

市販のツールを活用した、機関リポジトリにおける データ登録機能の運用改善案

外崎 みゆき、海尻 賢二

機関リポジトリにおける電子資料の提供において、全文検索機能は高く評価されている。しかし、一見同じように見える画像 PDF で登録された電子データの場合は全文検索の対象にならず、利用者が検索結果から漏れているが現状に気付く事が少ない。画像 PDF ファイルは市販のツールを利用して透明テキスト付の文字型 PDF に変換する事が可能であり、変換作業における作業負荷も少ない。市販のツールを使用して、画像 PDF ファイルを文字型 PDF に変換する仕組みを処理過程に組み込み、機関リポジトリにおいて、全文検索の対象となる電子資料を増やし、利便性を高める運用を提案する。

Improving the usability of the institutional repository system using off-the-shelf software

TOZAKI Miyuki, KAIJIRI Kenji

Full text retrieval is important for institutional repository system which is used to collect, preserve and disseminate the digital documents. However, a lot of PDF documents are submitted in graphical format. These PDF documents can not be used for retrieval and authors seldom recognize this fact. To solve this problem, one effective way is to use the common off-the-shelf software to convert graphical PDF files into searchable text version. In this paper, we will propose a solution to integrate the conversion into the text PDF into the data registration process. And this will expand the retrieval possibilities for these documents and improve the convenience of the system.

1. はじめに

現在、各大学では多くのコンピュータシステムを導入運用して業務支援、研究支援、教育支援などに使用している。それらのシステムの中には運用期間中に、社会情勢の変動に適応できなくなるケースや、インターネットを介した新しいサービスの要望が発生して、対応に苦慮するケースが見られる。

大学図書館においても例外ではなく、従来の資料の運用管理に加えて、インターネットを介したサービスやデジタルコンテンツの管理など、新しい要求への対応が急務となっている。一方で人員は削減の方向にあり、特に私立大学では業務の委託化傾向[1][2]が強く、新しいサービスを試行する事が難しい現状である。新機能を提供するにあたり、大学図書館が独自で開発を行う場合も、パッケージで提供されているシステムに対して、改修をおこなうことで独自の機能を追加する場合も費用が大きく、経費節減の中で実現は難しい。別の視点から捉えると、インターネット環境や情報関係の技術的進歩により、オープン・ソース、安価な市販ソフト、webAPI や標準プロトコルを利用するなど、機能内容も提供方法もさまざまではあるが、問題解決に利用できるシステム資源は増加している。また、実績のあるソフトウェアを利用することで、新規に開発する費用よりも安価に機能が追加できるケースも多い。最新の技術、市販やフリーのソフトウェアなど、資源を有効に活用する事で、システム運用を改良する方法が考案できれば、業務負荷の軽減やサービスの向上に貢献できる。

学術機関リポジトリ (Institutional Repository) とは、大学および研究機関で生産された電子的な知的生産物を捕捉し、保存し、原則的に無償で発信するためのインターネット上の保存書庫である(1)。

機関リポジトリに登録される電子資料(論文など)の多くは、印刷業者や出版社、または著者から提供される透明テキストが付与された文字型ファイルである。しかし、遡及目的や印刷物のみで提供される資料をスキヤニングで電子化した場合や、印刷機能のファイル出力を利用して PDF ファイルを作成した場合は画像 PDF ファイルとなる。二種類の PDF は見た目には分りにくいが、画像 PDF は本文が検索対象とならない。

画像 PDF ファイルを透明テキスト付の文字型 PDF ファイルに変換する事は安価な市販のソフトでも可能である。市販ツールを利用して画像の変換テストを行い作成される透明テキストを調査したところ、十分に利用に耐える精度が確認できた。現在、利用者または事務部署から機関リポジトリに論文を登録するために提供している web 版登録支援システムの処理過程に、文字型 PDF への変換処理を組み込むことで機能改善が可能になる。

