

ウェブ情報のアーカイブ化促進に資 する実証実験 概要

平成17年10月3日

総務省

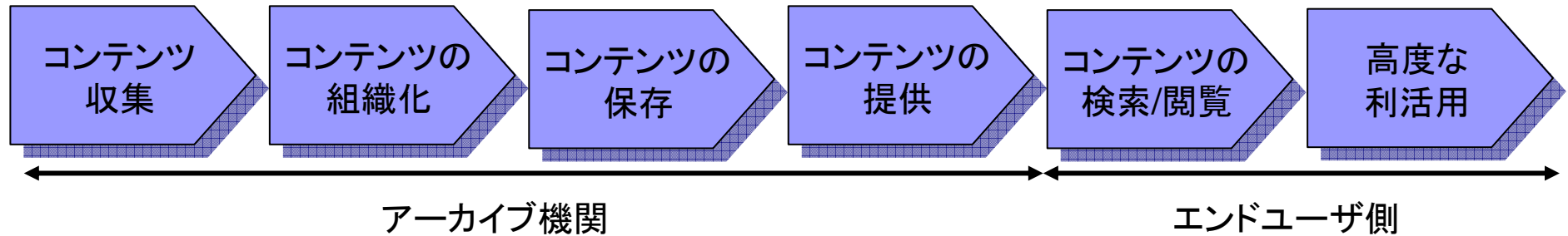


本実証実験の目的

- ウェブ情報にはデジタル時代の知識・文化が結集
- しかしながら、ウェブ情報は日々の更新、消去により散逸

⇒ 16年度の実証実験の成果を踏まえ、引き続き17年度においても実証実験を行い、デジタル時代の貴重な文化遺産であるウェブ情報のアーカイブ化促進や、様々な主体による複数アーカイブ間の連携利活用を可能とする技術・仕組みを構築・実証

