第19回図書館総合展

70周年からその先へ　～ユニバーサル・アクセスを目指して～
国立国会図書館の取組

次世代システム開発研究室の活動

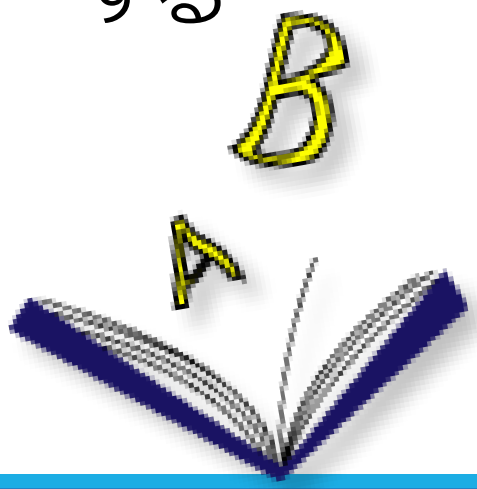
国立国会図書館電子情報部
電子情報サービス課次世代システム開発研究室
室長　原田久義

平成29年11月7日（火）

1. 次世代システム開発研究室（次世代室）の設置とその目的
2. 次世代室の体制
3. これまでの成果
4. 今後の課題・展望

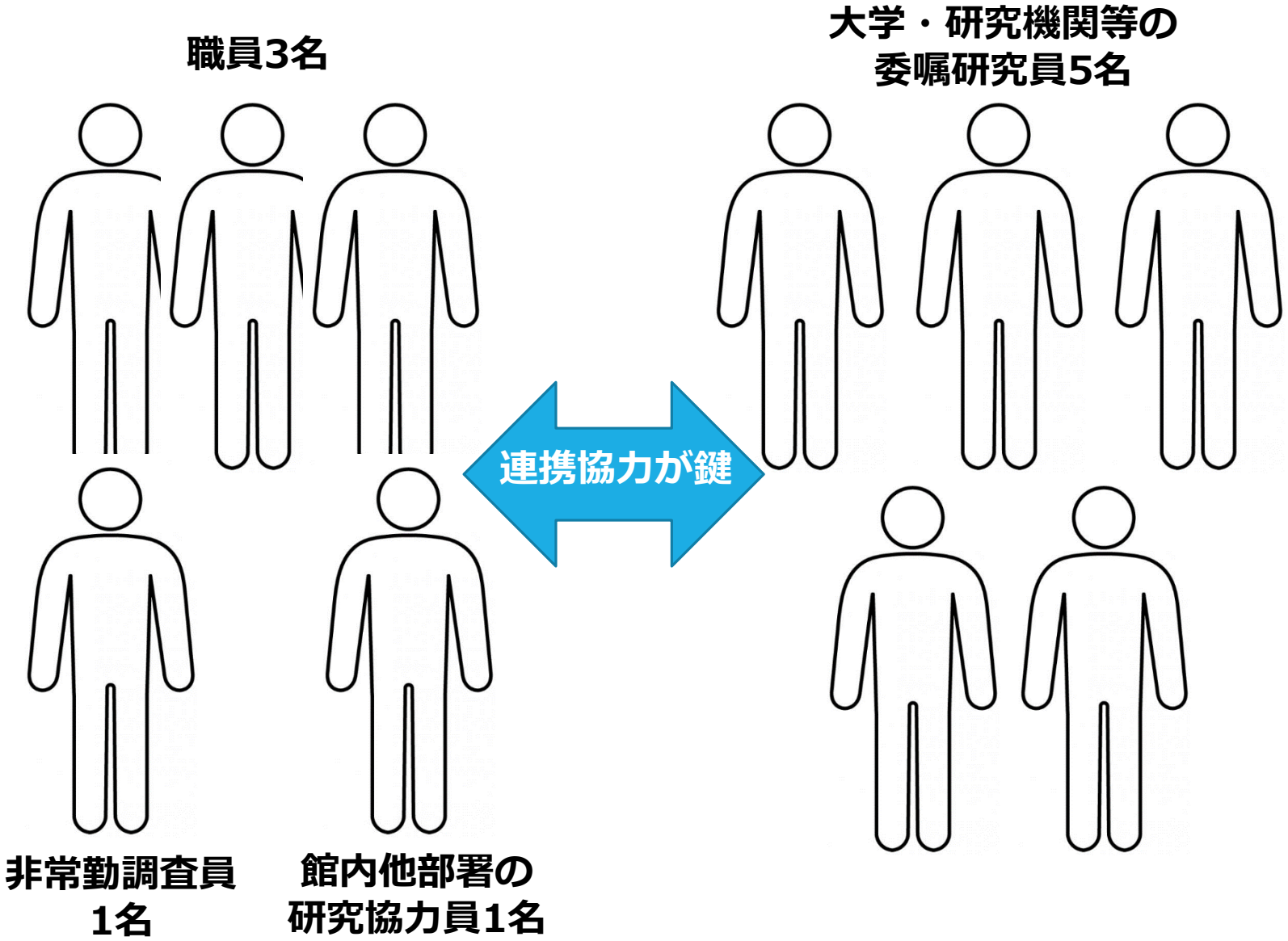
1．　次世代室の設置とその目的

- 2011年10月の機構改革における電子情報部の発足に伴い設置
- 先進情報技術を応用した新しい図書館サービスを実現するための次世代システムに係る調査研究・実証実験を行うことを目的とする