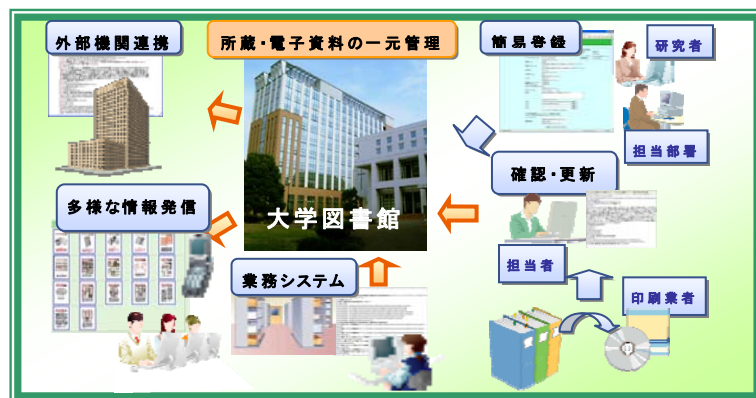


図1 機関リポジトリ運用イメージ

2. 登録支援システム

2.1 開発の経緯

人員減に悩む図書館において、新しいサービスの提供を行う場合、運用にともなう業務量の増加が大きな問題となる。機関リポジトリの構築においても例外ではない。運用において大きな作業ボリュームを持つのはコンテンツの収集と登録管理である。登録時の作業がしやすいようにシステム側でツールを提供する事で業務負荷が軽減されるように考慮する事とした。

本学では稼働当初より、各学部が定期的に発行している紀要を電子化して、機関リポジトリに収録している。

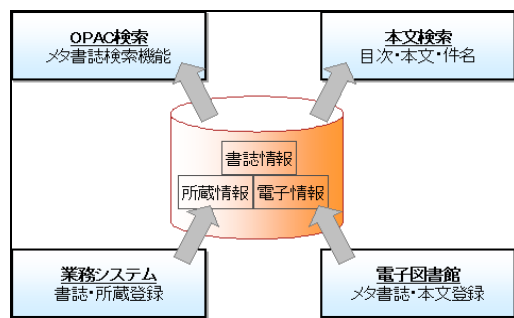


図2 機関リポジトリデータベース

登録は、印刷業者から納品される目次データに図書館で固有のデータを付与し、自動生成により仮書誌データの登録を行う。同時に、印刷業者から納品される論文電子データ(PDF)を登録し、リンク付け処理する仕組みである。

各学部の編集委員会の協力を得て、冊子体の紀要が納品される時にデジタルデータを手入れし、冊子体の巻号単位で登録するよう組織として基盤を整備してきた。システム支援の面では、冊子体単位でデータを登録する一括登録機能を提供している。

教員の自主的な参加(論文の登録)と利用は、機関リポジトリ事業を活性化させる重要な一環と考えている。そのためには、教員から直接登録できるインターフェースの整備が必要と思われた。つまり、研究者または担当部署より個別に登録するための環境である。これにはweb版の登録支援環境(簡易登録システム)を開発することで対応した。

簡易登録システムはweb環境を利用した電子資料の登録支援ツールである。電子データに必要なメタ書誌項目を書誌事項に詳しくない登録者がプルダウンメニューなどの簡便な選択操作で入力できる支援機能を完備している。

2.2 簡易登録システムの機能

簡易登録における処理は次の手順で行われる。まず、登録者はWebブラウザを通じて、メタデータ、本文(PDF、HTML、画像ファイル)を登録する。登録したメタデータ、本文は一時的にサーバ上で保管する。夜間バッチでデータをデータベースサーバに登録し、同時に仮書誌を作成する。登録結果はシステム管理者と目録業務担当者にメールで通知される。図書館の目録業務担当者は登録されたデータを確認、更新し、公開状態にして利用者に提供する。

2.3 簡易登録における課題

機関リポジトリに登録される電子資料の多くはPDF形式のファイルである。紀要論文などの学内発行物は印刷業者からPDFデータを手入れするため、文字型PDFを納品物として指定する事が可能である。

しかし、個別に登録される電子資料の中にはスキャナで画像処理を行い電子化された画像PDFが含まれる。

実際に検索を行った場合に、透明テキストの付与されていない画像PDFの場合は本文検索の対象にならない。書誌事項によるデータの切り出しのみでは、抽出率が低く、電子資料が埋没する危険がある。

