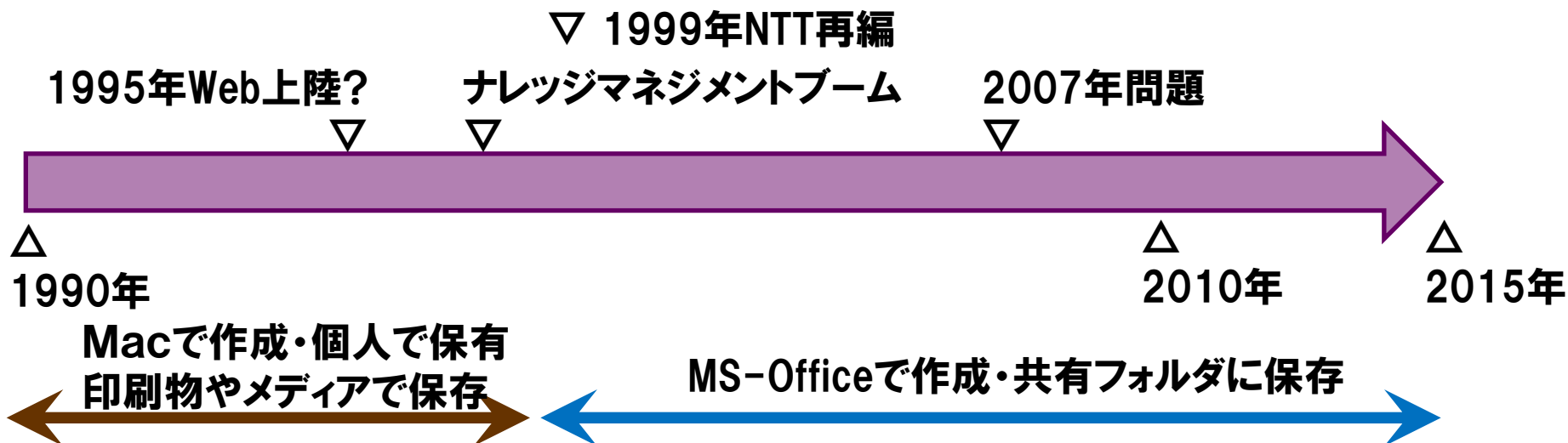
NTTセキュアプラットフォーム研究所

齊藤典明

Background

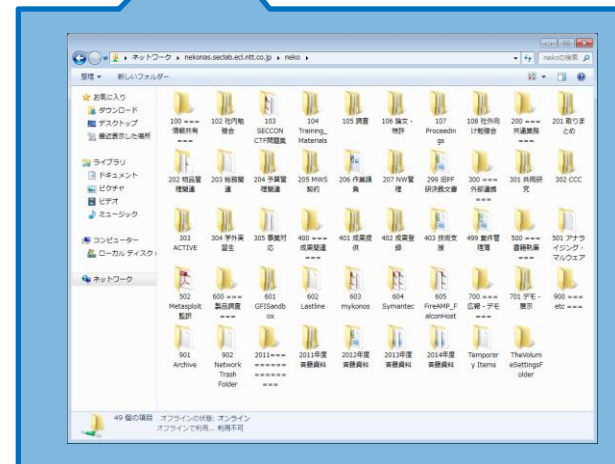
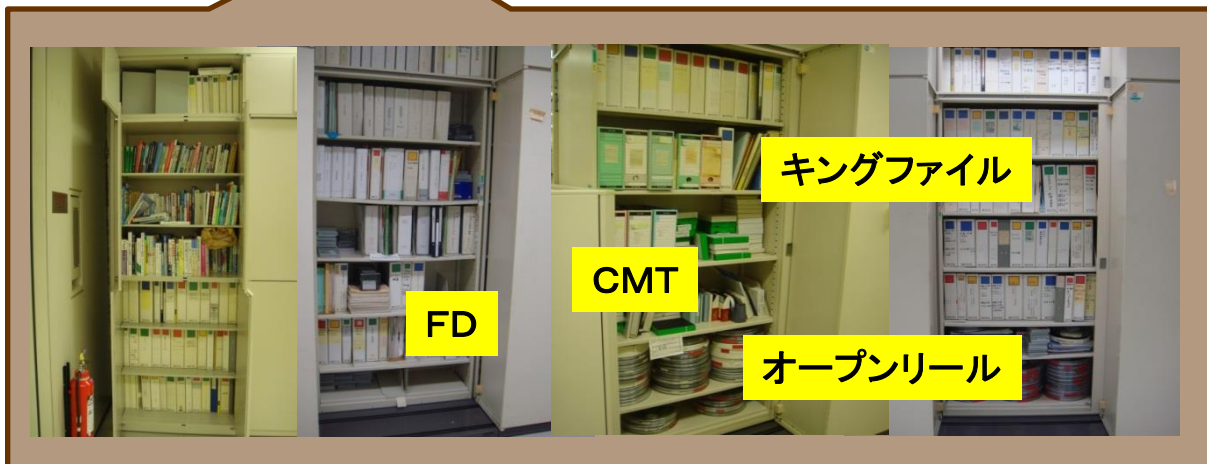