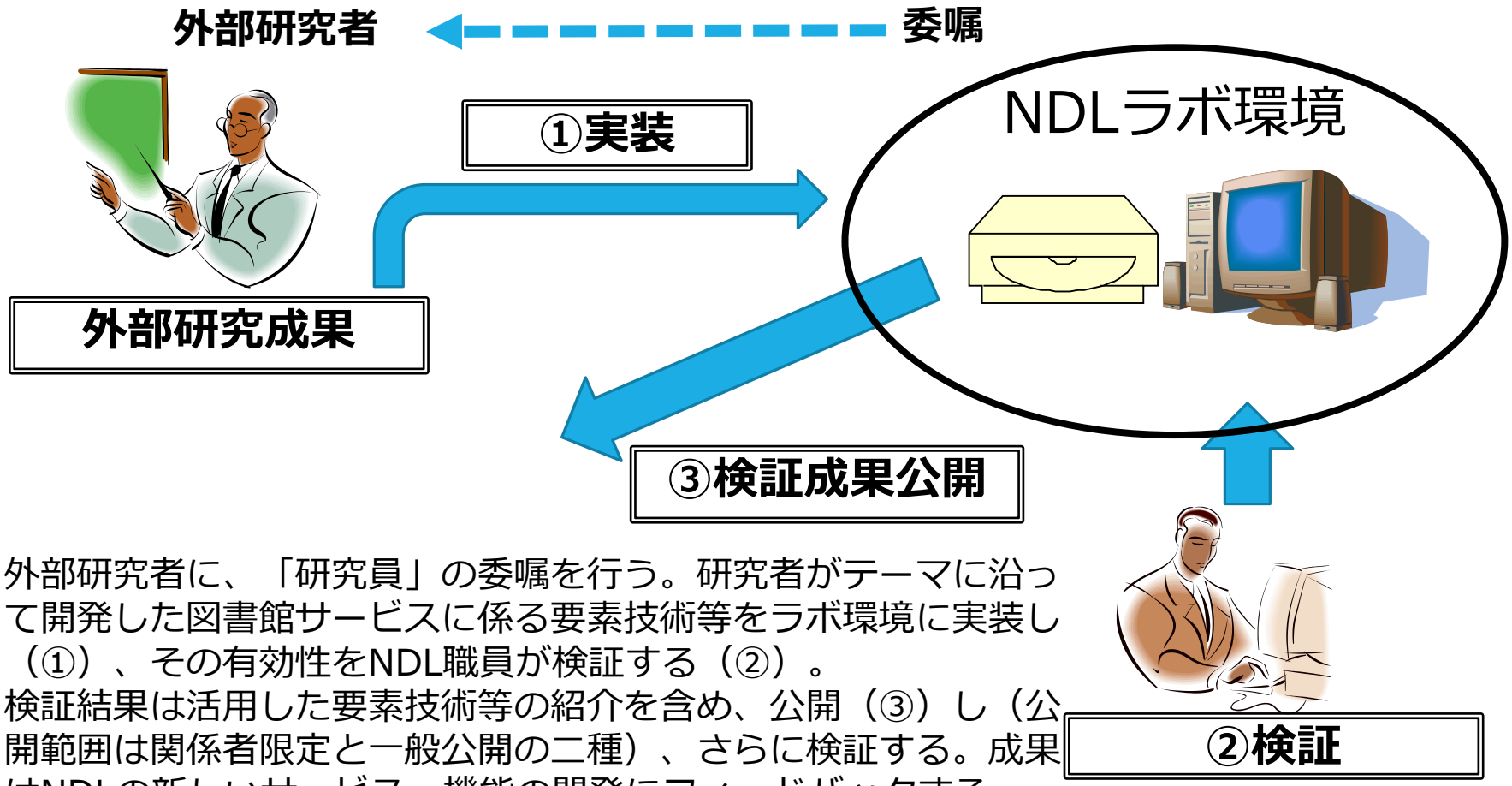
NDL Lab

2．　次世代室の体制



2.　これまでの成果

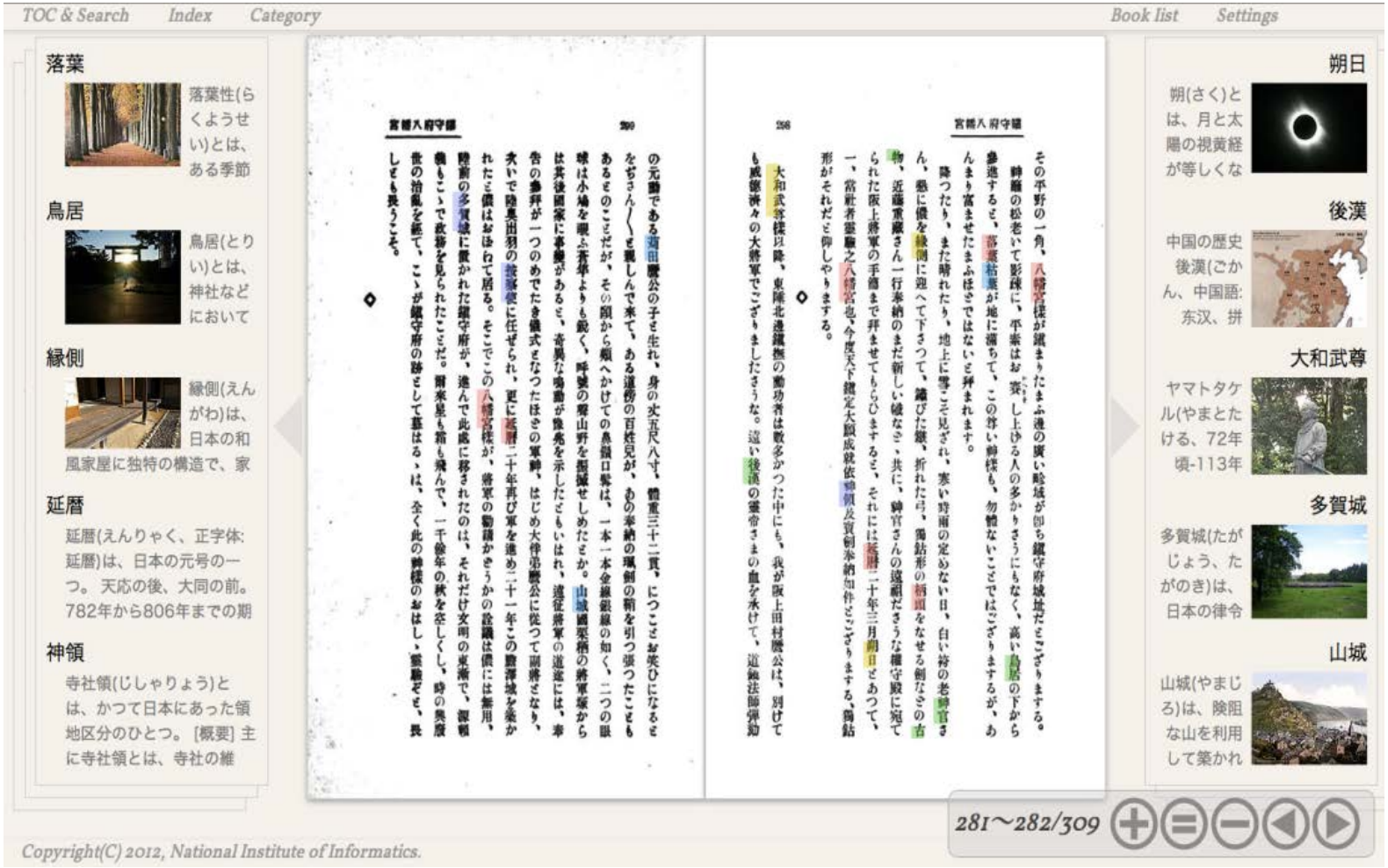
（１）NDLラボの開設



外部研究者に、「研究員」の委嘱を行う。研究者がテーマに沿って開発した図書館サービスに係る要素技術等をラボ環境に実装し（①）、その有効性をNDL職員が検証する（②）。
検証結果は活用した要素技術等の紹介を含め、公開（③）し（公開範囲は関係者限定と一般公開の二種）、さらに検証する。成果はNDLの新しいサービス・機能の開発にフィードバックする。

3.　これまでの成果

（2）電子読書支援システム