平成17年度実証実験の概要



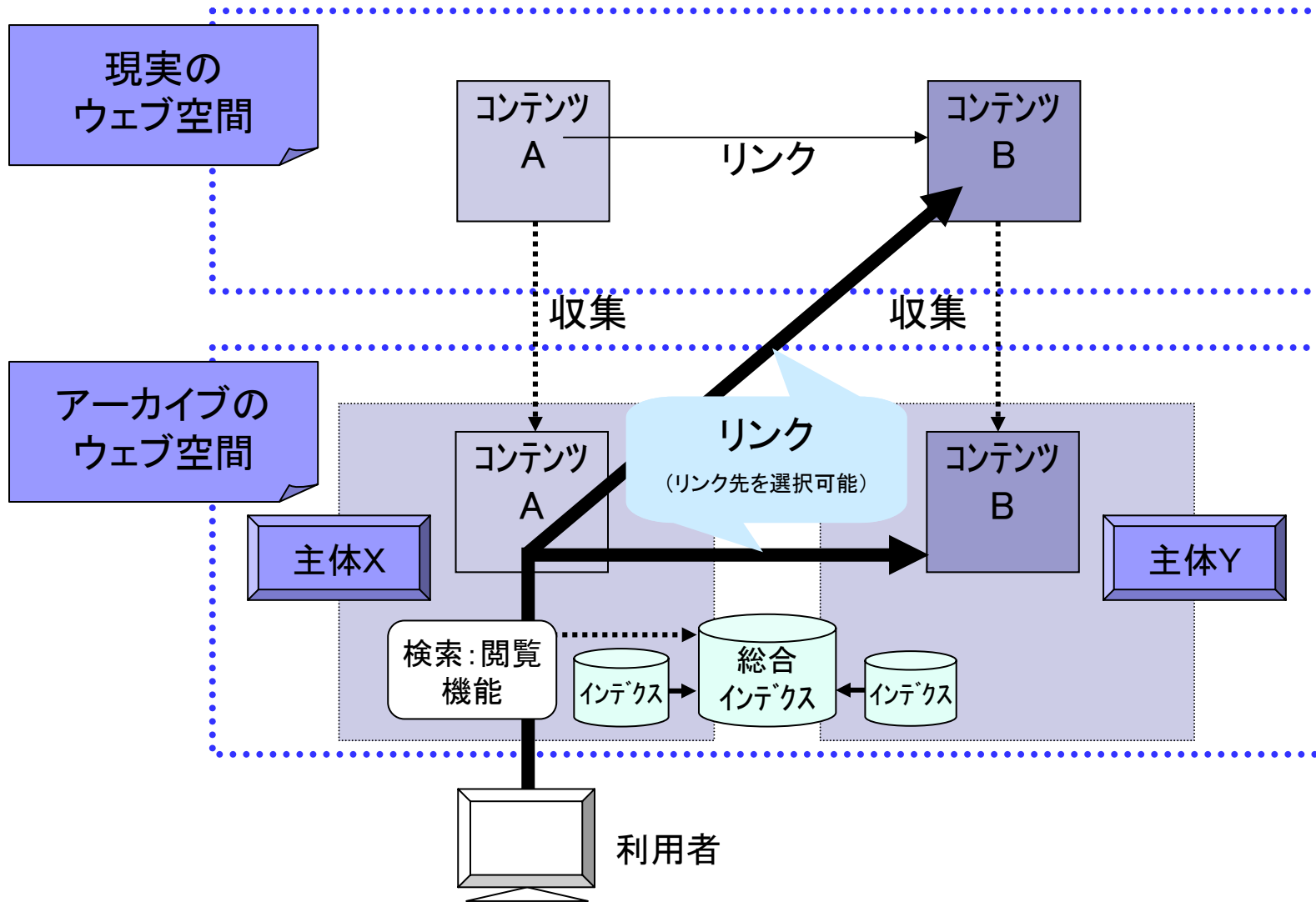
16年度実施項目

- (1) 異なるアーカイブ間の横断検索を実現する汎用的なメタデータ体系の確立
- (2) メタデータに基づく時系列別検索、特定コンテンツに係る時系列別表示の実現
- (3) コストの縮減、長期的な保存を可能とするアーカイブ技術の開発等
- (4) 同一のウェブ情報を重複して保存することを回避する技術に関する調査研究
- (5) 収集範囲拡大に関する調査研究

17年度実施予定項目

- (1) 複数ウェブ情報アーカイブ間のリンクナビゲーション機能の開発
- (2) ウェブ情報アーカイブにおける効果的な検索・閲覧インターフェイスの開発等
- (3) 業務作業負荷の軽減を可能とする情報の自動検出・抽出技術の開発等
- (4) ウェブ情報アーカイブに係るメタデータ体系のさらなる検証・確立

1. 複数ウェブ情報アーカイブ間のリンクナビゲーション機能の開発



アーカイブのウェブ空間におけるコンテンツAから、コンテンツBへリンクする。
(現実のウェブ空間のコンテンツBもしくは、アーカイブのウェブ空間のコンテンツBを選択可能)