本文検索が可能で多くの資源のなかに書誌事項のみでしか検索できない資料が混在する状況は好ましくない。今後の運用を考えると画像PDFファイルへの対処方法を早い時期から考える必要がある。

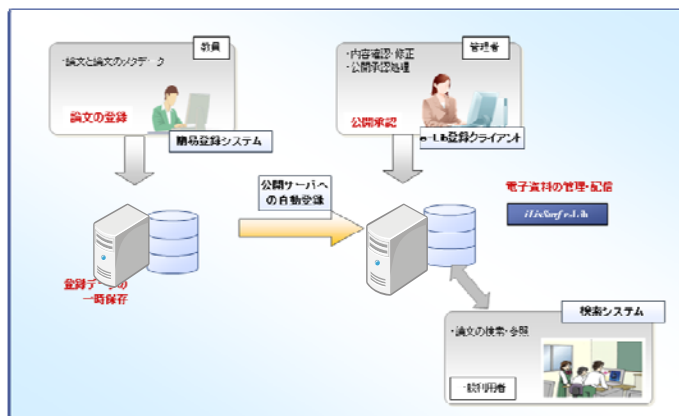


図3 簡易登録システム運用図

3. 市販のOCRソフトの評価

3.1 市販のOCRソフトについて

現在、OCRソフトはフリーソフト、製品版を含めて数多く流通しているが、PC上で処理するソフトウェアは比較的安価で入手可能である。その中には画像処理によりデジタル化したデータから文字を切り出す機能を有し、かつ、PDFデータの扱いが可能な商品がある。これらのOCRソフトの機能を利用すれば登録する画像データあるいは画像PDFファイルを透明テキストが付与されたものに変換できそうである。画像PDFをOCRソフトで処理して文書型PDF(透明テキスト付)を作成するという手法に着目し、これらのソフトの機能が利用に耐えるものかを調査することとした。

今回の調査には、機関リポジトリへの論文登録の支援を目的として行っているため、デジタル画像(jpg, mpg)ではなく、画像PDFファイルをメインに調査を進める。

3.2 画像処理における解像度

はじめに画像処理における解像度をどのように設定するのが妥当かという問題がある。解像度が高いとファイルサイズが大きくなり、低いと文字が明瞭に取り込めない。

今回の調査では一般的なPDF処理ソフトであるAdobe社のAdobe Acrobat 8

Professional(2)を使用し、日本語論文と英語論文をサンプルとして画像処理における解像度を変えて、ファイルサイズの比較(表1、表2)を行った。英文、和文とも解像度を600dpiとした場合にファイルサイズが著しく大きくなる。サイズとして実用可能なものは400dpiまでと思われる。

表1 日本語論文のPDFファイルサイズ

	スキャンした ファイルサイズ (KB)	AdobeAcrobat8 ProfessionalでOCR化し たファイルサイズ(KB)
印刷業者から納品されたPDF	487	—
200dpi	829	681
300dpi	1268	1028
400dpi	1620	1349
600dpi	2681	2173

表2 英語論文のPDFファイルサイズ

	スキャンした ファイルサイズ (KB)	AdobeAcrobat8 ProfessionalでOCR化し たファイルサイズ(KB)
印刷業者から納品されたPDF	110	—
200dpi	696	597
300dpi	1062	885
400dpi	1386	1155
600dpi	2213	1799

次に、OCRソフトを使用して解像度による読み取り精度を調査する目的で、解像度を変えてデジタル化を行い、ページに含まれる文字数を比較した。

表3 日本語論文のページ内文字数

	ページに含まれる文字数
印刷業者から納品されたPDF	828
200dpi	829
300dpi	828
400dpi	829
600dpi	828

表4 英語論文のページ内文字数

	ページに含まれる文字数
印刷業者から納品された PDF	1628
200dpi	1629
300dpi	1628
400dpi	1628
600dpi	1630

調査において日本と英語の例でそれぞれ、解像度ごとの抽出文字数を比較したが、文字数のみを見ると、日本語論文(表3)、英語論文(表4)とも解像度による文字の読み取り数には殆ど差が無い状態で、汚れを文字と認識するケースや全角文字を半角2文字に読み違える事は少ない。