3.　これまでの成果

（２）電子読書支援システム

- 一定の文字情報を基に連想検索を行い、他の情報源（Wikipedia等）から関連する参考情報を候補として利用者に提示し、読書や調査の支援に資することを目的とする
- 対象コンテンツは、近代デジタルライブラリー（現・国立国会図書館デジタルコレクション）から著作権保護期間が満了した資料を選定
- 国立情報学研究所コンテンツ科学研究系阿辺川武特任准教授の協力を得て実施
- 2013年5月一般公開

3.　これまでの成果

（３）NDLラボサーチ

3.　これまでの成果

（３）NDLラボサーチ

- 当館作成書誌（図書、雑誌、新聞、記事等）を検索対象とし、国立国会図書館サーチの検索キーワードのオートサジェスト、検索結果に関連したタグクラウドによる絞り込み等の実験的な機能を追加したシステム。検索機能の高度化を企画し試行するための環境として使用
- 2013年9月一般公開

3.　これまでの成果

（４）翻デジ

3.　これまでの成果

（４）翻デジ

- 近代デジタルライブラリー掲載の著作権保護期間満了の資料を対象として、クラウドソーシングにより本文テキストの共同入力・校正作業を行うプラットフォーム
- 当館のデジタル化資料の更なる活用策として、多数のボランティアが本文テキストの入力・校正を行うことで、キーワードによる全文検索を可能とし、利用者の利便性を向上させることを目指す
- 人文情報学研究所永崎研宣主席研究員の協力を得て実施
- 2014年3月一般公開