Background



残っていても今では使えないものが多い (Many items are left but unusable now)

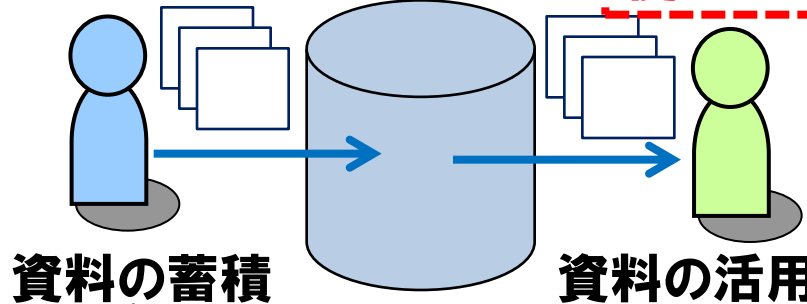
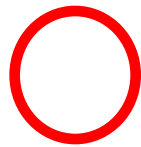
古い資料は今でも使える (Old documents are still usable)



1グループあたり1,000~5,000ファイル/年、1~10GB/年で増加 (1,000~5,000 files/group/year, 1~10GB/year increase)

蓄積資料はなぜみつからないのか？(その1:一般問題)

蓄積する活動は定着している



古い電子資料は手元にあるのに使われていない



気づかない？
あきらめてしまう？
利用価値がない？

分類の問題

(その時その時で思ったようにフォルダに分類して蓄積)

利益不均衡の問題

(活用を意識して蓄積時に不足情報を入力するのは難しい)

記憶の問題・利用者の問題

(どこにしまったか思い出せない)
(どこにあるかわからない)

手がかり不均衡の問題・検索の問題

(蓄積時と活用時でラベルが異なる)
(検索キーワードがわからない)

探すための負荷の問題

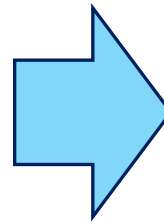
(大量の資料⇒記憶・視覚・操作の負荷)