読取間違いの発生に着目して比較するとデジタル化における解像度では日本語論文(表5)、英語論文(表6)ともに解像度により結果に差が発生する。

表5 日本語論文誤読文字数

	誤って判断された文字数	記号などを除いた読取間違い文字数	全体の文字数から正しく読込めた文字数の比率(%)
200dpi	66	9	98.9
300dpi	60	3	99.6
400dpi	58	1	99.9
600dpi	59	2	99.8

表6 英語論文誤読文字数

	誤って判断された文字数	記号などを除いた読取間違い文字数	全体の文字数から正しく読込めた文字数の比率(%)
200dpi	11	2	99.8
300dpi	8	2	99.8
400dpi	6	0	100
600dpi	7	2	99.9

日本語のドキュメントについては、殆どが「,」「。」などの句読点を、全角と半角で読み違えたもので、これらのものはシステムに登録される際に、正規化されるので運用上は問題ないとする。どの解像度もかなりの読取精度

であると考え、200dpiでは、「ベ」→「ペ」、「ブ」→「プ」、「深」→「探」、「懂」→「悼」、「憬」→「惜」と読み間違えていた。300dpiでは、読み間違いが減少し「ブ」→「プ」、「憬」→「憶」、「1」→「l」となっている。この「憬」→「憶」の読み間違いは、400dpi、600dpiでも同じように発生していた。400dpiでは、この間違いのみであった。600dpiでは、この間違いに加えて、それ以外の解像度では発生していなかった、「0」→「0」と判断しているものがあった。

次に英語のドキュメントとなるが、日本語のものに比べて読み間違いがかなり少なくなっている。こちらも殆どが「'」などの句読点を、全角と半角で読み違えたものである。それら以外では、200dpiでは、「u」→「n」、「o」→「0」の2つ、300dpiでは「rm」→「nn」(前述の表では2つの間違いとカウントしている)、400dpiでは、間違いなしであった。600dpiでは、解像度が上がったため、汚れを文字と認識していた。

今回のサンプルでは、日本語、英語と何れの解像度でも実用化は可能範囲であるが400dpiは読み間違いも少なく、PDFのファイルサイズもさほど大きくならないので適切である。

3.3 OCRソフトによる比較

3.2にて、今回のサンプルに関しては400dpiが適切な解像度と判断した。ここでは、この解像度のPDFを用いて、Adobe Acrobat 8 Professional (Adobe社)以外のソフトウェアでは、どの程度の精度で文字認識が行われるかを検証したい。

比較は英文字の読み取りで評価の高い、e.Typist v. 11.0 (3) (メディアドライブ社)をメインに検討し、その他として読取革命(4) (パナソニック社)を使用した。スキャナのOCR機能を試してみたが、解像度なども限定され、画像取り込みが本来の目的ではないこともあり、調査の対象から外した。

[e.Typist]

表7 日本語論文誤読文字数[e.Typist]

	誤って判断された文字数	記号などを除いた読取間違い文字数	全体の文字数から正しく読込めた文字数の比率(%)
400dpi	13	4	99.5
600dpi(参考)	12	3	99.8

表 8 英語論文誤読文字数[e.Typist]

	誤って判断された文字数	記号などを除いた読取間違い文字数	全体の文字数から正しく読込めた文字数の比率 (%)
400dpi	9	3	99.8
600dpi (参考)	6	0	100

「e.Typist」は先の Adobe Acrobat 8 Professional と異なり、OCR 専門のソフトウェアである。その為、認識範囲の指定や認識後に画像ファイルと比較して一文字ずつ確認、修正を行うことが可能となっている。また、処理の段階で PDF 文書が複数で構成されている場合は、これを 1 ページずつ分割しての作業となる。透明テキストをつけた PDF は全体をまとめて 1 つの PDF ファイルに出力することは可能であるが、Adobe Acrobat 8 Professional と比較すると全体の作業効率は悪くなる。