3.　これまでの成果

（5）国デコ Image Wall

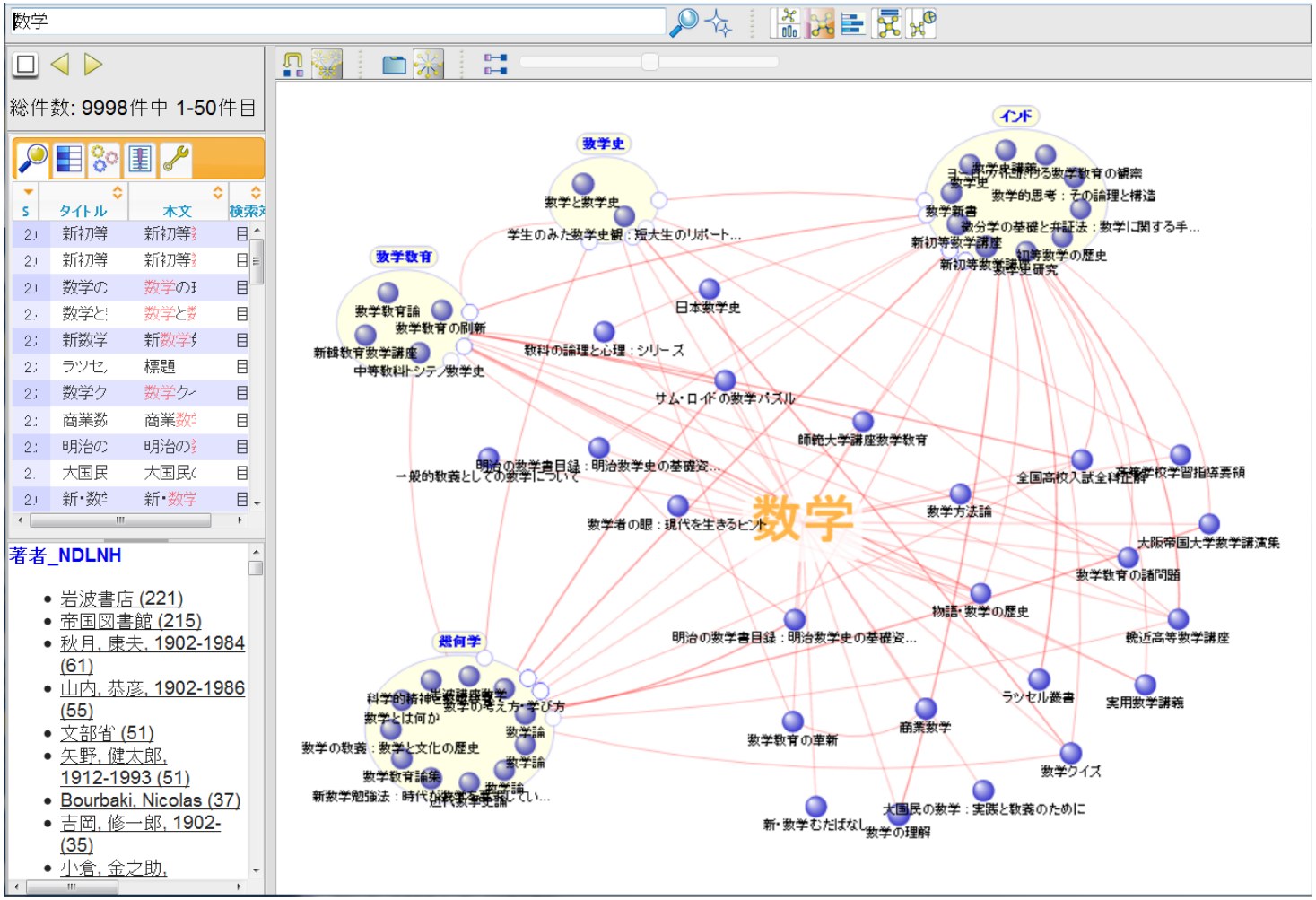
3.　これまでの成果

（５）国デコ Image Wall

- クラウドソーシングを活用して国立国会図書館デジタルコレクションから絵や図を含むページを抽出し、IIIFに対応した画像を一覧表示
- 人文情報学研究所永崎研宣主席研究員の協力を得て実施
- 平成28年8月に公開