組織で必要とされる資料の調査

調査1

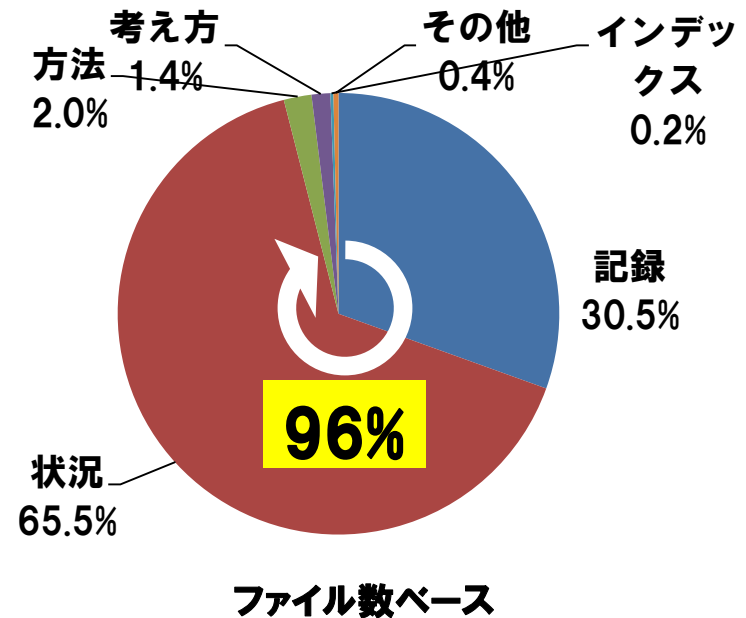
管理職アンケートで組織で継承すべき知識の種類

- 体系化された知識(暗号理論)
- スキル(論文の書き方)
- 記録(会議資料)
- 考え方(コンセプト)
- 状況(調査資料)
- 方法(事務手続き)
- インデックス情報(場所・人)



調査2

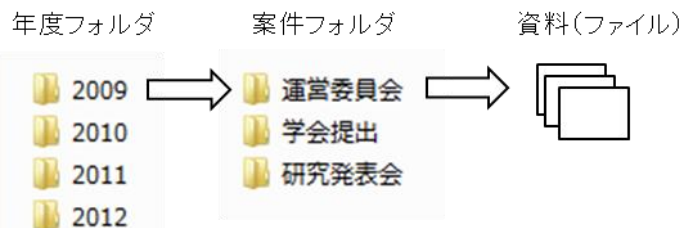
共有フォルダの運用で実際に蓄積された資料の分類



**「状況」と「記録」がほとんど
→会議が資料生成・蓄積のタイミング**

資料の保存方法の調査

調査3 共有フォルダの運用方法



(A) 長期継承に適した構造



(B) 長期継承に適していない構造

「長期継承」に適した議論がない

調査4 世の中一般の整理方法

方式	観点1	観点2	観点3
	活用促進 目的か？	第一階層が 年単位か？	長期保存 目的か？
整理術1	○	×	—
整理術2	○	×	—
整理術3	○	×	—
整理術4	○	×	—
整理術5	○	×	×
整理術6	○	○	×
整理術7	○	×	—
整理術8	○	×	—
整理術9	○	×	×
整理術10	○	×	×

蓄積資料はなぜみつからないのか？(その2:個別の問題)

10年以上の長期運用フォルダの特徴

蓄積資料が増える → データ量 GB~TB
ファイル数 10,000個~
フォルダ数 1,000個~

活動内容が変化 → 分類が変化
使わなくなったフォルダも多数

利用者が替わる → 蓄積した人と利用する人が異なる
(どこに何が入っているかは誰もわからない)