日本語のドキュメントの読取 (表 7) に関しては、Adobe Acrobat 8 Professional に比べて、記号類の半角/全角の誤りが殆ど無かった。読取の間違いとしては、400dpi では「、」→「紀」と判断していた。また、「1)。」を「%」と判断している。参考のために、600dpi の PDF を処理してみたが、「1)。」を「%」と判断したのみであった (上記の表では、3 個とカウントしている)。Adobe Acrobat 8 Professional では、すべての解像度の PDF で発生していた「憬」という文字の読み間違いは、このソフトウェアでは発生していない。英語のドキュメントの読取 (表 8) に関しては、Adobe Acrobat 8 Professional と同じく、記号の半角/全角の読み間違いが殆どであった。400dpi の PDF では、用紙の汚れを文字と認識しているが、運用上は殆ど問題ないと思われる。600dpi では、この用紙の汚れを読み取ることはなかった。

このソフトウェアは、OCR 専用ソフトウェアであるため、読取を本格的に行うには適していると考えられる。また画面上で読取を行う範囲を目視できるため不要と思われる箇所は、認識から外すことも可能である。しかし、複数ページの場合、範囲指定など煩雑な操作が多くなり、作業時間がかかるため、大量のドキュメントを処理するのには不向きである。

[読取革命]

「読取革命」は、画像 PDF を読込んで OCR 処理をする機能を持たないため、スキャナを用いて TIFF や JPEG の形式で出力し、その画像を OCR 処理、透明

テキスト付 PDF にするという処理を行った。

表 9 日本語論文誤読文字数 [読取革命]

	誤って判断された文字数	記号などを除いた読取間違い文字数	全体の文字数から正しく読込めた文字数の比率 (%)
400dpi	9	4	99.5

表 10 英語論文誤読文字数 [読取革命]

	誤って判断された文字数	記号などを除いた読取間違い文字数	全体の文字数から正しく読込めた文字数の比率 (%)
400dpi	4	0	100

ソフトウェアの性質か、改行の後ろや文字間の隙間部分を空白文字として認識していることが他のソフトウェアと大きく違うようである。その為、著者である「伊藤」という文字を「伊△藤」(△は全角空白文字)と文字認識しているため、「伊藤」では検索が行えない可能性がある。

日本語のドキュメントの読取 (表 9) では、「べ」→「ペ」、「、」を読込めていない、存在しない「=」を読込んでいる、という誤りがあった。英語のドキュメント (表 10) に関しては、他のソフトウェアと同様に、全角の記号類を半角と読み間違えたもののみであり、運用上は影響がないと思われる結果であった。

今回の検証では TIFF や JPEG にする際にページ毎に分割されてしまったため、作業効率がかかなり悪いものになってしまった。尚、複数の画像ファイルを読取認識させても出力の際は、1 つの PDF にすることが可能である。

3.4 OCR ソフトの適性

これらの調査結果より、現時点では、通常の運用では Adobe Acrobat 8 Professional を使い、複雑なレイアウトの文書や認識させたくない箇所があるものなどは e.Typist などを補完的に用いる運用が OCR 化には適正と思われる。

文字誤読における問題は、利用する検索エンジンに組み込まれた正規化機能により、解消されるケースがある。例えば、記号は正規化により削除されるため、紙面の汚れを記号として読み取った場合などは誤読による影響はない。また、カタカナや英文字の全角と半角の誤読も正規化により解消され

る。問題となるのは日本語特有の破裂音と濁音の表記である。今回のテスト結果では、誤読は1%台までであり、正規化による解消を考えると十分に検索索引の切り出しには利用可能な精度が得られたと判断している。

4. 簡易データ登録システムへの連動

市販のOCRソフトウェアの価格がさほど高額ではなかったため、今回はフリーソフトを調査の対象から外し、商品として利用の多い実績のあるソフトウェアの中から、システム連携の候補を選択する。

OCR化の制度や手間を考慮して、Adobe Acrobat 8 Professional を使用することとした。

4.1 機能改善の提案

4.1.1 登録処理後のファイル入替

データベース登録処理が完了した後に、PDFファイルダウンロードを行い、パソコン側でOCRツールを使用して透明テキストを付与する。新規に作成した文字型PDFファイルを元の画像PDFファイルの代わりに登録する処理方法が図4に示した案である。

