全文テキスト化実証実験参加協力会社との定例会（第１回）

日時：平成２２年１０月２９日（金）１０時 ～１１時
場所：国立国会図書館新館３階研修室

議事次第

１　報告

○ 平成 22 年度全文テキスト化実証実験の概要
国立国会図書館総務部企画課長　田中久徳

○ 予備的調査及びプロトタイプシステムの概要
株式会社三菱総合研究所

２　質疑応答
３　閉会

配布資料

- 資料１　　平成 22 年度実施全文テキスト検索の実証実験の状況
- 資料２－１　予備的調査の概要
- 資料２－２　全文テキスト化システムプロトタイプの概要について
- 資料２－３　全文検索・表示システムプロトタイプ構築等作業の提案概要

（次回予定）

- 次回（11 月中開催予定）は、以下の点について報告予定
  ① 進捗報告
  ② 評価の実施方法及び項目
  ③ 有識者検討会
  ④ その他

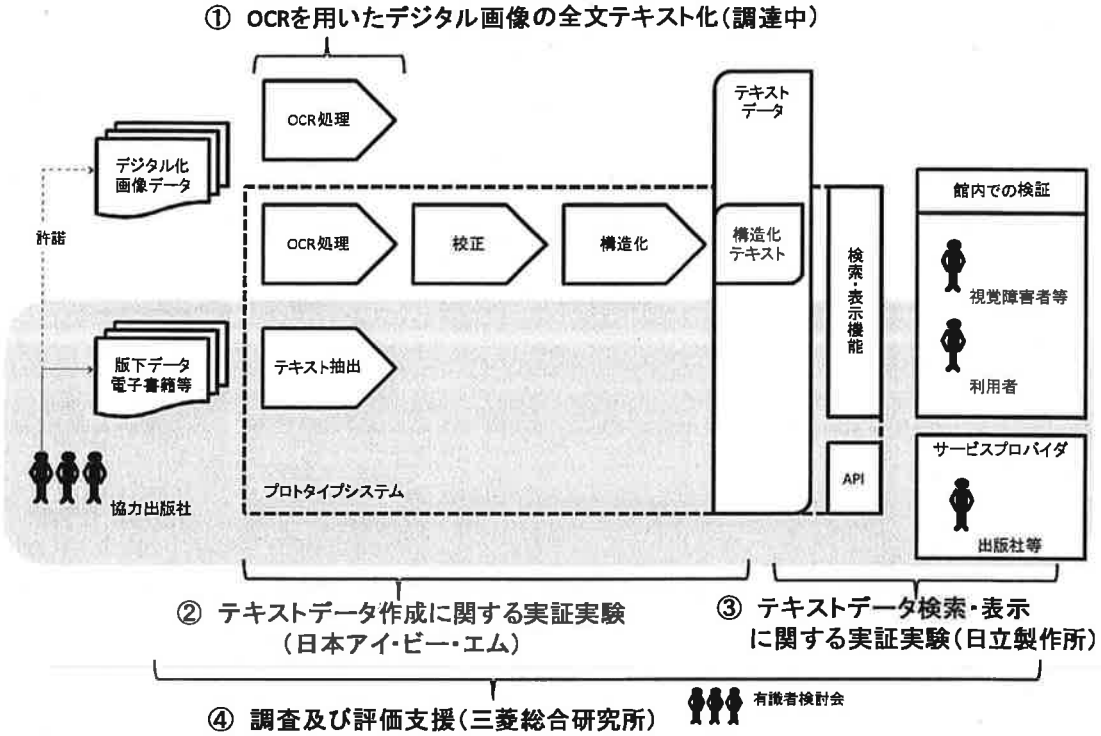
問い合わせ先
国立国会図書館総務部企画課
全文テキスト化実証実験担当
TEL：03-3506-5297（直通）
E-mail：digi-jimu@ndl.go.jp

資料１

平成 22 年 10 月 29 日
国 立 国 会 図 書 館

平成 22 年度実施全文テキスト検索の実証実験の状況

1 事業の全体



調達件名	受託業者	スケジュール	対象数量
① OCR を用いたデジタル画像の全文テキスト化	調達中	11～1 月：作業	2 万冊 （主に戦前期図書）
② 全文テキスト化システムプロトタイプ構築	日本アイ・ビー・エム	10～1 月：構築 2～3 月 ：試行	3,000 ページ （対象資料選定中）
③ 全文検索・表示システムプロトタイプ構築	日立製作所		①＋②＋ 出版社提供データ
④ 全文テキスト化実証実験に係る調査及び評価支援	三菱総合研究所	9 月～：作業	－

2 参加協力会社の概要

（ア）参加協力内容

- 実証実験用の出版データの提供　（10 月中）
- 実証実験に係る定例会への参加　（月次）
- プロトタイプシステムの試行及び評価（2～3月）

※参加協力会社には、最終報告書を配布予定

(イ) 参加協力会社数

・ 39 社 ＜http://www.ndl.go.jp/jp/aboutus/digitization_fulltext.html＞

(ウ) 出版社提供データ数(10 月 29 日時点)

・ タイトル数:300、冊数:600、ページ数:200,000

(エ) 提供データ形式

・ PDF 、XMDF、TEXT

3 データの取扱い及び授受について

(ア) データの取扱い

以下の条件のもと、出版社からデータを提供していただく予定である。

・ 国立国会図書館は、データ提供者の許可なく、データを当実証実験の用途以外に使用しない

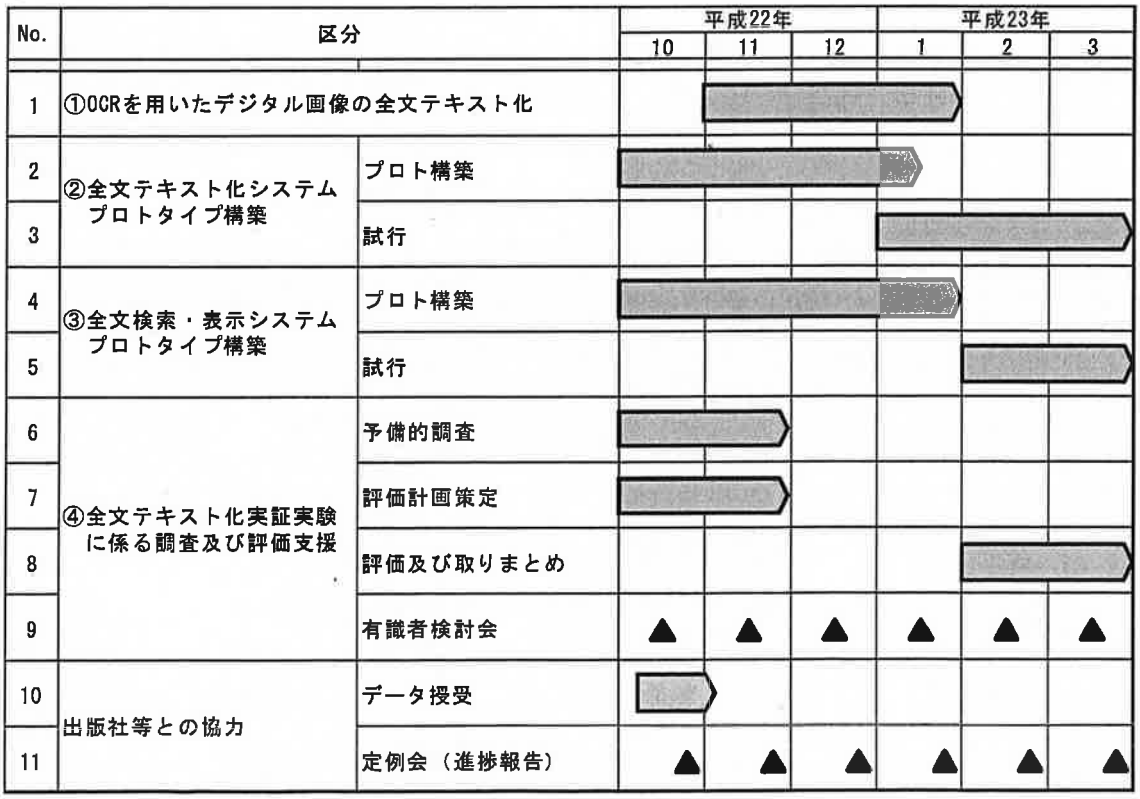
・ 国立国会図書館は、故意又は過失によりデータ漏えい等が発生した場合を除き、データに関し権利侵害等の問題が生じたときは、一切の責任と負担を負わない