2. ウェブ情報アーカイブにおける効果的な 検索・閲覧インターフェースの開発等

平成16年度:時系列別検索・閲覧に関する基本的な機能の開発

平成17年度:時系列別検索・閲覧に関する拡張的な機能の開発

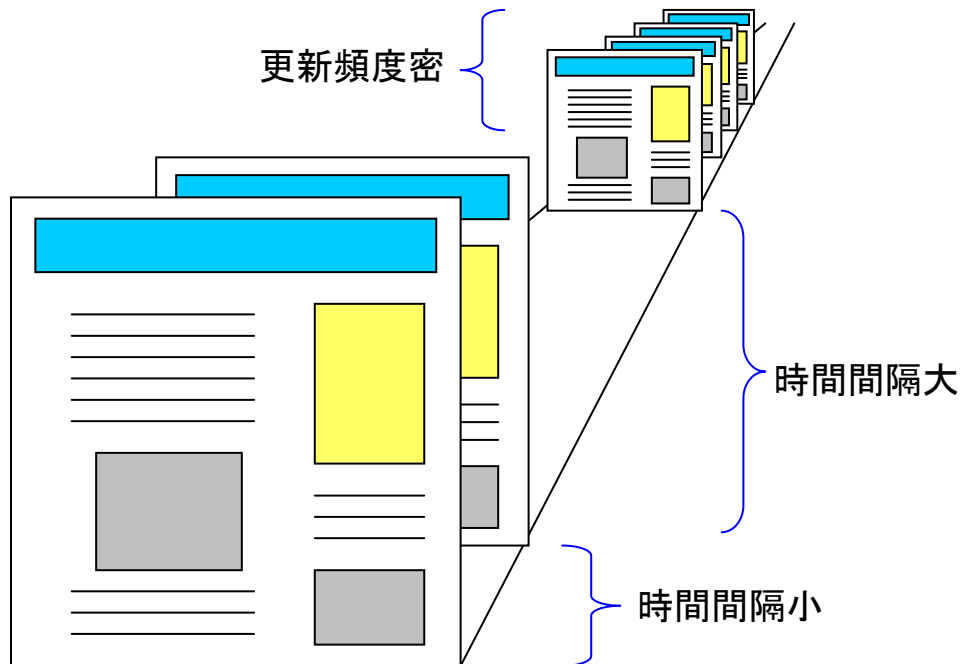
(時系列に見た際の変化をより視覚的に表示する技術の開発)

時間をさかのぼりながらページの移り変わりを自動的に検出

2-1. ページ更新頻度の可視化

三次元仮想空間上の距離で時間の長短を表現

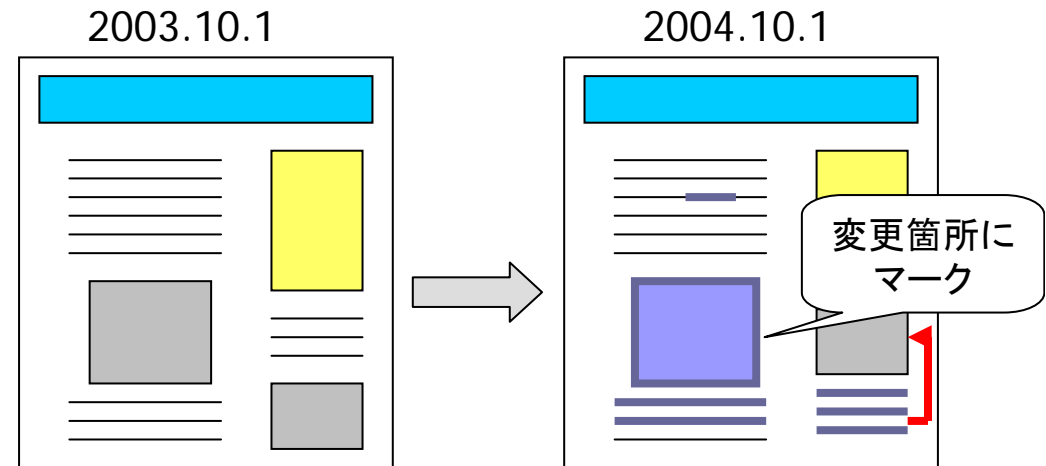
ウェブページ間の時間の前後関係だけでなく、
更新頻度、更新時期を直感的に把握できる