ただし、機関リポジトリにおける本文切り出し機能は管理ツールにおける登録、または、サーバの夜間バッチにより提供されている。ファイル転送処理だけで入れ替え処理を行っても検索対象にはならない。

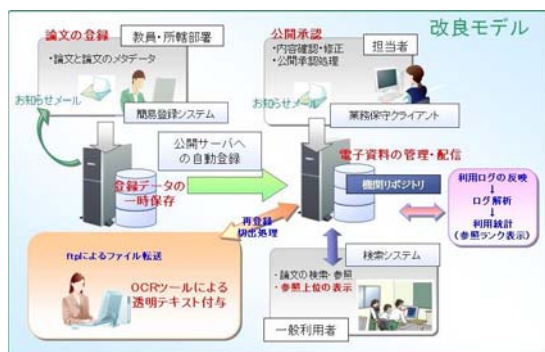


図4 登録後処理案

4.1.2 簡易登録処理工程への組み込み

簡易登録システムでは利用者から登録されたファイルが仮データとともに中間サーバに取り置かれ、その後夜間バッチでデータベースに登録される。

この仕組みに着目して処理の途中行程にOCR処理を組み込む方法が図5に示した案である。中間データとして登録されている画像PDFファイルを文字型PDFファイルに差し替え、サーバ登録時の切出機能を活用する方法である。

利用者が簡易登録システムを利用してPDFファイルをサーバに仮登録する工程において、夜間バッチで機関リポジトリのデータベースに登録する処理が行われるまでの時間を利用してOCR化処理を組み込む。

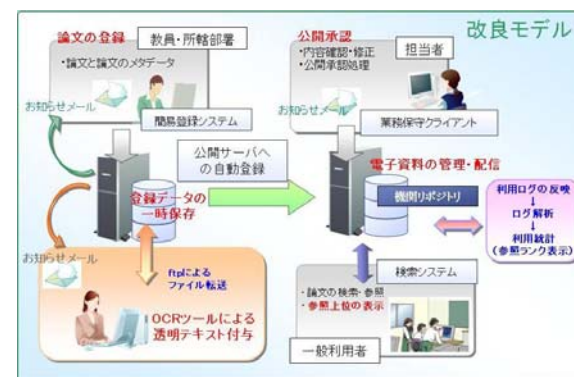


図5 処理工程中組込案

4.2 OCR化工程の選択

登録処理後にファイルをダウンロードして処理する方法の場合は、時を選ばず、また、まとめて複数のファイルを対象とした処理が可能である。しかし、データベースに登録された電子資料のなかから、画像PDFのファイルを見極めて抽出する作業は困難である。実際にPDFファイルのファイルヘッダーをバイナリで比較したが、画像PDFと文字型PDFを切り分ける決め手は見つからなかった。

利用者には画像PDFも文字型PDFも画面表示において見分けが付きにくい。利用者が気付かないうちにファイルが置き換わり、全文検索の対象となるケースではかえって改造により混乱が生じる可能性がある。

画像PDFで登録されているデータを後日、文字型PDFに加工して差し替えるという処理を否定するものではないが、画像PDFファイルを判別して一括

して抽出する手法が見つからないため自動処理としてシステム化する事はできない。一部、画像 PDF に加工して既に登録が行われた遡及の電子資料が存在する。これらのファイルが一括処理の対象となるが、今後の運用においては実用化が難しい。

一方、簡易登録工程への組込みについては処理を行うタイミングが難しい。しかし、簡易登録システムでメールを利用した登録通知機能があるので、ファイルが登録されたことを知ることは可能である。

状態管理を行うことで、夜間バッチまでの時間を利用して OCR 化を行う方法が負荷の少ない、リスクも少ない処理になると考える。

処理のタイミングが厳しいことへの対応策としては、データベースへの登録工程を翌日の処理とするなど、夜間バッチの時間を変更するなど一案である。また、中間サーバのディレクトリに一時保存する処理であることから、テキスト PDF に変換する処理前のファイルと処理後のファイルの保存場所（ディレクトリ）を変える方法なども考えられる。