・ データ提供者は、国立国会図書館に対し、データの消去を求めることができる

(イ) データの授受

・ 媒体に格納の上、10 月中に提供してもらう旨、依頼済

4 スケジュール



資料2－1

全文テキスト化実証実験

予備的調査の概要

2010年10月29日

株式会社三菱総合研究所

1. 電子書籍に係る動向調査の作業内容及び方法

(1) 事業者に関する調査

- 出版業界、印刷業界等に属する事業者に対して、電子書籍出版への対応状況、採用技術等の動向について調査を実施。
- 調査項目（予定）
  - 電子書籍への対応状況について
  - 電子書籍の採用技術について
  - 最近の取り組み事項におけるトピック
  - 国立国会図書館様の全文テキスト化の取り組みについての要望・ニーズ　など

(2) 事業者団体に関する調査

- 出版業界、印刷業界等の企業で構成される事業者団体に対して、業界全体としての電子書籍出版への対応状況、採用技術等の動向について調査を実施。
- 調査対象（候補）
  - 日本電子書籍出版者協会（EBPAJ）
  - 電子書籍を考える出版社の会（eBP）
  - 日本電子出版協会（JEPA）
  - 電子出版制作・流通協議会（AEBS）
- 調査項目（予定）
  - 団体の概要について
  - 業界における電子書籍への対応動向について
  - 国立国会図書館様の全文テキスト化の取り組みについての要望・ニーズ

（次ページへつづく）

1. 電子書籍に係る動向調査の作業内容及び方法

- 調査項目（続き）
  - 旧字・新字、外字の扱い、文字コード
  - 日本語表示固有の表現
    - 図表
    - ルビ
    - 縦書き　など
  - 数式等の特殊表現　など

(3) 調査方法

- 文献調査により、該当する分野に関する全体的なサーベイを実施した上で、その中でも特徴的であり、さらなる調査が必要なものについて、インタビュー等による個別の確認を実施。
- 海外事例については、調査期間を勘案して、公表されている文献ベースでの調査を原則とし、現地訪問調査等は実施しない予定。

2. 全文テキスト化に係る技術・製品動向調査の作業内容及び方法

(1) 全文テキスト化に係る技術動向調査

- 全文テキストを活用したシステム構築（全文テキスト化、全文検索・表示、視覚障がい者向け読上げ等の機能を有するシステムの構築）に当たって、実装可能又は将来的に実装可能となることが予想される技術の動向について、調査を実施。
  その際には、アクセシビリティ対応の技術及び製品も調査対象に含む。
- 調査対象（予定）
  - 電子書籍のフォーマットについて
  - OCR技術の動向について
  - 情報検索技術の動向について
  - 言語横断検索、検索結果及び検索結果の要約の自動翻訳の動向について
  - 検索結果の自動要約について
  - アクセシビリティについて

(2) 全文テキスト化に係る製品動向調査

- 全文テキストを活用したシステム構築に当たって、実装可能又は将来的に実装可能となることが予想される製品（パッケージソフトウェア）の動向について調査を実施。
- 調査対象（予定）
  - OCRソフト、ドキュメントリーダ
  - 検索ソフトウェア
  - 配信ソフトウェア
  - 読み上げソフトウェア
  - テキストマイニングツール

（次ページへつづく）

2. 全文テキスト化に係る技術・製品動向調査の作業内容及び方法

(3) 調査項目

- 調査項目(予定)
  - 技術・製品の概要について
    - 何を実現する技術・製品なのか
    - どのような特長がある技術・製品なのか　など
  - 全文テキスト化実証実験における活用局面について
    - テキストデータ作成、またはテキストデータ検索・表示いずれに対して適用可能な技術・製品なのか
    - その技術・製品を活用することで、どのようなことが実現可能なのか　など

(4) 調査方法

- 調査方法については、「1 (3) 調査方法」(2ページ)を参照。
- なお、調査の際には、H21年度に事務局が実施した調査の結果も活用。
- また、商用製品に限定することなく、オープンソースソフトウェアや教育研究機関等の研究成果も含めて調査を実施。

3. 全文テキスト化に係る先進事例調査の作業内容及び方法

(1) 調査対象

- 全文テキストを活用したデジタル化資料配信事例、電子図書館事例、アクセシビリティ対応事例、全文テキストの構造化及び全文検索に関する研究開発事例等について調査を実施。
- 調査対象(予定)
  - デジタル化資料配信事例
  - 電子図書館事例
  - アクセシビリティ対応事例
  - 全文テキストの構造化及び全文検索に係る研究開発事例

(2) 調査項目

- 調査項目(予定)
  - 事例の概要について
  - 今後の展開と課題について
  - 全文テキスト化実証実験への応用について

(3) 調査方法

- 調査方法については、「1 (3) 調査方法」(2ページ)を参照。
- また、ミュージアムやアーカイブ等における電子化取組事例も含めて調査を実施。

(次ページへつづく)

4. 実証実験に係る予備的調査の結果取りまとめ

■ 1～3に示した予備的調査の結果について、全文テキスト化実証実験に係る予備的調査報告書(案)として取りまとめる。

□ 全文テキスト化実証実験に係る予備的調査報告書 目次(案)

◆ 電子書籍に係る動向
- » 出版社における電子書籍に関する取り組み
- » 印刷業者における電子書籍に関する取り組み
- » 書店における電子書籍に関する取り組み
- » システムベンダーにおける電子書籍に関する取り組み
- » 事業者団体における電子書籍に関する取り組み

◆ 全文テキスト化に係る技術・製品動向
- » 電子書籍のフォーマットの動向
- » OCR技術の動向、OCRソフト、ドキュメントリーダに係る製品動向
- » 情報検索技術の動向、検索ソフトウェアに係る製品動向
- » 言語横断検索、検索結果及び検索結果の要約の自動翻訳の動向
- » 検索結果の自動要約
- » アクセシビリティに関する動向
- » 配信ソフトウェアに係る製品動向
- » 読み上げソフトウェアに係る製品動向
- » テキストマイニングツールの動向

◆ 全文テキスト化に係る先進事例
- » デジタル化資料配信の事例
- » 電子図書館の事例
- » アクセシビリティに対応した事例
- » 全文テキストの構造化及び全文検索に係る研究開発事例

