



公文書管理における 情報科学技術 (ICT) 利活用

立命館大学情報理工学部
上原哲太郎

R 文書のデジタル化の利点と課題

RITSUMEIKAN

主な利点

- 検索が容易
- 再利用が容易
- 破棄が容易
- 物理的移動がないので保管作業が容易
- 保管に場所をとらず大量の保管が可能に
- 文書に付加情報を紐づけるのが容易

主な課題

- 改ざんや消去が容易
- 改ざんや消去の検出が困難
- 原本と複製の区別ができない
- 利用者にある程度のリテラシーを要求する
- 長期保管の技術的課題



文書のデジタル化の方向性

ファイル名：桃太郎.txt
文書・テキスト形式
作成者：山田太郎
管理課：XX課
作成日：H30/12/1
保管期限：H35/3/31

ファイル名：桃太郎.xml
文書・XML形式
作成者・山田太郎
管理課・XX課
作成日・H30/12/1
保管期限・H35/3/31

桃太郎
昔々或る処に
おじいさんと
おばあさんが
おりました

桃太郎
昔々或る処に
おじいさんと
おばあさんが
おりました

桃太郎
昔々或る処に
おじいさんと
おばあさんが
おりました

```
<表題>桃太郎</表題>  
<章>1</章>  
<登場人物>  
<人>おじいさん</人>  
<人>おばあさん</人>  
</登場人物>
```

手書き
文書

清書

メタ情報
付加

文書の
構造化

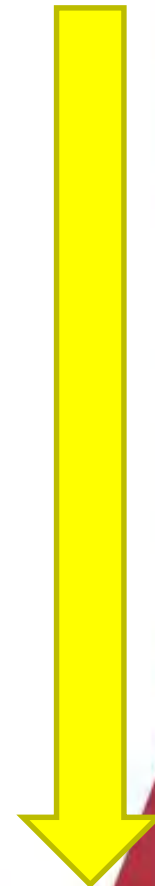
Beyond Borders

文書の「内容」と「表示形式」の分離
文書に対する「機械可読な意味づけ」

R 文書のデジタル化の段階

RITSUMEIKAN

- 第1段階：清書としてのデジタル化
 - 紙の文書作成事務を手書きから置き換え文書作成効率と可読性を高める
- 第2段階：メタデータの付加と利用
 - 文書ファイルに機械可読可能な付加情報を加えて利活用し文書管理を容易にする
- 第3段階：文書そのものの構造化
 - 文書の本文も機械処理可能な構造化された状態で作成し全体の管理や処理を容易に



システム導入だけでは不十分
従事者のITリテラシを要する

R 公文書管理が抱える課題：私見

RITSUMEIKAN

- デジタル化としてはまだ道半ば
 - 清書段階が多くメタ情報付加も不十分・非統一
- 公文書管理として求められる要求に応える作業が現場に大きな負担
 - 管理作業の多くが人手に頼っており手順が煩雑
 - 改ざん・消去・隠ぺいの懸念に対する透明性確保要求に対し技術運用とも対応困難
- 文書公開請求に対し検索と「墨塗り作業」が大きな負担



公文書管理へのICT活用

- **前提：「清書段階」の文書管理**
- **文書のデータ形式が統一されていない
(場合によっては文書がテキストでなく画像)**
- **メタ情報も不十分で期待できない**
- **解決すべき課題**
- **文書作成・保管・廃棄にかかる事務の効率化**
- **公文書公開請求にかかる事務の効率化**
- **文書消去改ざんが無いことの証明**


R 文書管理事務の効率化

RITSUMEIKAN

- **文書管理システムの導入**
- **文書ファイルの体系的保管と検索、廃棄フローを実現**
- **電子決済と一体型／分離型両方ある**
- **効果**
 - **システム内ではメタ情報がつく**
 - ファイル名・表題・作成者・作成日・管理課・保存年限・決済履歴など
 - **検索も（文書のデータ形式が正しければ）容易**
 - 画像内の文字などはまだ技術的課題がある
- **課題**
 - **文書作成事務そのものがやや煩雑に・リテラシ必要**
 - **システムが想定しない「例外」に弱い**
 - 特殊な形式の文書／テレワーク対応／緊急時対応／システム障害など