工程の選択については運用テストを充分に行い、リスクを考慮して選択する必要がある。

5. システム組込の評価と運用

5.1 対費用効果

この機能を組み込むことで資料の OCR 化を学外の業者に外注することなく、学内での処理が可能となる。学生アルバイトによるデジタル化作業については、作業の指導と管理が必要となるため、図書館業務が委託体制で運用されている現行の人員体制では、専任職員の業務負荷が高まり難しいと判断した。

デジタルコンテンツの作成単価は元となる資料の形態が冊子体であるかパンフレットであるかという条件や、色彩が白黒 2 階調かカラーなのか、写真を含むかなどの印刷物の内容で異なる。A4 サイズの資料 1 枚当たりの処理価格は 60～140 円が標準的な費用である。更に OCR 処理を追加し、書誌データまで作成を依頼すると、デジタルコンテンツ作成費用は 1 件当たり数百円から千数百円となる。利用者提供のファイルを処理工程の中で変換し文字型 PDF に変更する提案は費用軽減目的で充分評価できると考えている。

5.2 実運用にむけて

現行のシステムに OCR 化工程を組み込むことで、画像 PDF ファイルを文字型 PDF ファイルに変換する処理を行なう方法は、運用上の大きな負荷が発生せずにデータの品質向上が見込まれる。また、利用予定の OCR 化ソフトは市

販されている安価で実績のある Adobe Acrobat 8 Professional であるため、運用面では、実現が充分可能な提案である。

テスト環境を構築したのち、テストデータを使用して一連の登録処理を行い、動作確認をした結果、大きな問題は発生していない。現在、本環境にも登録受付メール配信機能を組み込むなど、環境の整備を行っている。

5.3 今後の課題

実環境での適用については、テストデータによる登録調査を行った上で、本稼働となるが、メール通知により手動で OCR 化のタイミングを図る手順が実運用において問題とならないか、慎重に経過を観察したい。

経過を観察したうえで、次にバッチ時間の見直しを検討したい。また、ディレクトリ管理を見直して、仮データ（OCR 化前のデータ）を登録するディレクトリと、バッチ対象データ（OCR 化後のデータ）を登録するディレクトリを変更するなどの方法も実用化において検討が必要と考えている。作業担当者の負荷を軽減する目的で、ディレクトリに登録された PDF ファイルを自動的に Windows サーバに転送する仕組みを実装する事も有効と思われる。

仮登録されたデータより画像 PDF を選択し、該当するファイルのみを抽出してパソコンへ転送し、パソコンにおいて OCR 化処理を自動実行して、処理後の PDF ファイルをパソコンからサーバに転送する。これらの一連処理の自動化が今後の課題である。

今回は十分に時間がかけられなかったが、ファイルヘッダーの解析結果より、文字型 PDF か、画像 PDF かの判別が可能になれば、処理対象をファイル単位で判別して、自動抽出によりパソコンに転送する事が可能になる。パソコン上で行なう OCR 化処理の自動化も、使用するソフトウェアによっては実現できる可能性がある。

OCR 化処理をサーバ上で行なう方法を、パッケージ提供メーカーと共に検討は行った。しかし、販売されている Linux サーバ用の OCR 処理モジュールは業務用で高額であり、組み込みには現行のパッケージ改修費用も必要となるため費用対効果は得られないと判断した。今後も調査を続け、低価格でサーバモジュールが提供されれば処理工程を見直したいと考えている。

5.4 展望

インターネット時代の大学図書館において、電子ジャーナル[3]、機関リポジトリなどのデジタルコンテンツの提供は、重要な要素である。多くの学会がオープンアクセスへの取り組みを表明し、学会への投稿論文は著者の所属する機関のリポジトリへの登録を認める方向にある[4]。論文の公開・非公開