◆ まとめ
- » 全文テキスト化実証実験に向けた技術・製品・事例の活用についての検討

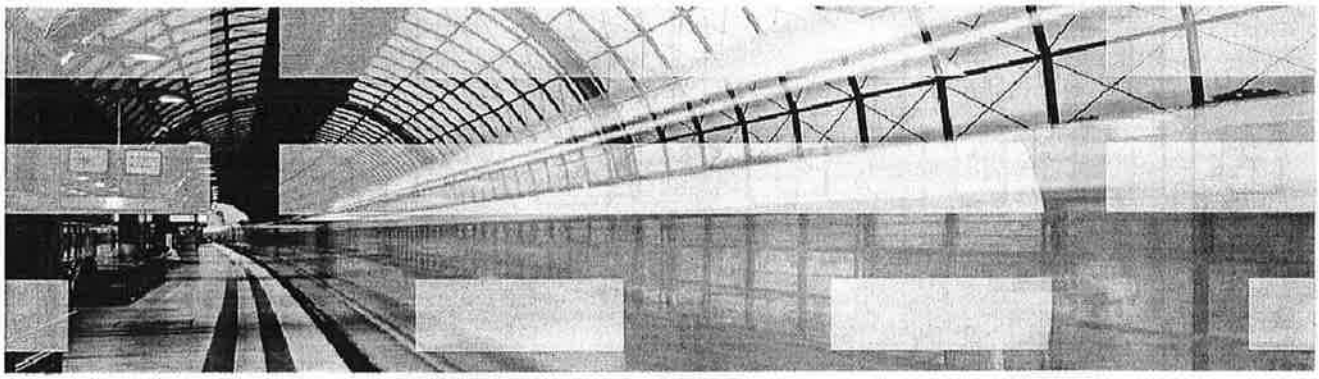
※)小項目は例示。

国立国会図書館
National Diet Library

IBM

資料2－2

全文テキスト化システムプロトタイプの概要について（ご説明資料）



2010年10月29日
日本アイ・ビー・エム株式会社

IBM

全文テキスト化システムプロトタイプ構築の作業スコープ

- 全文テキスト化システムプロトタイプ構築においては以下の作業を実施するものとします。

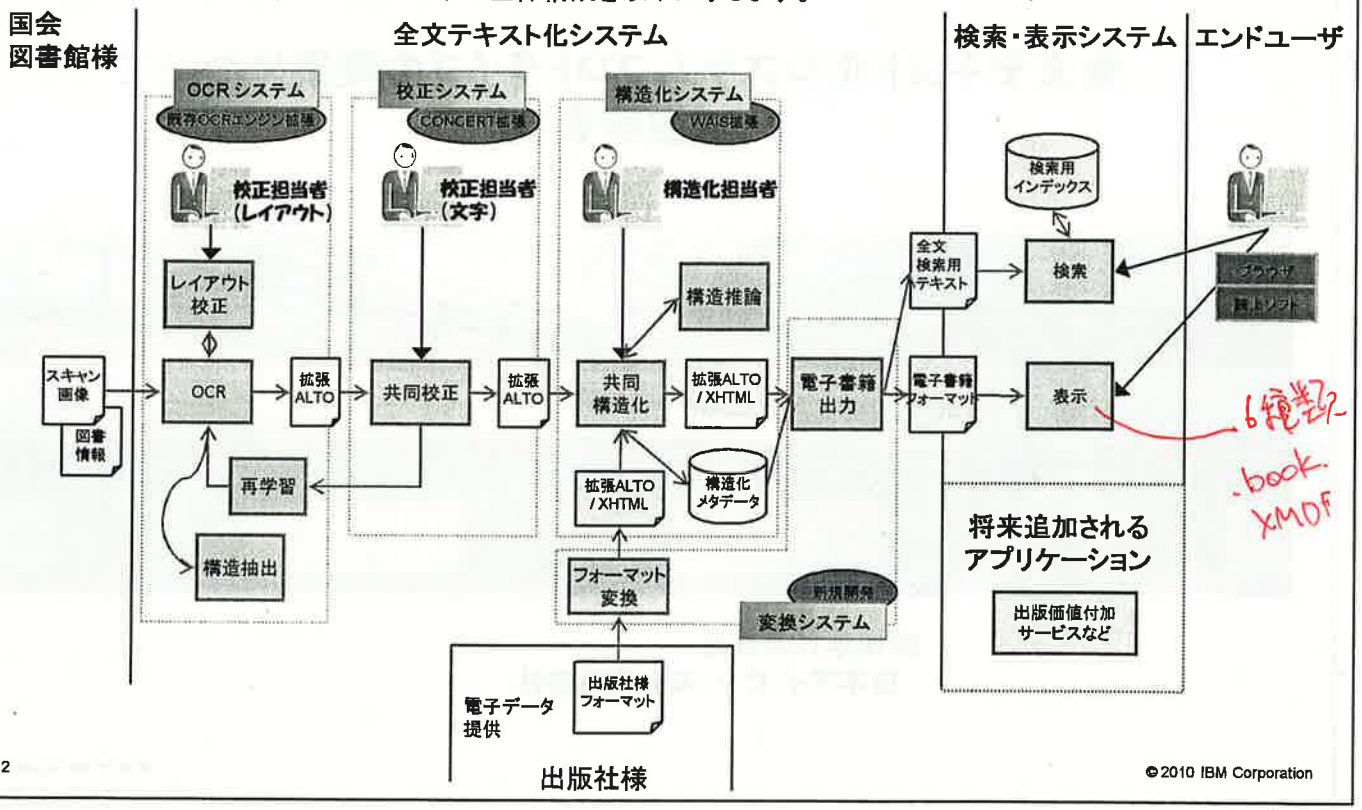
(1)標準フォーマットによる全文テキスト化システムの構築
a.日本語対応のOCR出力標準フォーマット (ALTO等の標準的なフォーマットの拡張版を想定。以下「標準中間フォーマット」という。)の検討・定義・検証
b.全文テキストデータを構造化するために付与するメタデータ(以下「構造化メタデータ」という。)のフォーマット(以下「構造化メタデータフォーマット」という。)の検討・定義・検証
c.OCR、共同校正機能、共同構造化機能の連携システムの構築
d.標準中間フォーマットから全文テキスト化された書籍を、利用者に見やすい形で表示するための国立国会書館指定の複数のフォーマット(以下「電子書籍フォーマット」という。)への出力システムの構築
e. 出版社から提供される電子書籍データや版下データを標準中間フォーマットに変換するシステムの構築

(2) 共同作業支援システム(共同校正機能及び共同構造化機能)によるコスト削減等の効果の検証
a.共同校正機能による校正作業の効率化、高度化の検証
b.共同構造化機能による構造化作業の効率化、高度化の検証

(3) 半自動校正システム、半自動構造化システムによるコスト削減等の効果の検証
a.OCRの再学習システムの構築とそれによる校正作業の効率化、高度化の検証
b.構造情報(例:見出し情報)を半自動的に付加する機能と、それによる構造化作業の効率化、高度化の検証
c. 読み上げ順序の視覚化システムの構築と構造化作業の効率化、高度化の検証

システム機能に係る実現方式
システム全体構成

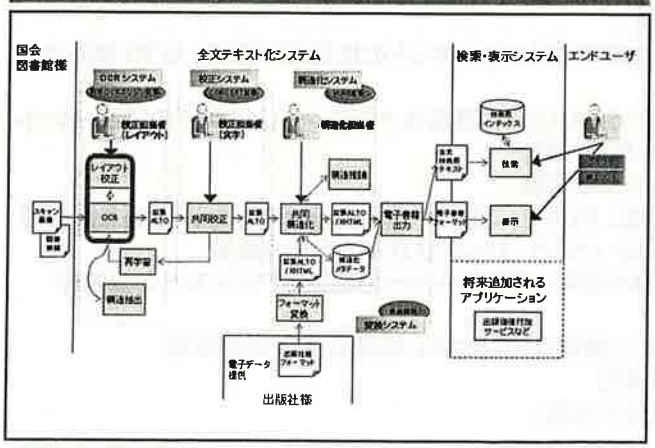
- 全文テキスト化プロトタイプでのシステム全体構成を以下に示します。



システム機能に係る実現方式
OCR機能
機能概要

- 全文テキスト化プロトタイプで提供する「OCR機能」のプロトタイプの目的、機能の実現方式、機能概要は以下の通りです。

OCR機能

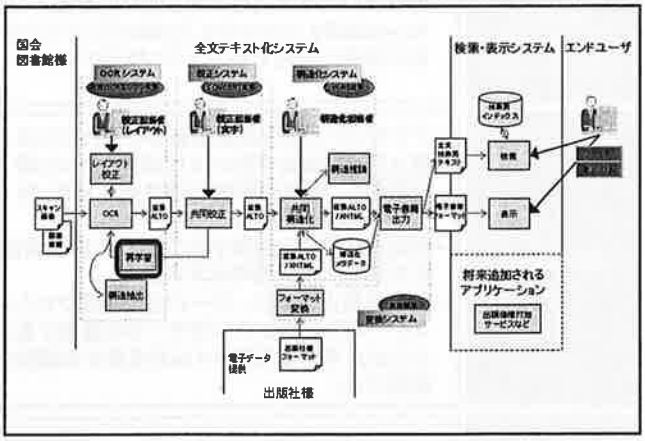


プロトタイプの目的	・資料データのデジタル画像から、テキストファイルへの変換を行うOCR処理において、高精度なレイアウト認識と文字認識を実現する
機能の実現方式	・メディアドライブ社のOCRソフトウェア製品である、WinReaderエンタープライズをエンジンとしてOCR機能を実現する。
機能概要	・高精度OCR認識の実現（JIS第1水準・JIS第2水準） ・高度なレイアウト解析能力を保有 ・文字当たりの認識精度は国内最高水準の99.53%を誇る（メーカー公称値） ・サーバ型OCRエンジン製品の採用による並列処理の実現 ・クライアント画面はWindowsのリモートデスクトップ機能を通じて利用することで、任意のPCからの操作環境を提供

システム機能に係る実現方式
OCR再学習機能
機能概要

- 全文テキスト化プロトタイプで提供する「OCR再学習機能」のプロトタイプの目的、機能の実現方式、機能概要、省力化のための工夫は以下の通りです。

OCR再学習機能

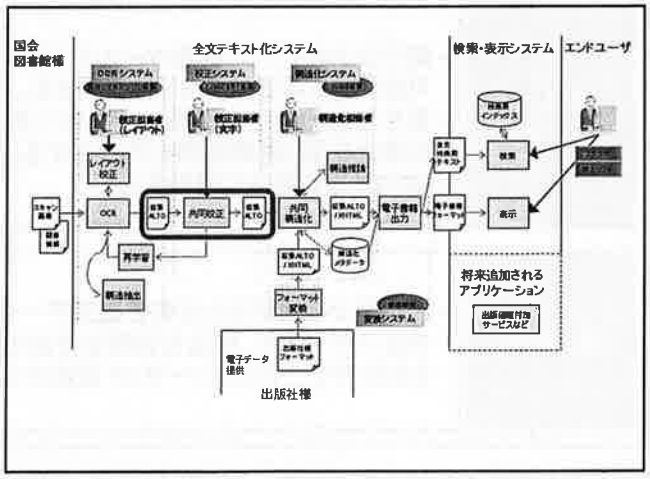


プロトタイプの目的	・共同作業者が実施する校正作業を通じて、文字の誤認識時の校正データを学習し、OCRの認識率を向上させる。
機能の実現方式	・メディアドライブ社のOCRソフトウェア製品である、WinReaderエンタープライズが実装している辞書自動学習機能を利用する。
機能概要	・共同校正機能において発生する校正データを用いて、OCR辞書を自動学習させる。 ・OCR再学習のための辞書データについて標準フォーマットを策定し準拠する。 ・メディアドライブ社の独自技術である自動学習機能を用いることで、複数人が校正作業を行った結果を自動的にOCR辞書に学習可能とする。
省力化のための工夫	・OCRの辞書(読み取った情報を文字として識別するために参照する辞書)を自動更新。(自動学習機能は国内製品ではメディアドライブ社のみ実現)