2-2. 変更箇所の強調

時間の前後でウェブページの内容を比べて、変更箇所を抽出

時間軸上の前後のウェブページを比較する際に、
変更箇所が一目で確認できる



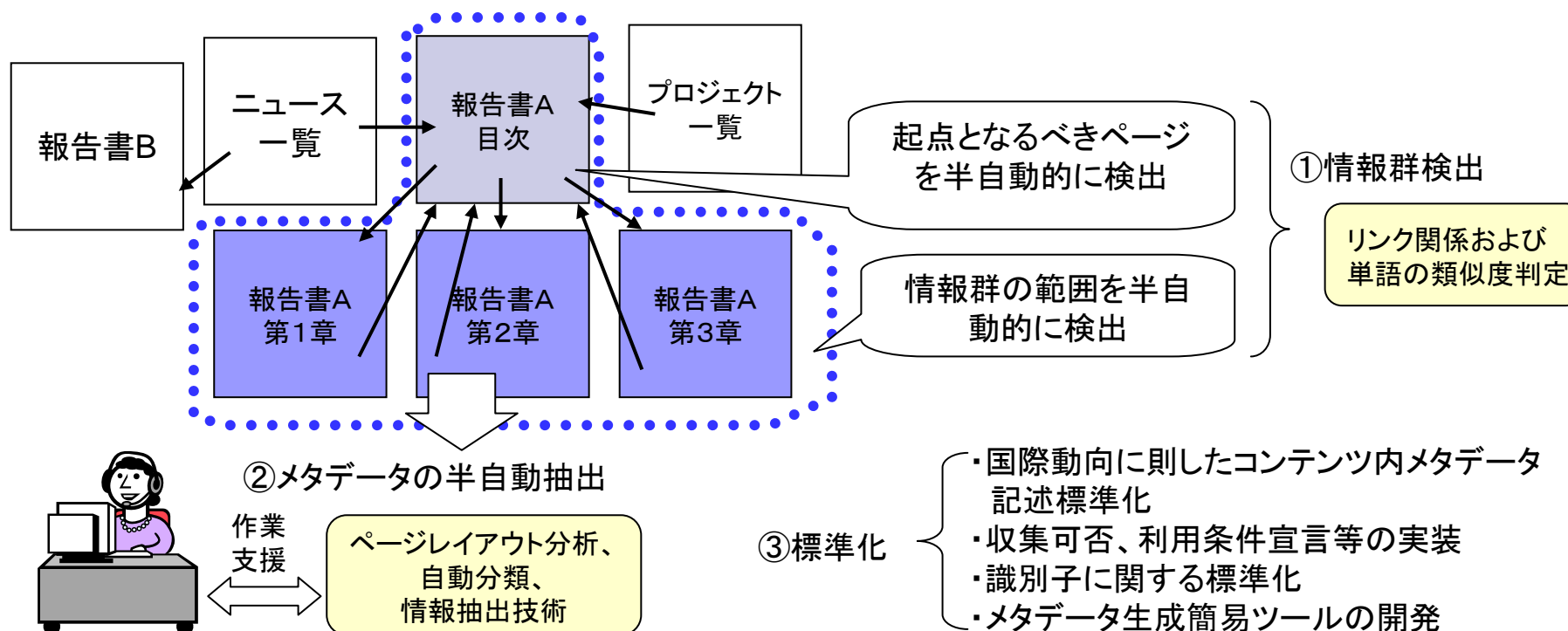
平成16年度の調査研究結果を踏まえ、ユーザーの利便性を考慮したポータルサイト等のUIの再検討も実施

3. 業務作業負荷の軽減を可能とする情報の自動検出・抽出技術の開発等

比較的細かい粒度でアーカイブを行う場合、

- ① [情報群検出] 複数ページから構成されるコンテンツの範囲、起点を同定。
- ② [抽出] Dublin Coreを基本とするメタデータを、ページコンテンツから半自動抽出。
- ③ [標準化] コンテンツ、メタデータの作成・記述の標準化の提案に向けた検討。コンテンツ作成者向けメタデータ生成簡易ツールの開発。

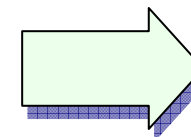
メタデータ付与
業務の軽減



4. ウェブ情報アーカイブに係るメタデータ体系のさらなる検証・確立

平成16年度：
基本的なメタデータ体系の構築と複数アーカイブ間の横断検索

平成17年度：
実証項目1, 2, 3や最新動向をふまえた、メタデータ体系の検討



詳細なメタデータ体系の検討・確立