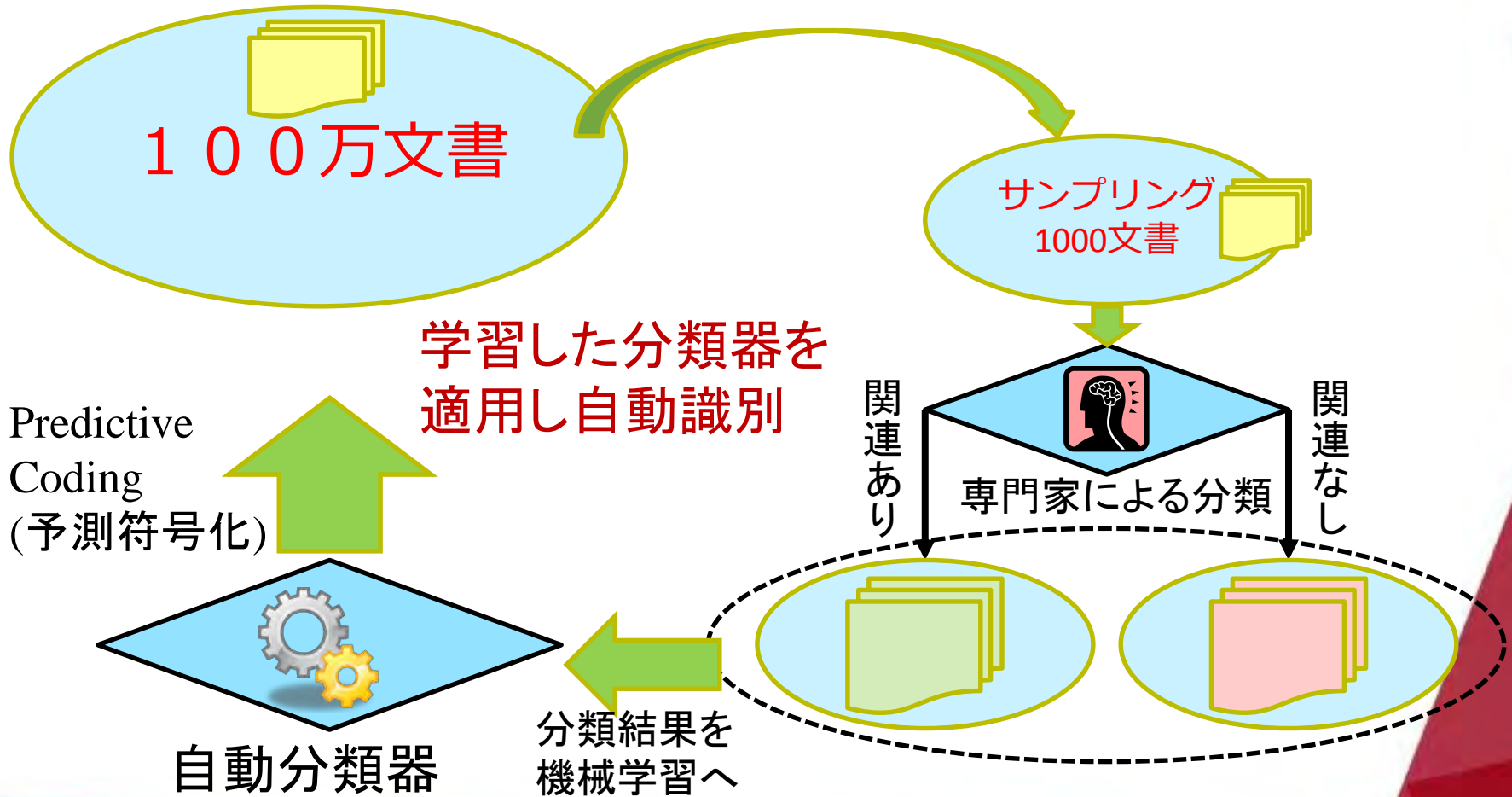


文書公開事務の課題：検索

- 現在文書管理システムでできるのは主に「キーワード検索」
 - しかしキーワード検索は検索語の選び方が難しい
 - 不要な文書も多く検索してしまう場合
 - 適切な検索語が不明な場合（「XXに関する文書」など）
 - テキストだけではなく画像も検索対象の場合
- 
- 自然言語処理技術・AIを用いた解決は可能か？



デジタル・フォレンジックで 活用が進む機械学習による文書分類



Beyond Borders

Discovery制度がある国で活用が盛ん

文書公開事務の課題：



画像に対する文字検索

- 近年のAI関連技術の向上で
画像内の文字認識精度は劇的に向上
- 但し相応の計算能力を要する = コスト増大

公文書

テキストデータ

公文書

画像データ



文書公開事務の課題：「墨塗り」

- 公開文書の墨塗り部分に必要な情報が隠れていないか？

- 「墨塗り署名」
特殊な電子署名技術

- 文書に改ざんがないこと
- 墨塗り部分に特定の言葉が隠れていないことを証明できる

技術的には可能だが運用が課題

「桃太郎に関する文書」

桃を割ると
中から大きな
[redacted]が
出てきました

「桃太郎」
ではないか？
(実際は金太郎)

R 改ざん検出・消去隠ぺいの検証

- 暗号技術が大きく貢献できる
- 技術も普及しており利用は比較的容易
国もGPKIを整備済み
- 基本的な安全性確認も国の主導で実施
(暗号技術検討会CRYPTREC)



R 電子署名を活用する際の課題

RITSUMEIKAN

- 鍵ペア（秘密鍵・公開鍵）の管理コスト
 - 職員全員に個別に持たせて管理するコスト
 - 事故による漏洩や紛失への対応コスト
- 署名を「いつ」行ったかの証明は必要か
 - タイムスタンプ署名の活用
- 署名したものが存在しない場合は？
 - システム的対応（デジタル・フォレンジック）
 - ブロックチェーンの導入

消去ファイルの
検索や復元

どちらも公開鍵暗号技術に基づく R ブロックチェーンとPKI

RITSUMEIKAN

公開鍵認証基盤 (PKI)

- 既に整備済みで普及
法的根拠も問題なし
- 標準規格があり
様々なデータの
暗号化と署名に利用可
- 鍵ペア管理がやや煩雑
- 認証局維持が高コスト
- 文書存在の証明には
別システムが必要

ブロックチェーン

- 規格化はまだこれから
- 改ざん不能なデータを
公開できるのが特徴
- タイムスタンプ署名と同等
文書存在証明が容易
- 公開データは消去困難で
「誤り」が訂正できない
- システムコストが不明
- 暗号通貨と無関係に系を
維持する利点が不明

R まとめ

- 既にICTには公文書管理に使える種々の技術があるがそれを運用する職員のリテラシの底上げが不可欠で人材育成が必要
- 一定のシステム投資も将来必要
- 公文書管理のニーズが民間と異なるため独自の文書管理システムにならざるを得ない