システム機能に係る実現方式
共同校正機能
機能概要

- 全文テキスト化プロトタイプで提供する「共同校正機能」のプロトタイプの目的、機能の実現方式、機能概要、校正作業の分担・共同化の実現レベルは以下の通りです。

共同校正機能

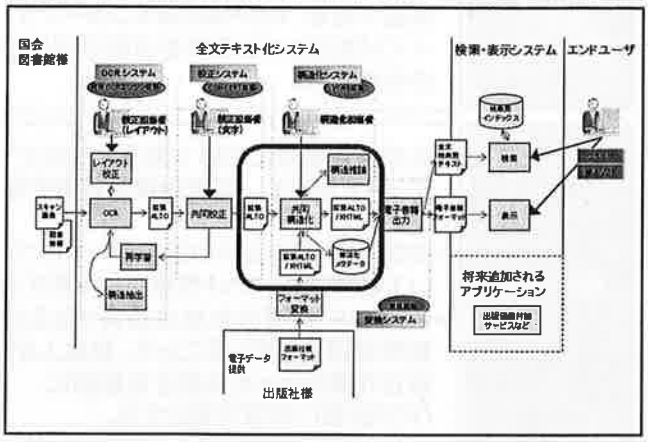


プロトタイプの目的	・日本語特有のレイアウトに対応し、複数人で並行して共同校正が可能なシステムを実現し、その効果を検証する。
機能の実現方式	・海外で多数の電子化プロジェクトでの利用実績があるCONCERT (COoperative eNgine for Correction of ExtRacted Text) (*)を元に日本語の共同校正システムを構築する。 (*) CONCERTの電子化プロジェクトでの利用実績 ・欧州連合(EU)の電子化プロジェクトIMPACT ・UCLAのニュース映画のラベルの電子化 ・オーストラリア、イスラエルの国勢調査のフォーム ・USの税申告のフォーム等 がある。
機能概要	・OCR機能によってテキスト化された標準中間フォーマットのデータを、複数の構成担当者が分担・共同して校正できる。
校正作業の分担・共同化の実現レベル	・同時に同一ページを編集(校正作業)できるのは一人のみとする。 ・同時でなければ、同一箇所を複数人が編集できるようにする。

システム機能に係る実現方式
共同構造化機能
機能概要

- 全文テキスト化プロトタイプで提供する「共同構造化機能」のプロトタイプの目的、機能の実現方式、機能概要、構造化作業の分担・共同化の実現レベルは以下の通りです。

共同構造化機能

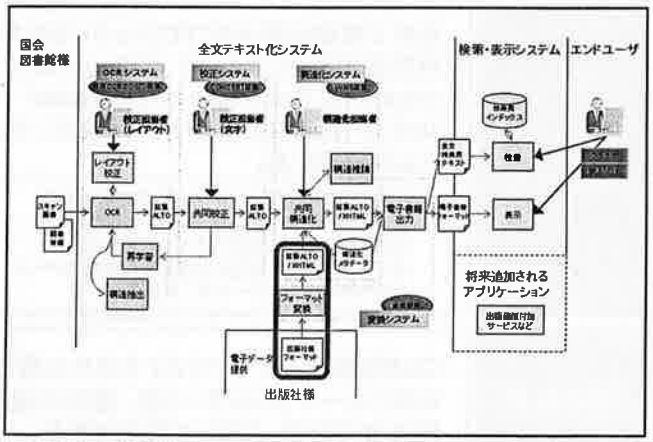


プロトタイプの目的	• 電子書籍のアクセシビリティを最低限確保するために必要な見出し情報や読み上げ順序といった構造情報を、効率よく付加できるような仕組みが実現可能であることを実証する。
機能の実現方式	• 弊社ソフトウェア資産であるWAIS(Web Accessibility Improving System)システムを電子書籍に対応可能なように拡張する。
機能概要	• アクセシビリティを最低限確保するために必要な見出し情報や読み上げ順序といった構造情報を、複数の構造化担当者が分担・共同して構造化できる。 • 構造推論機能と連携することで、一部の構造化作業を大幅に効率化する。 • 電子書籍出力時に、サーバ上で中間ファイルフォーマットに対して、メタデータを適用することにより、電子書籍ファイルを生成する機能を実現する。
構造化作業の分担・共同化の実現レベル	• 同時に同一ページを編集(構造化作業)できるのは一人のみとする。 • 同時でなければ、同一箇所を複数人が編集できるようにする。

システム機能に係る実現方式
電子書籍フォーマット変換機能 (1/2)
機能概要

- 全文テキスト化プロトタイプで提供する「電子書籍フォーマット変換機能」のプロトタイプの目的、機能の実現方式、機能概要は以下の通りです。

電子書籍フォーマット変換機能

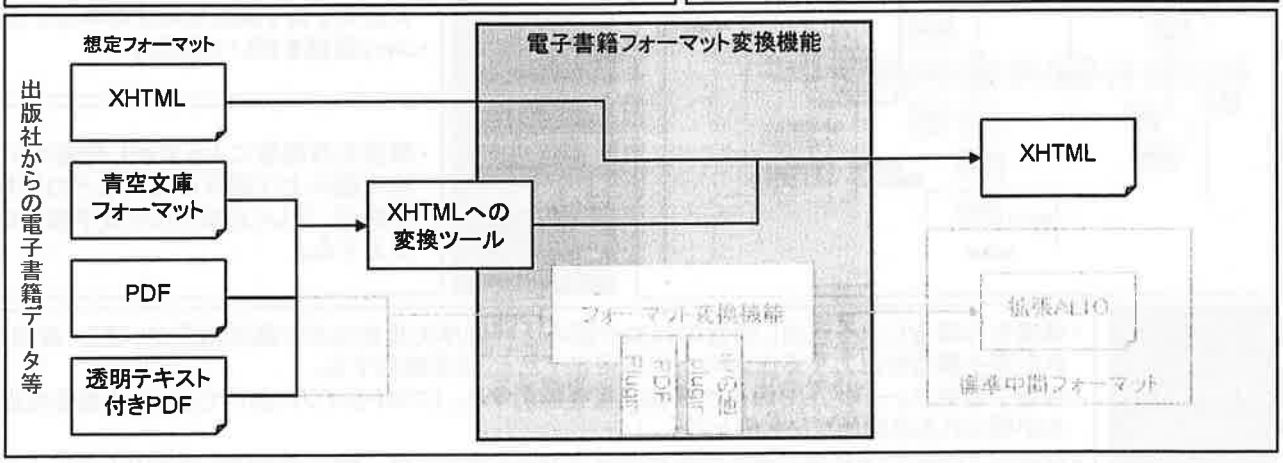


プロトタイプの目的	• 出版社から提供される電子書籍データを、構造化情報を付加できる形のデータにフォーマット変換する仕組みが実現可能であることを実証する。 • 各電子書籍フォーマットでの提供への対応の難易度を確認する。 • 適切なフォーマット変換ツールが利用可能である場合には、そちらを利用し、実用化に際しての有効性を検証する。
機能の実現方式	• 適切なフォーマット変換ツールが利用可能である場合はそちらを利用する。 • 電子書籍出力機能の実装方式に一部共通化を図り開発期間を効率化する。 • Java言語を用いて開発する。
機能概要	• 出版社から提供される電子書籍データや版下データを、構造化情報を付加できる形のデータにフォーマット変換する。