使用頻度 → 頻度は一律ではない
直近の資料のほうが頻度が高い

⇒蓄積構造が不適切→古い資料が新しい資料の活用を阻害

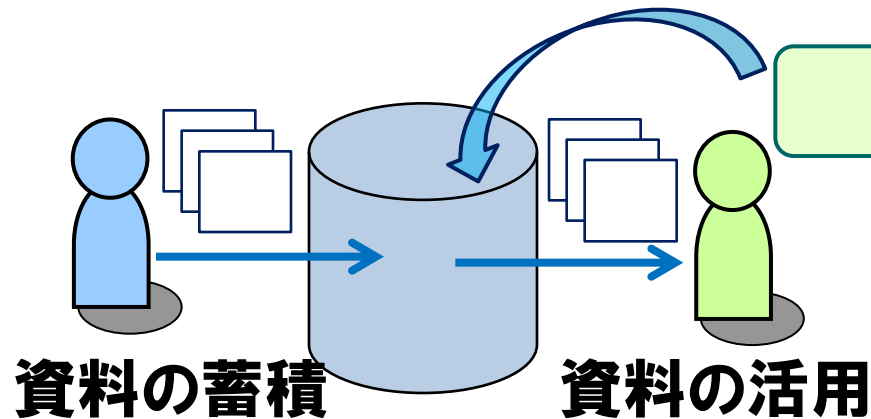
資料活用の観点で整理する方法が必要

長期的な蓄積と継承に向けた提案

その1: 共有フォルダの第一階層を時系列にすると良い

その2: 共有フォルダ用のインタフェース

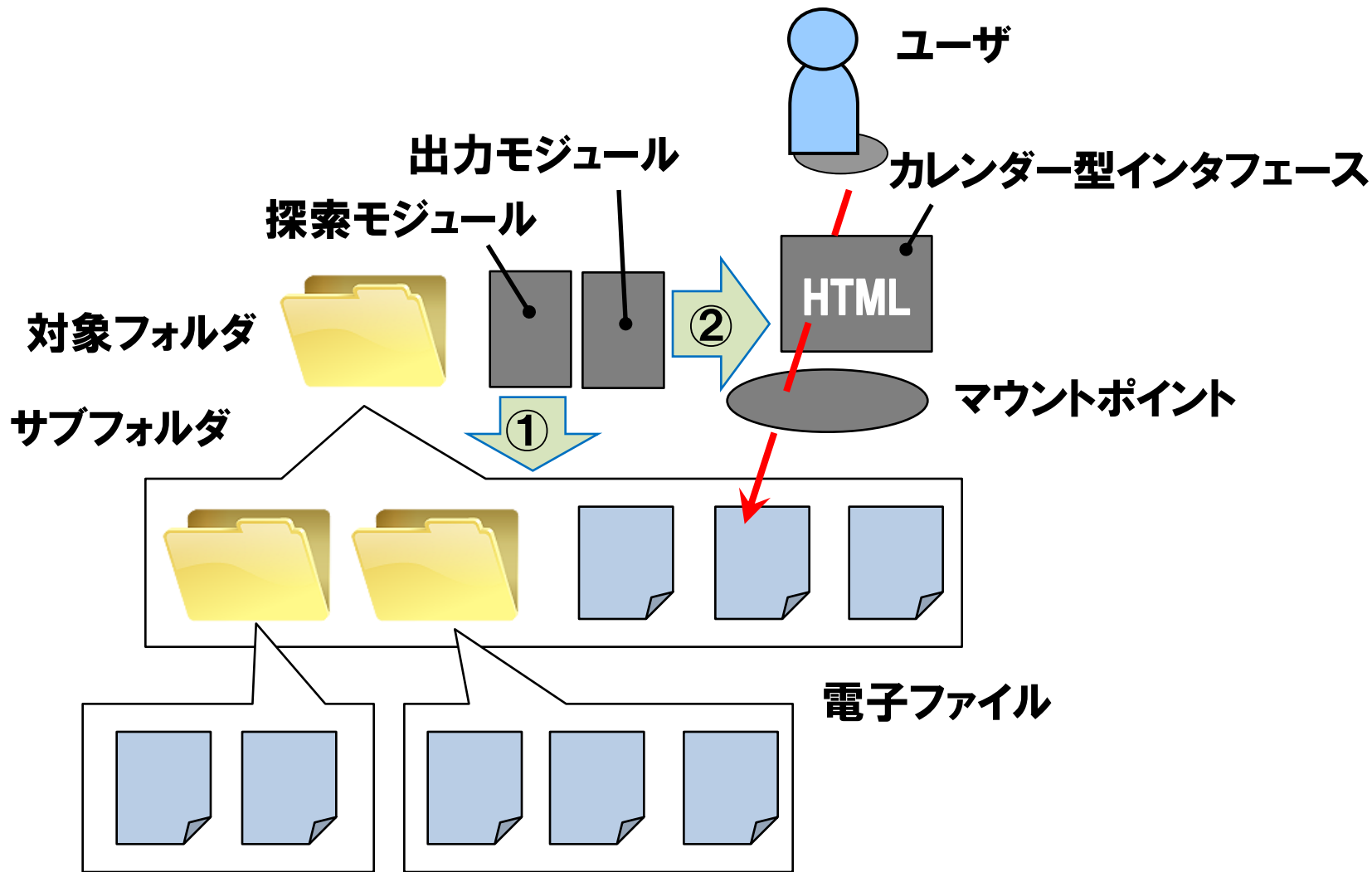
**資料は組織活動(会議などのイベント)で生成・蓄積
イベントベースで取り出す方法ができると有望？**