3.　これまでの成果

（６）書誌情報検索・可視化システム



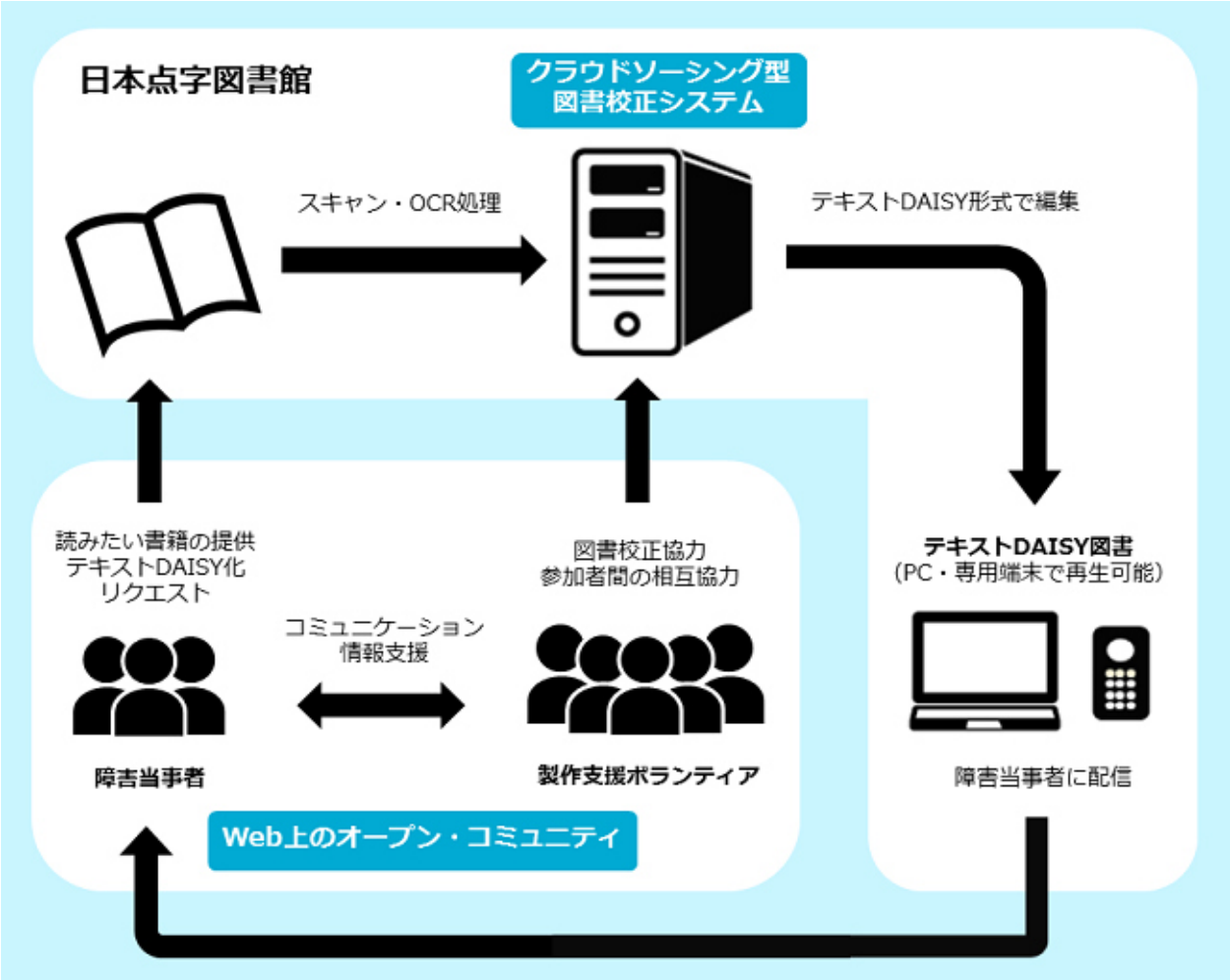
3.　これまでの成果

（６）書誌情報検索・可視化システム

- 美馬秀樹東京大学大学院工学系研究科特任准教授を中心として開発された、文書間の意味的な関連に基づいてデータマイニングを可能とする検索・可視化システム 「MIMAサーチ」を活用
- 当館でデジタル化した図書の書誌データ（約92万件）及び目次データ（約71万件）を対象に検索・可視化
- 2015年7月一般公開

3.　これまでの成果

（７）視覚障害者等へのテキスト化資料提供実験



3． これまでの成果

（７）視覚障害者等へのテキスト化資料提供実験

- 社会福祉法人日本点字図書館と当館が、平成27年3月に覚書を締結し、視覚障害者等へテキストデータを提供するため、共同校正システムを用いた実験を実施中
- 平成27年4月～6月にかけて当館デジタル化資料から5タイトルを選定し、テキストDAISYを製作した結果、テキストDAISY、EPUB、プレーンテキストの製作に同システムが有効であることを確認した
- 以後、日点が同システムを用いたテキストDAISYの製作を継続するとともに、日点以外の機関による同システムの利用の働き掛けを行っている