システム機能に係る実現方式
電子書籍フォーマット変換機能 （2/2）
機能概要

- 「電子書籍フォーマット変換機能」の入力、出力、実装イメージは以下の通りです。

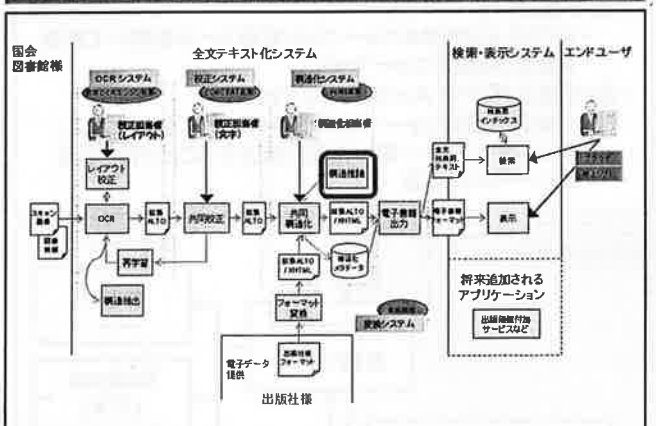
入力	出力
• 出版社から提供される電子書籍データや版下データ等 　• PDF、XMLベースの電子書籍フォーマット及びXHTML変換可能なフォーマットを想定 ※上記以外の形式のファイルフォーマットについても、期間に応じて積極的に対応を検討する。	• 構造化情報を付加できる形のデータ 　• 拡張ALTO（標準中間フォーマット） 　• XHTML



システム機能に係る実現方式
構造推論機能
機能概要

- 全文テキスト化プロトタイプで提供する「構造推論機能」のプロトタイプの目的、機能の実現方式、機能概要は以下の通りです。

構造推論機能

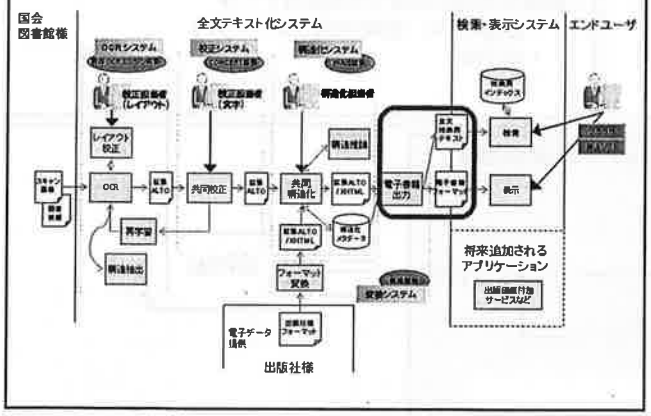


プロトタイプの目的	• 構造化作業を半自動化する仕組みとして、テキスト中の構造情報を推論する機能を実現し、それによって、どの程度の省力化が可能となるかを検討する。
機能の実現方式	• ユーザインタフェースに関しては、既存のWAISシステムの一部を拡張する。 • 共同構造化システムの一部として、構造化作業の一連の流れの中で適宜、バックグラウンドで実行されるものとする。 • 推論ロジックに関してはJava言語を用いて新規に開発する。
機能概要	• テキスト中の構造情報を推論し、共同構造化機能において、該当箇所を提示、一括での編集を可能とする。 • 推論する構造としては、アクセシビリティの観点から重要な見出し構造を対象とする予定である。

システム機能に係る実現方式 電子書籍出力機能 (1/2) 機能概要

- 全文テキスト化プロトタイプで提供する「電子書籍出力機能」のプロトタイプの目的、機能の実現方式、機能概要は以下の通りです。

電子書籍出力機能



機能の実現方式	• 適切なフォーマット変換ツールが利用可能である場合はこれを利用する。 • 電子書籍出力機能の実装方式に一部共通化を図り開発期間を効率化する。 • Java言語を用いて開発する。
機能概要	• 構造化作業者による見出し情報の付加や読み上げ順序の指定などの編集結果が、正しく反映された電子書籍を出力する。

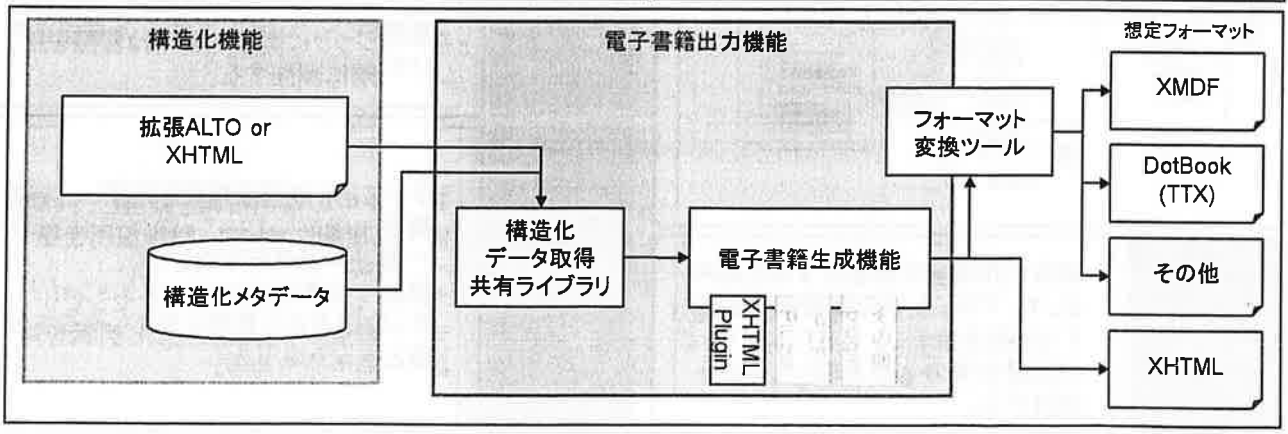
プロトタイプの目的	• 構造化作業者による見出し情報の付加や読み上げ順序の指定などの編集結果が、正しく反映された電子書籍を出力する仕組みが実現可能であることを実証する。 • 各電子書籍フォーマットでの出力の難易度を確認する。(プロトタイプにおいて必ずしも完全な出力が得られる必要はないものとする。) • 既存のフォーマット変換ツールの使用が妥当である場合には、そちらを利用し実用化に際しての有効性を検証する。

システム機能に係る実現方式 電子書籍出力機能 (2/2) 機能概要

- 「電子書籍出力機能」の入力、出力、実装イメージは以下の通りです。

入力	出力

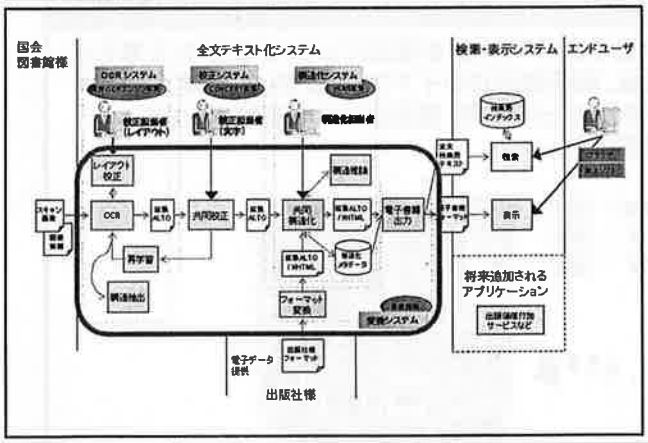
• 標準中間フォーマット(拡張ALTO)ファイル、または、XHTML ファイル
• 構造化メタデータ
  • OCRの読取結果に間違いが見つかった場合でも、既に行った構造化作業を再度やり直さなくてすむような構造化メタデータの設計を目指す
	• 電子書籍フォーマット
	  • XHTMLより既存のフォーマット変換ツールを用いて変換可能な6種類のフォーマット
	• 全文検索用テキストフォーマット
	  • 上記電子書籍フォーマット6種類のいずれか、もしくは、そのいずれかの一部の情報を抽出することで得られるフォーマットを想定