**イベントの管理
⇒カレンダー形式**

- A 時間情報の抽出**
- B 活動項目の抽出**

システム構成



- ・時間情報 → タイムスタンプから収集 ただし 対象を絞る

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)	×
自動更新ファイル	OSやアプリケーションの動作で自動的にタイムスタンプが更新されるファイル (emlファイルなど)	×

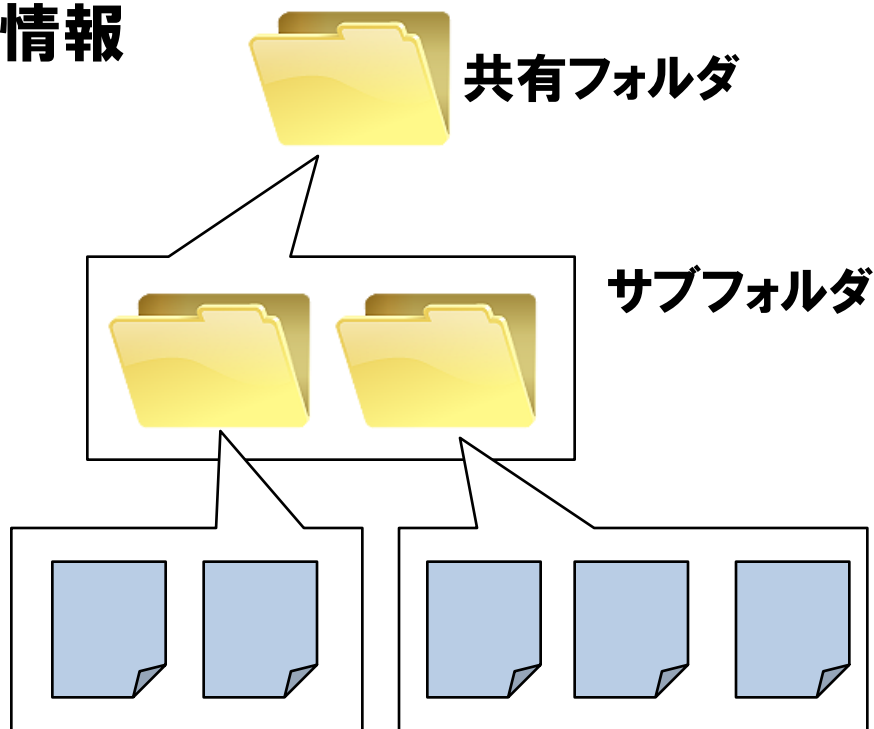
活動項目の抽出手法(着眼点)



・フォルダ名は意味を持って命名される

→ ファイルに対するメタ情報

→ 活動項目のカテゴリ



活動項目の抽出手法(分類方法)



- 共有フォルダ整理法(10種) + 例題フォルダ(約4,000個)から検討

- ・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

- ・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

- ・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

- ・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

- ・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

- ・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

- ・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

- ・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いれる

- ・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

- ・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

活動項目の抽出手法(分類方法)



- ・共有フォルダ整理法(10種) + 例題フォルダ(約4,000個)から検討

フォルダ名の分類方法

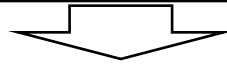
分類	具体例	記述	要否
時系列	年、日付、時期、順番 など	5/10	×
分類方法	組織名、ロケーション名、担当者名、整理方法 など	8/10	×
注釈・記号	資料の扱い方(期限など)・整理上の記号 など	4/10	×
内容・イベント	資料内容の性質や使われた活動名 →自由記述で生成 (各ルールで規定されない部分)	0/10	○