4. 今後の課題・展望

①ジャパンサーチ（仮称）

次世代の分野横断型検索・ナビゲーションサービスのプロトタイプを構築中

②研究データの保存と利活用

図書館における研究データの取り扱いについて、方向性を検討するための調査研究

③電子情報の長期利用保証

パッケージ系電子出版物のマイグレーション等の試行

④AI

AIを活用した書誌データ等の自動付与及びAIの当館の業務プロセス・サービスへの適用

4．　今後の課題・展望

（1）ジャパンサーチ（仮称）のプロトタイプ構築

「知的財産推進計画2017」等に基づき分野横断型検索サービスのプロトタイプを次世代室において構築中

①コンテンツの所在等の明確化

メタデータの横断的な検索機能を提供し、我が国の様々な分野のコンテンツの所在情報等を明らかにする

②データ提供機関へのアクセス促進

ジャパンサーチを経由して、個別のデータ提供機関へのアクセスにつなげる

③データの利活用の促進

ライセンスの明示、APIの提供等により、データ利活用を容易にする。研究・教育、防災、ビジネスでの活用に加え、海外発信機能の強化につなげる

④データ提供機関への支援

データ提供機関にとって便利な仕組みを提供することで、そのサービスやシステムを支援する

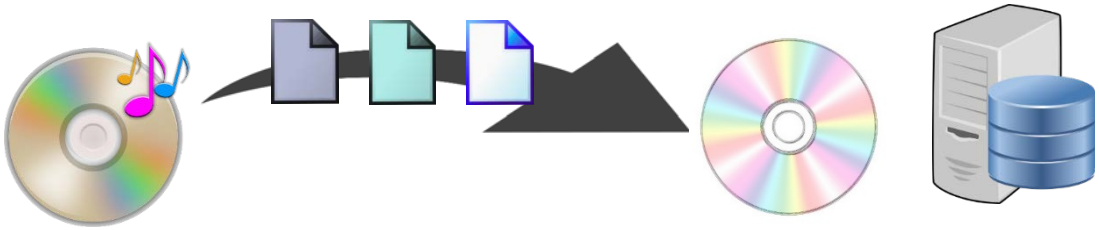
⑤新規ビジネス・サービスの創出等

以上を通じて、データに関する新規ビジネス・サービス創出及びデジタルコンテンツの保存・共有・活用を促進する

４．　今後の課題・展望

（2）パッケージ系電子出版物のマイグレーション等の試行

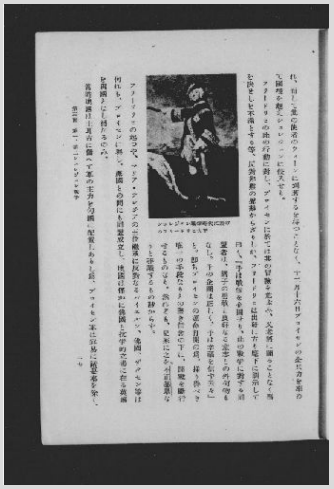
- 当館は、CDやDVDなどの「パッケージ系電子出版物」と呼ばれるデジタル形式の資料も、納本制度により多数収集している
- しかしながら、デジタル形式の資料には、媒体のぜい弱性、再生装置の入手困難化及び再生ソフトウェア等技術の陳腐化という課題がある
- 長期にわたる利用を保証するため、媒体の劣化などが原因で再生できなくなる前に、内部のデータを取り出し、必要であれば保存に適したフォーマットに変換し、別の媒体に移行するマイグレーション作業等を試行する



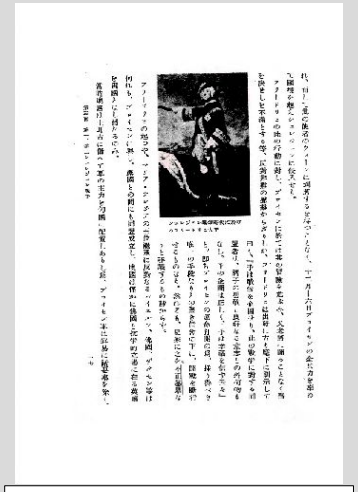
4． 今後の課題・展望-AIの活用-

（3）デジタル化資料の背景白色化の実験

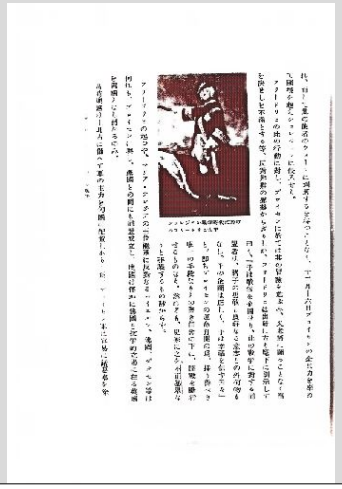
- デジタル化された書籍の画像を対象に、文字以外の背景部分を白くすることで可読性を上げることが目的
- コントラスト調整などの作業を人手で行った画像と元画像のペアを3000枚程度用意し、pix2pixの手法で学習を行っている