システム機能に係る実現方式
ワークフロー管理機能
機能概要

- 全文テキスト化プロトタイプで提供する「ワークフロー管理機能」のプロトタイプの目的、機能の実現方式、機能概要は以下の通りです。

ワークフロー管理機能

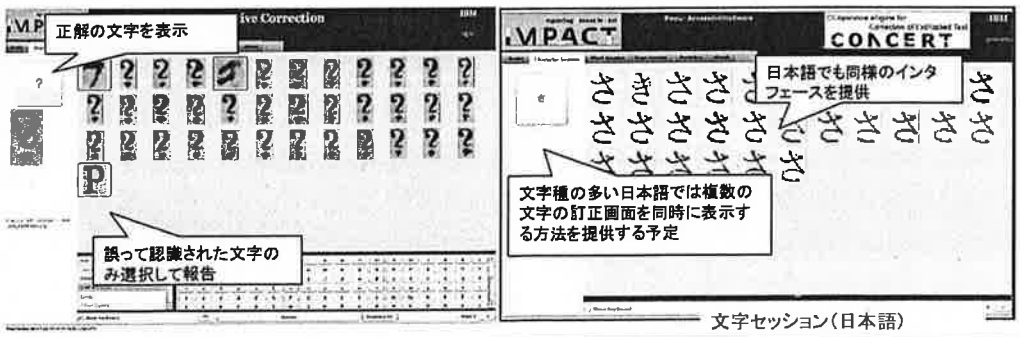


機能の実現方式	・弊社特許「コンピュータ・システムにより実行されるアクセシビリティ・メタデータの作成・拡張・検証を支援する方法」（特願2009-168195）を参照し、校正、構造化の作業割り振り機能を開発。
機能概要	・全文テキスト化システムで取り扱うデータの集中管理を行う。（詳細は次ページ参照） ・書籍プロジェクト管理 ・レポジトリ管理 ・ワークフロー管理
プロトタイプの目的	・管理者及び複数の編集者が書籍校正及び構造化作業を共同作業により効率的に行うことができることを実証する。 ・共同編集作業においてどのような課題があるかを評価することにより、本格稼動システムのセキュリティ要件等を事前調査する。 ・どの程度のシステム処理能力が必要とされるかを測定することにより、必要なシステム規模等を事前調査する。

システム機能に係る付加価値提案
利用者の操作性や利便性の向上に資する工夫

- 今回の全文テキスト化プロトタイプにおいて、利用者の操作性や利便性の向上に資する工夫を以下の通り挙げさせて頂きます。

機能分類	利用者の操作性や利便性の向上に資する工夫
共同校正機能	文字単位の修正機能として、文字セッションを提供し入力エラーが少なく高速に校正ができるインタフェースを提供する。



IBM

システム機能に係る付加価値提案
付加価値の創出を可能とする工夫

- 今回の全文テキスト化プロトタイプにおいて、付加価値の創出を可能とする工夫を以下の通り挙げさせて頂きます。

機能分類	付加価値の創出を可能とする工夫
OCR機能	メディアドライブ社の独自技術であるレイアウト解析機能により、非定型の文章を自動的に解析し、文字領域、表領域、画像領域のレイアウトを解析して分離可能とする。これにより文章のレイアウトを含めた再構築、高精度な文字認識を実現し、元原稿の再現性を高めることに貢献する。  本機能を有効活用することで、今回の弊社提案におきましてはデジタル化対象資料として「雑誌原資料（～2000年）」を加えて、“④戦後期雑誌（～2000年）のデジタル化画像”のテキスト化実現性検証作業を行います。

HITACHI
Inspire the Next

資料２－３

国立国会図書館殿

「全文検索・表示システムプロトタイプ構築等作業」の提案概要

平成22年10月29日
株式会社　日立製作所

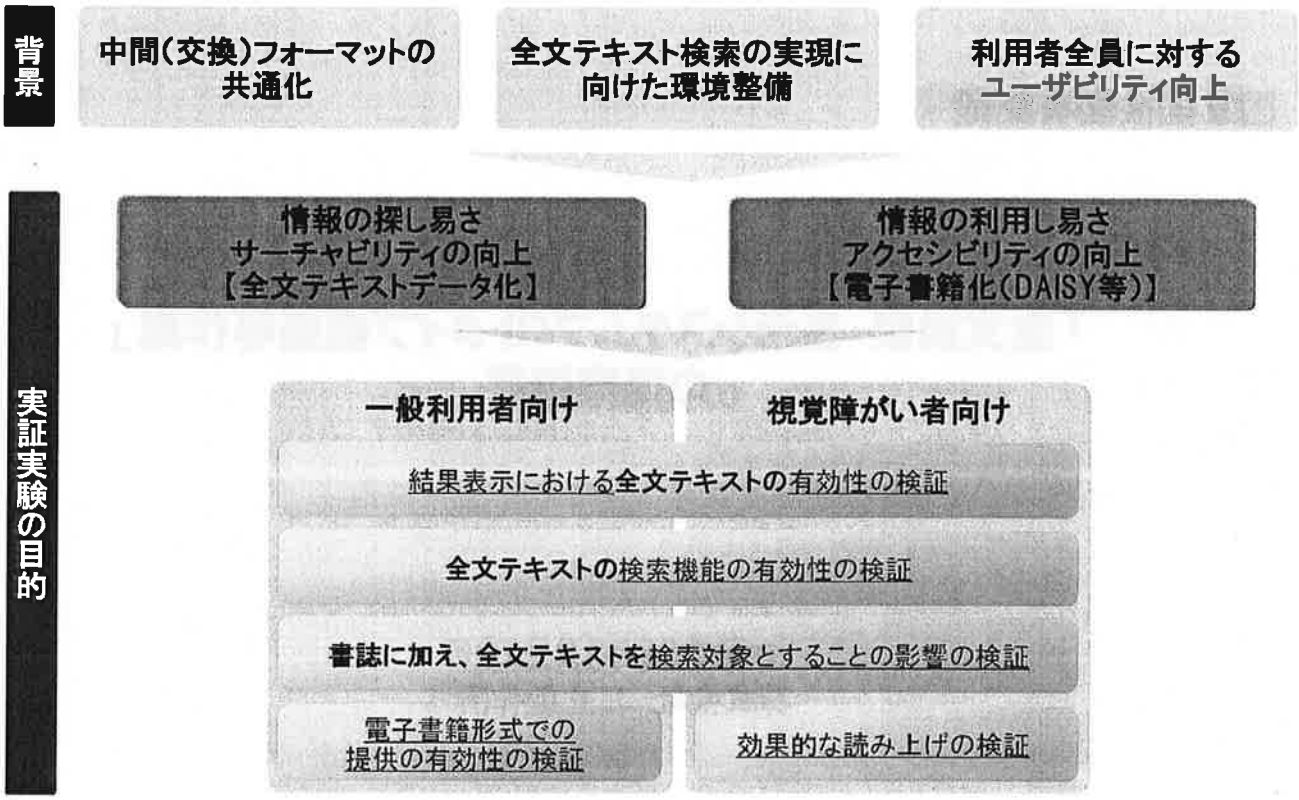
HITACHI
Inspire the Next

「全文検索・表示システムプロトタイプ構築等作業」の提案概要

Contents

1　本実証実験の目的

HITACHI
Inspire the Next



HITACHI
Inspire the Next

「全文検索・表示システムプロトタイプ構築等作業」の提案概要

2　検索・表示に関する機能

2　検索・表示機能

HITACHI
Inspire the Next

本実証実験での提案機能を以下に示します。

#	画面	提案機能	実現方法
1	検索画面	自然文検索	任意の文章を検索条件として検索を行い、文章の内容が似ている書籍を取得
2		構造指定検索	テキストの構造を活用して、検索範囲を限定して検索
3		サジェスチョン	検索式の入力時に検索語を補完 入力語に関連する別の検索語を提示
4		難易度検索	検索条件の一つとして書籍の難易度を選択
5	検索結果一覧画面	もしかして検索	検索語の綴りの誤りをチェックして、スコア順に並び替える
6		ランキング	書誌と全文テキストのデータの特徴を考慮し、書誌に適度な重みをつけて スコアを算出
7		連想検索	選択した資料群から特徴語を抽出し、連想検索エンジンを使って類似資料を検索
8		スニペット	全文検索エンジンの機能を利用して、検索キーワードの前後の文脈を 全文テキストから取得して表示
9		難易度表示	本テキストから本文の難易度を判定した結果を表示