カレンダー型インタフェース上の配置で解決

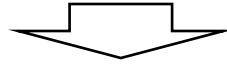
活動項目の抽出手法(アルゴリズム)



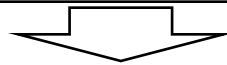
フォルダ全数の調査



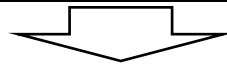
フォルダ名の分類によるカテゴリ抽出:
時系列、整理方法、注釈・記号を表現する単語を削除し、
残りの内容・イベントを表現する単語をカテゴリ名として抽出



抽出カテゴリで網羅できるファイルの割合で絞込み:
0.5%未満、20%以上のファイルを指定するものを除外



目視で単語の修正



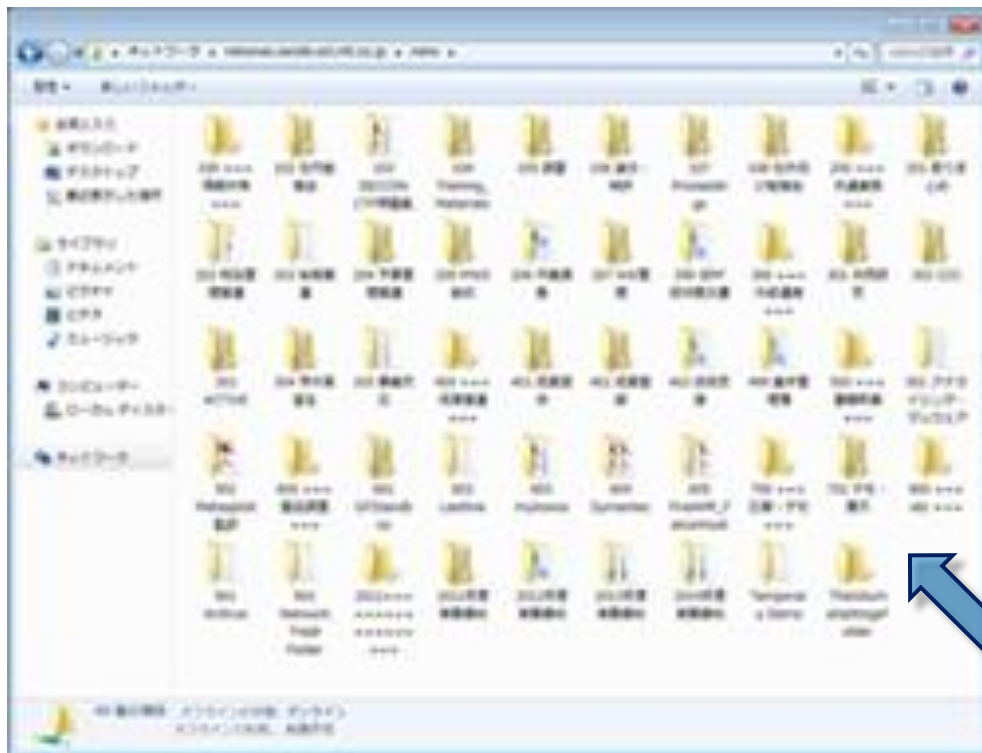
抽出されたカテゴリに属さないファイルを「その他」分類



活動項目のカテゴリ

Before

- 共有フォルダ：厳しい制約がないので気楽に資料を提供
- 一定の整理はされているけど使えない → 混沌としている



10~20人程度で
10年以上使用した状態

フォルダ総数 8,885個
ファイル総数 46,887個
データ量 102GB
フォルダ階層 15階層

第一階層：49個のフォルダ

例：去年のデモ資料を探したい
49個の中からフォルダを選ぶ
実はデモ資料は分散して
入っていてなかなか見つからない

After



init

2015											
01	02	03	04	05	06	07	08	09	10	11	12
2014											
01	02	03	04	05	06	07	08	09	10	11	12
2013											
01	02	03	04	05	06	07	08	09	10	11	12
2012											
01	02	03	04	05	06	07	08	09	10	11	12
2011											
01	02	03	04	05	06	07	08	09	10	11	12
2010											
01	02	03	04	05	06	07	08	09	10	11	12
2009											
01	02	03	04	05	06	07	08	09	10	11	12
2008											
01	02	03	04	05	06	07	08	09	10	11	12
2007											
01	02	03	04	05	06	07	08	09	10	11	12

時間軸からの入り口

2015											
01	02	03	04	05	06	07	08	09	10	11	12
2014											
01	02	03	04	05	06	07	08	09	10	11	12
2013											
2012											
2011											
2010											
2009											
2008											