によりダウンロード数は明らかな差があり[5]、オープンアクセス・オプションとその被引用に対する効果も報告されている[6]。

各大学において、研究者の多くが機関リポジトリに論文登録を行うことで、機関リポジトリは大学における情報公開の場として、また研究成果の発信の場として活用されている。

システム連携による研究者支援の取り組みも行なわれ、Web of Scienceが提供する研究者の業績リスト公開サービス (ISI Web of Knowledge(5)) などは著明である。国内では、信州大学における研究者情報 (SOAR-RD(6)) と機関リポジトリに登録された研究成果の連携について、岩井雅史が取り組みを報告している[7]。ReaD (研究開発支援総合ディレクトリ(7))、研究者リゾルバー(8) (国立情報学研究所) など、研究者と研究成果を関連付けるサービスも提供を開始している。

インターネットを介して図書館で提供する、資料の検索および提供サービスは多くの改善が可能と思われる。システム間の連携、データベースの連動などにより、研究者が求める資料や学生が必用とする教材を安価に、迅速に提供する事が可能となる。

研究者ワークフローは堀切近史の述べるように[8]、従来とは大きく変わってきた。研究活動における電子資源の活用は、より大きな位置づけを持つようになり、研究支援のために、電子論文情報の効果的な提供が図書館に要求される。

インターネットを介して提供されている多くのサービスシステムでは、標準プロトコルや webAPI などの連携機能を提供している。また、全国大学共同電子認証基盤構築事業 (UPKI : University Public Key Infrastructure) (9) による全国規模の認証基盤構築の取り組みも始まっている。これらを有効に利用し、更にオープン・ソースや市販のツールを活用するなどの方法を駆使して、インターネットを活用した連携により、安価な費用で今後の図書館サービスの向上を目指したい。

【参考資料】

- (1) 学術機関リポジトリによる学内学術情報の発信強
<http://www.nii.ac.jp/metadata/oai-pmh/ir.pdf>
- (2) Adobe Acrobat 8 Professional (Adobe 社)
<http://www.adobe.com/jp/products/acrobat/>
- (3) e.Typist v.11.0 (メディアドライブ社)
<http://pac.mediadrive.jp/>

- (4) 読取革命 (パナソニック ソリューションテクノロジー株式会社)
<http://panasonic.co.jp/pss/pstc/products/yomikaku/>
 - (5) 研究者の業績リスト公開サービス ISI Web of Knowledge
<http://isiwebofknowledge.com/>
 - (6) 信州大学学術情報オンラインシステム SOAR
<http://library2.shinsu-u.ac.jp/soar/>
 - (7) 研究開発支援総合ディレクトリ ReaD
<http://read.jst.go.jp/>
 - (8) 国立情報学研究所 研究者リゾルバー
<http://rns.nii.ac.jp/>
 - (9) UPKI : University Public Key Infrastructure
<https://upki-portal.nii.ac.jp/>
- [1] 石井 奈穂子 . 立命館大学におけるアウトソーシングの導入 : アウトソーシング(業務委託)を活用した専任職員業務の専門性の確立 , 情報の科学と技術 57(7), 337-340, 20070701
 - [2] 飯澤 文夫 . 明治大学図書館におけるアウトソーシング : サービスの拡大・多様化・高度化に向けて, 情報の科学と技術 57(7), 331-336, 20070701
 - [3] 植松 貞夫 . 大学図書館における電子ジャーナルとその展望 , 図書館雑誌 103(11), (1032) 756?758, 2009/11
 - [4] 時実 象一 . オープンアクセス : 機関リポジトリの最近の動向 , 情報の科学と技術 59(5) pp.231-237 20090501
 - [5] 林 和弘 . 日本のオープンアクセス出版活動の動向解析 , 情報管理 52(4) pp.198-206 2009
 - [6] 時実象一 . オープンアクセス・オプションとその被引用に対する効果 , Current awareness (299) pp.10?13 2009/3/20
 - [7] 岩井 雅史 . 研究者情報との連携による機関リポジトリの戦略的発信 : 信州大学の取り組み , 情報の科学と技術 59(1), 18-22, 20090101
 - [8] 堀切 近史 . 研究者ワークフローと論文情報 (<特集>e-Research と学術出版) , 情報の科学と技術 59(1), 23-28, 20090101

(関連ウェブサイト参照日は全て 2010 年 01 月 10 日)