2　検索・表示機能

HITACHI
Inspire the Next

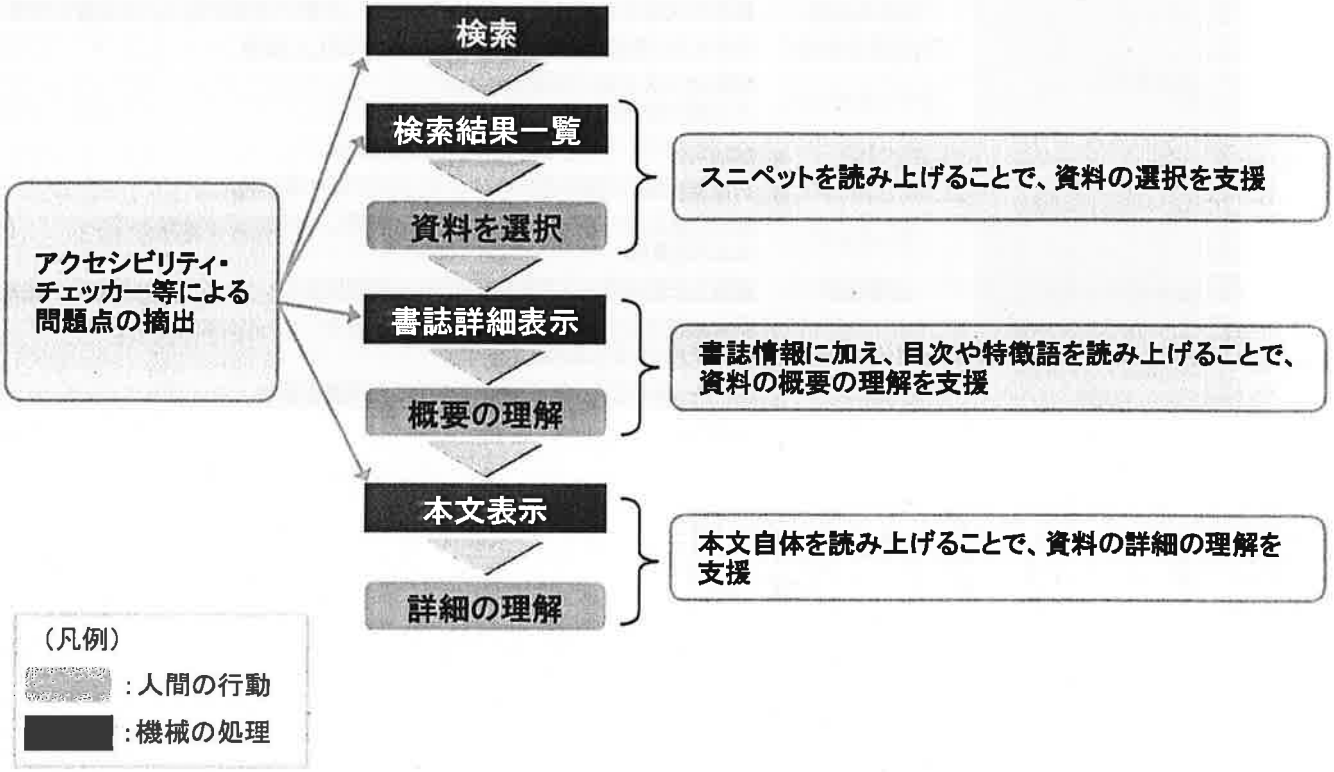
本実証実験での提案機能を以下に示します。

#	画面	提案機能	実現方法
10	書誌詳細画面	目次表示	全文テキストまたは情報探索プロトタイプのメタデータから抽出して表示
11		タグクラウド	連想検索エンジンを利用して全文テキストの特徴的な単語を取得して表示
12		内容ベースの レコメンド	参照した資料の書誌IDを元に連想検索エンジンを使って類似資料を検索し、 お薦め資料を表示
13		固有名表示	典拠データやWikipediaの情報から作成した固有名表現辞書を利用し、 全文テキストから固有名を抽出して表示
14		参考文献リンク	本文中に記載されている文献を書誌情報として表示
15		文脈検索	全文検索エンジンの機能を利用して、リアルタイムに全文テキストから 検索語を含む文脈を取得して表示
16	本文表示画面	目次・本文 リンク	全文テキストの中の目次に設定されたリンクを利用し、本文該当箇所へ移動
17		検索語可視化	全文テキストの構造ごとに検索語の出現数を数え上げ、リアルタイムに表示
18		ハイライト表示	検索語との一致箇所を全文テキスト中から検索してリアルタイムに表示