2007											
2006											
2005											
2004											
2003											
2002											
2001											
2000											
1999											
1998											
1997											

ここはダミーです

08
8+
files.txt
files.txt
files.txt
files.txt
files.txt
files.txt
files.txt
files.txt
files.txt

init

2015											
01	02	03	04	05	06	07	08	09	10	11	12
2014											
01	02	03	04	05	06	07	08	09	10	11	12
デモ・展示							8+				
2013											
01	02	03	04	05	06	07	08	09	10	11	12
2012											
01	02	03	04	05	06	07	08	09	10	11	12
2011											
01	02	03	04	05	06	07	08	09	10	11	12
デモ・展示					12+						
2010											
01	02	03	04	05	06	07	08	09	10	11	12
デモ・展示	19+	39+	10+	1+	1+						
2009											
01	02	03	04	05	06	07	08	09	10	11	12
デモ・展示							1+	4+	12+	28+	
2008											

...1997年まで

同じカテゴリについて複数年で比較

Before → After の比較

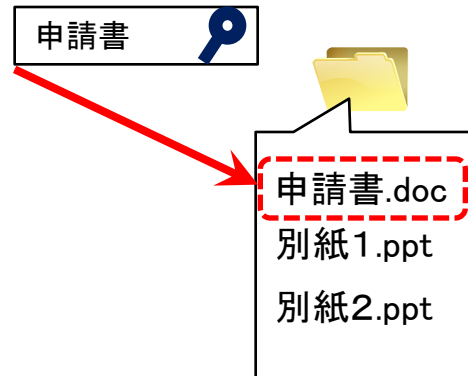


数値的にはいい感じに再分類された

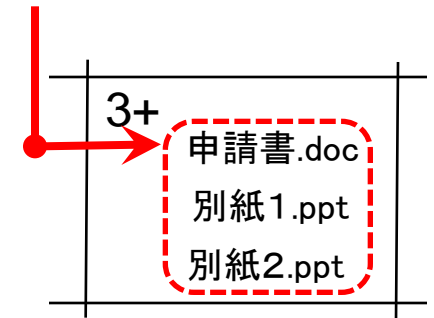
対象	ファイル数	フォルダ数	階層
共有フォルダ	46,887個 (全て)	フォルダ種類=8,725個 フォルダ総数=8,885個 末端フォルダ=7,357個 (平均ファイル数=6.3個)	最大15段 平均5.1段
提案方式	21,691 (資料のみ)	種類:カテゴリ106個+年19個=125個 総数:カテゴリ602個+年19個=621個 末端セル数=2,874個 (平均ファイル数=7.5個) (「その他」は7%=平均6.7個/セル)	3ステップ

■効果：提案方式で混沌とした共有フォルダでも・・・

- ・全体を見通せる
- ・活動項目ベースで探せる
(複数の単語から探せる)
- ・関連資料も探せる



ファイル検索



カレンダー型インタフェース

■今後の課題：

- ・手作業の部分を減らす(体系化する)
- ・抽出単語の妥当性／チューニング方法
- ・ファイル分類に偏りがある場合の再分類方法