元画像



人手で修正
した画像



学習結果を
用いた修正画像

人手による修正内容をおおよそ再現できている

4． 今後の課題・展望-AIの活用-

（4）NDC(日本十進分類法)の自動付与

目的
「書籍タイトル等の書誌データから特徴を抽出して書籍のNDCを自動で付与する」

書誌データ

書籍タイトル:
「経済学および課税の原理」

NDC分類: 331.323

分かち書き & 名詞の抽出

「経済学」
「課税」
「原理」

word2vec(CBOW) で各単語を50次元にベクトル化

「経済学」→[-0.102, -2.15, …,4.45]
「課税」 →[1.15, -3.31, 2.47, …, 0.167]
「原理」 →[-0.979, -0.537, …, 0.180]

平均を取る

「経済学および課税の原理」を表す50次元のベクトル
→[0.0233, -1.99, 2.45, -3.20, …, 2.67, 1.60]

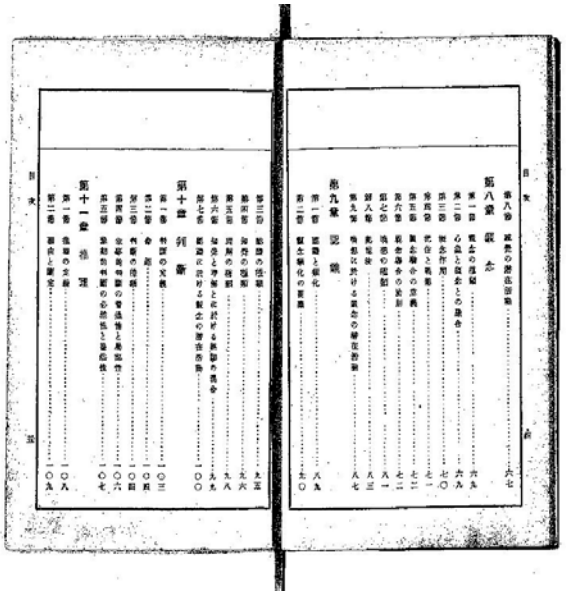
このベクトル表現を使って、NDCの分類番号を予測する

未知の書誌に対する精度と再現率は
上一桁(大分類)で70%程度
上二桁(中分類)で60%程度

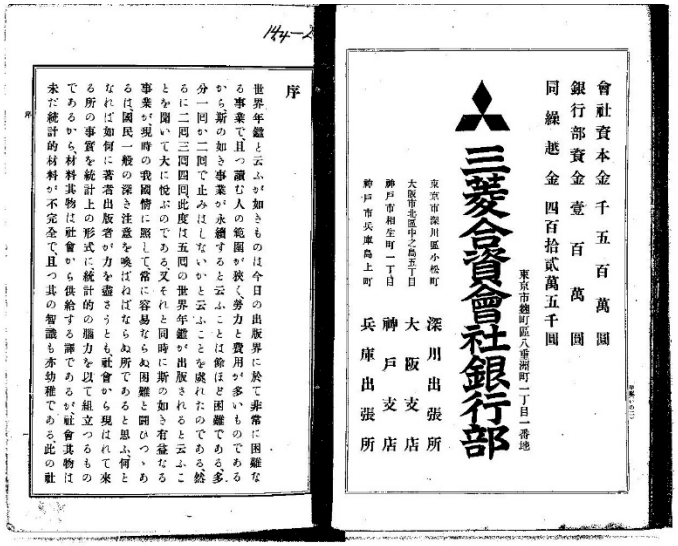
4.　今後の課題・展望-AIの活用-

（5）目次ページの自動認識

- 国立国会図書館デジタルコレクション所蔵資料から目次ページの自動認識を行う
- 目次/非目次ページを40000枚ずつ学習に使用
- テスト用データセットに対する精度は90%程度



目次ページの例



非目次ページの例

4.　今後の課題・展望

以上ご紹介した取組を通じて次の目標の達成を目指します。

- ナショナルアーカイブを支える恒久的保存基盤の具体化
- デジタルコンテンツ・データの利活用の促進
- 紙とデジタルを統合した業務・サービスの達成

ご協力・ご支援をお願いいたします。