2-1　アクセシビリティへの配慮

HITACHI
Inspire the Next

検索から電子書籍の視聴までの一連の操作における視覚障がい者等向けのアクセシビリティに配慮します。



2-2　検索画面

HITACHI
Inspire the Next

サジェスチョン

検索式の入力時に検索語を補完

構造指定検索

全文テキスト中の検索する範囲を指定

難易度検索

書籍の難易度の選択が可能

自然文検索

任意の文章を検索条件として検索を行い、文章の内容が似ている書籍を取得

2-3　検索結果一覧画面

HITACHI
Inspire the Next

もしかして検索

もしかして: 図書館

ヒット件数が0件の場合、検索語の綴りの誤りをチェックして、正しい検索語を表示

ランキング

書誌と全文のデータ量を考慮してスコア付けし、検索語に関連の強い順に資料を表示

スニペット

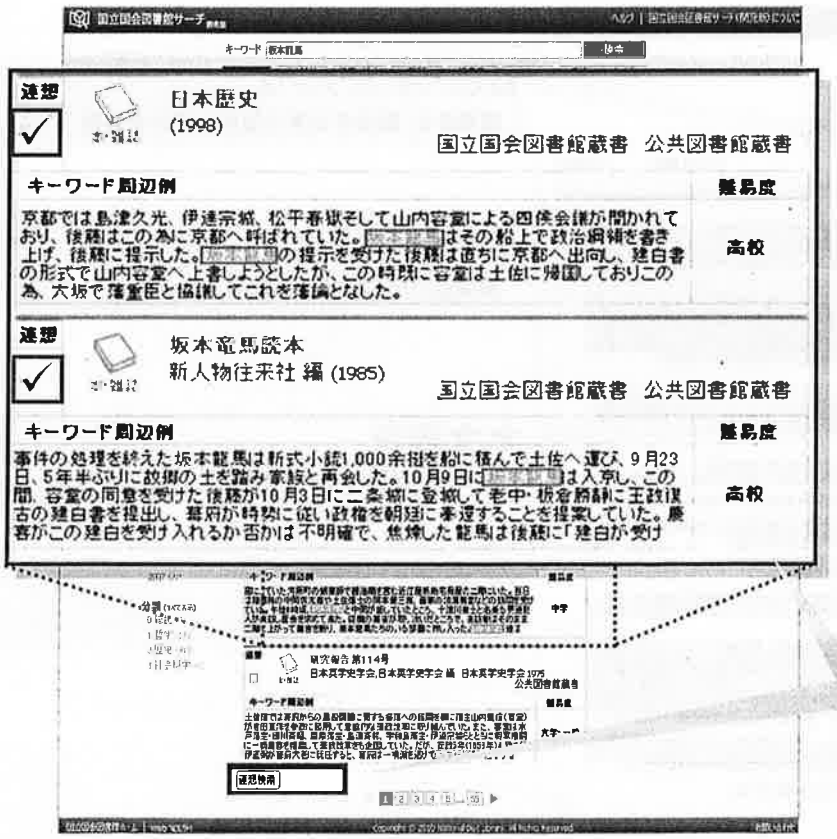
本文中の検索語前後の文章を表示

難易度表示

本テキストから本文の難易度を判定した結果を表示

2-4　検索結果一覧画面

HITACHI
Inspire the Next

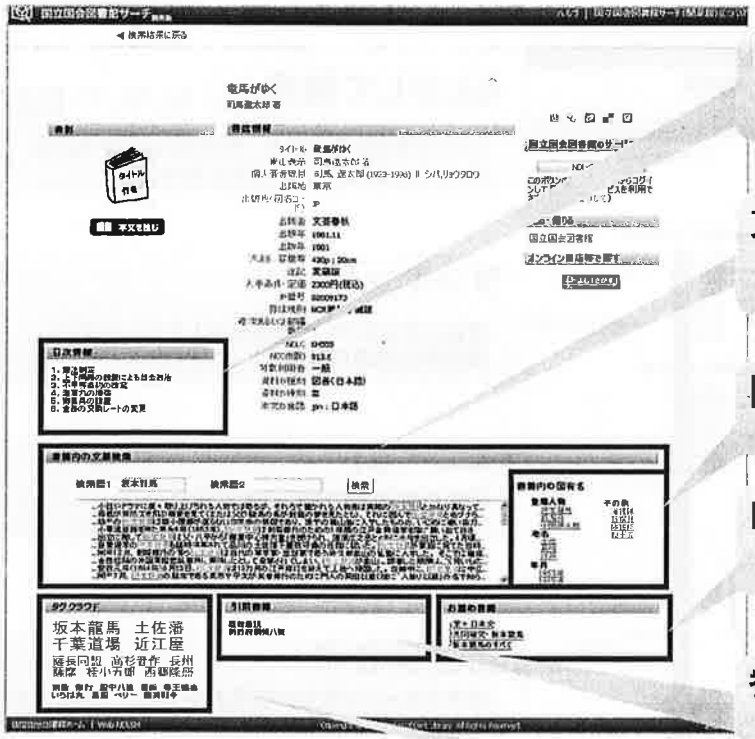


連想検索

資料を選択して似ている資料を検索

2-5　書誌詳細画面

HITACHI
Inspire the Next



目次表示

全文テキストまたは情報探索プロトタイプのメタデータから抽出して表示

文脈検索

任意の検索語を入力して、どのような文脈で検索語が使われているか表示

固有名表示

本文によく登場する人物名や地名を表示

内容ベースのレコメンド

参照した資料の書誌IDを元に連想検索エンジンを使用して類似資料を検索し、お薦め資料を表示

参考文献リンク

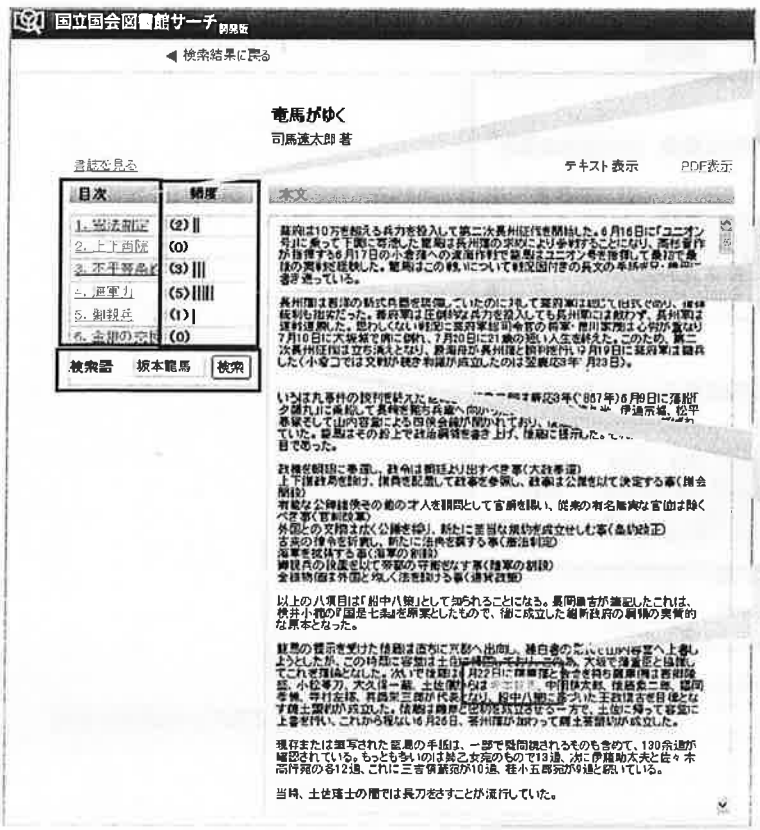
本文中に記載されている文献を書誌情報として表示

タグクラウド

本文中の特徴的なワードを表示

2-6　本文表示画面

HITACHI
Inspire the Next



目次・本文リンク

目次から、該当する本文箇所へ表示を移動

検索語可視化

検索語の出現回数を可視化

本文検索

本文中を任意のワードで検索

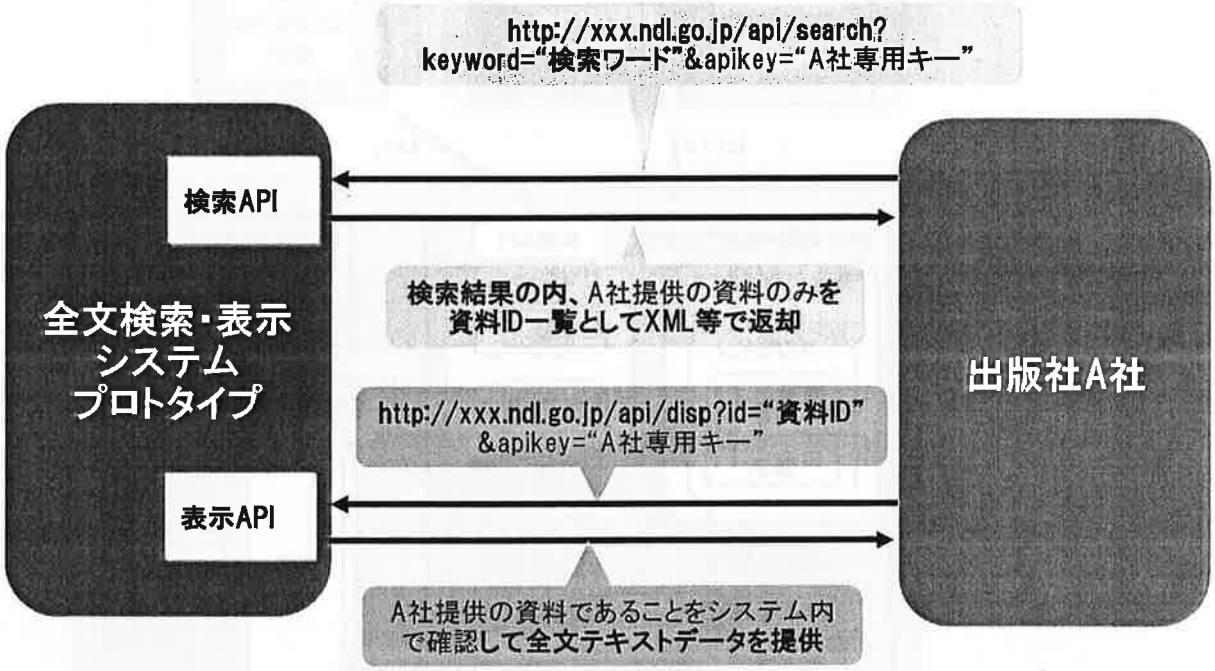
ハイライト表示

検索語をハイライト表示

2-7 出版社向け検索・表示API

HITACHI
Inspire the Next

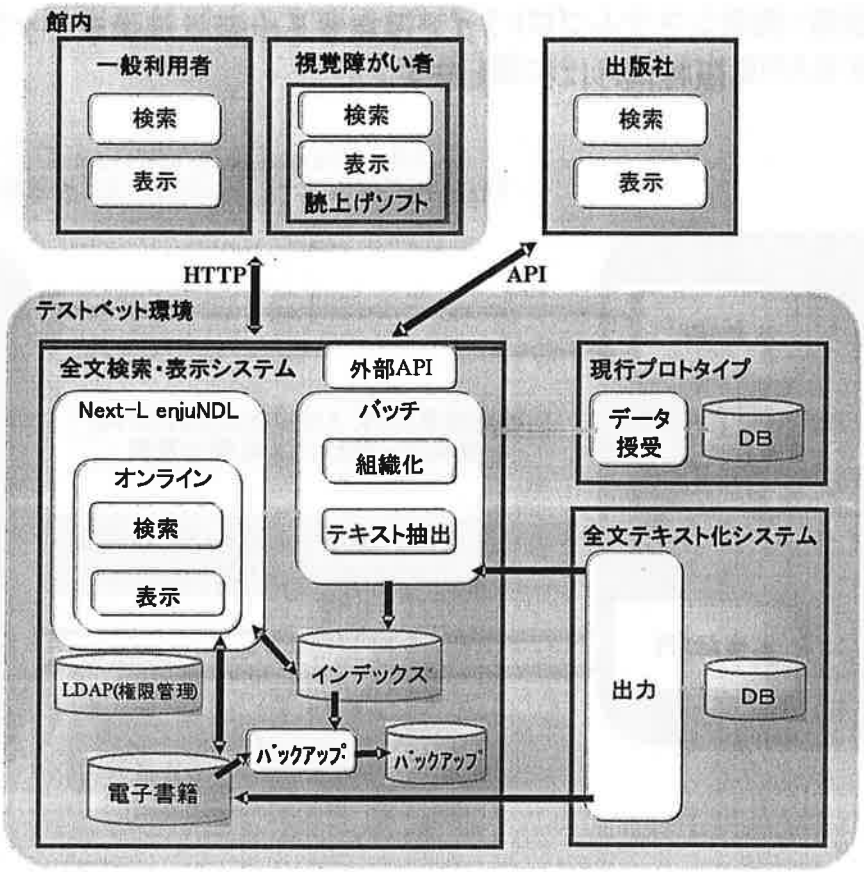
全文検索・表示システムプロトタイプで保有する出版社提供データを検索し、取得するAPIを出版社向けに提供します。



HITACHI
Inspire the Next

HITACHI
Inspire the Next

参考資料(全体構成)


館内
一般利用者
検索
表示
視覚障がい者
検索
表示
読上げソフト
出版社
検索
表示
HTTP
API
テストベット環境
全文検索・表示システム
外部API
Next-L enjuNDL
オンライン
検索
表示
バッチ
組織化
テキスト抽出
現行プロトタイプ
データ授受
DB
全文テキスト化システム
出力
DB
LDAP(権限管理)
インデックス
バックアップ
バックアップ
電子書籍