

マークシート方式による答案返却システムの実装について

梶取和明^{*†}

On implementing a marksheet system for returning exam papers

Kazuaki Kajitori^{*}

Abstract : The functionality of digitizing handwritten exams on sheets is definitely a necessary part of e-portfolio. We show such a functionality can be developed without much efforts and even a simple implementation of it is actually practical in the sense that there are satisfactorily few errors in recognition of student's ids and processing time is also tolerably short. We present the points of the system we developed showing what are important for such systems in practical uses and hope that it can be expanded to an e-portfolio system keeping its simplicity and user friendliness.

ASFA Key words : Web, scanner, examination, information systems

はじめに

教育・学習におけるeポートフォリオは、個人の学習のデータをオンライン上で保存・管理・利用する仕組みであり、近年主体的学習に必要な仕組みとして注目を集めている。これらのデータは学習者による学習活動の振り返り・自己および他者評価に積極的に利用されるものとされる¹⁾。

eポートフォリオで扱う学習データとしては、学習者がコンピュータ上で作成したものでないものも含まれる。たとえば、多くの記述式の試験答案は手書きで書くことが普通である。本論ではeポートフォリオの一部分を形成する手書きデータをコンピュータに取り込んで利用する方法について述べる。

手書きのデータをeポートフォリオの対象にするためには、効率良く電子化する手段が必要である。これには、普通の紙に学生に書かせてその用紙を回収し教師がスキャンする方法がある。他には、手書き答案を学生サイドでデジ

タル化して送る方法もある²⁾。後者は電子ペーパーやデジタルペンを用いて学生に手書き入力をさせ、その電子化筆跡をサーバーで回収するものなどで、すでに実用化している大学もあり³⁾少なくとも将来的には試験を含む多くの手書き電子化用途に対して有望であると考えられる。しかし前者の従来の答案用紙をスキャンする方法は費用が安く済む上に比較的簡単に実現できる利点がある。

本論では、教師が学生から集めた手書き答案用紙をスキャナで読み取ってサーバーにアップロードしたものを学生が参照できるようにする手書き答案用紙オンライン返却システムの構築を対象とする。このようなシステムを実装するに当たって必要なこととしては、

- ・ 答案を書いた本人を識別できるように学籍番号を読み取れる
- ・ 答案を書いた本人だけがオンライン上で閲覧できる（ログイン式にする）

^{*}水産大学校水産流通経営学科 (Department of Fisheries Distribution and Management)

[†]別刷り請求先 (corresponding author) : kajitori@fish-u.ac.jp

といったことが挙げられる。そのような要件を満すシステムはすでにいくつか^{4,5)}存在しているが、システム実装のポイントや運用上の問題点を論じたものはないようである。本小論では、自前の手書き答案オンライン返却システムを開発することによって比較的手軽に実装できることを示すとともに、技術的にそして運用上どのような問題点があるかを提示する。

システム仕様

実装するシステムは教師ユーザーと学生ユーザー双方ともウェブインターフェイスを使うウェブアプリケーションとし、以下の記述のためにこのシステムを ms と呼び、ms を実装したサーバーを ms サーバーと呼ぶ。ms は以下のような仕様とする：

1. 手書き答案用紙として学籍番号用のマークシート欄を設けた A4 縦の様式を用意し、裏表に記述可能とする。裏にはマークシート欄は設けない。表と同じ学籍番号を対応させる。
2. 受験者全員の答案はまとめてスキャナで、一つの PDF ファイルかページ数分の JPEG ファイルに読み込まれ、JPEG ファイル群は一つの ZIP ファイルにまとめるものとする。
3. ms は電子化した答案ファイル (PDF か ZIP) を ms サーバーにアップロードする機能を有し、教師ユーザーと科目名との対応を行えるようにする。
4. アップロードされた答案ファイルの各ページにマークシートから読み取った学籍番号を対応させる。読み取った学籍番号と元のページのマークシート部分の対応を表示する。
5. ms サーバーが表示した学籍番号と元のページのマークシートの対応が正しいかブラウザ上で確認し、誤りは正した上でサーバーに submit する機能を有する。
6. ms サーバーは誤りの訂正を反映させた上で学籍番号ごとにページを PDF ファイルにまとめて保存する。
7. 答案を書いた本人だけがオンライン上で閲覧できる。指定された URL からログイン画面に入り学籍番号の入ったアカウント名でログインすることで、その学籍番号の答案ファイルのダウンロードリンクを貼ったページにアクセスできる。

答案用紙の全体像 (の一部) は Fig.1 を参照されたい。Fig.2 は全てのページ (JPEG 画像ファイル) に対し読み取った学籍番号と元画像を対応させた HTML 表の一部である。1 にあるとおりページが用紙の裏の場合は、HTML 表の元画像欄にマークシートは表示されない (前ページと同じ学籍番号を対応させる)。

OCR ではなくマークシート式にしたのは、マークシートの認識率のよさと実装の簡単さによる。また、OCR では人間でも読み難い (当然 OCR でも読めない) 手書き数字を扱うこともあるが、このとき氏名欄に氏名があったとしてもそれを学籍番号に対応させるには手間がかかる。一方マークシートでは機械で読み難い学籍番号は人間が確認段階で容易に読み取れることが多い。

読み取った JPEG ファイル群を ZIP にまとめてサーバーにアップロードする場合は、サーバーで ZIP から戻したときにスキャナで読み取った順で処理されることを保証しておく必要がある。解答用紙の裏に学籍番号を対応させるには裏にマークシート欄がないことを確認した上で直前に読んだページの学籍番号を対応させるので、処理する順番が重要であるからである。

上記仕様の 2 はシステム外の機能で仕様を満すスキャナを用意する。ここでは富士通の ScanSnap IX500 を想定する。このスキャナは A4 紙を連続で両面スキャンでき結果を PDF ファイルにまとめるか、ページごとの JPEG ファイル群にすることができる。JPEG ファイル群の zip 化は OS に応じたアプリケーションを使うものとする。

Fig.2 の確認画面で「学籍番号」とあるのはマークシートから読み取った学籍番号である。元画像の学籍番号と読み取った「学籍番号」が違う時は、教師ユーザーは訂正欄を正しい学籍番号に書き換え、違わないときは訂正しないこととする。

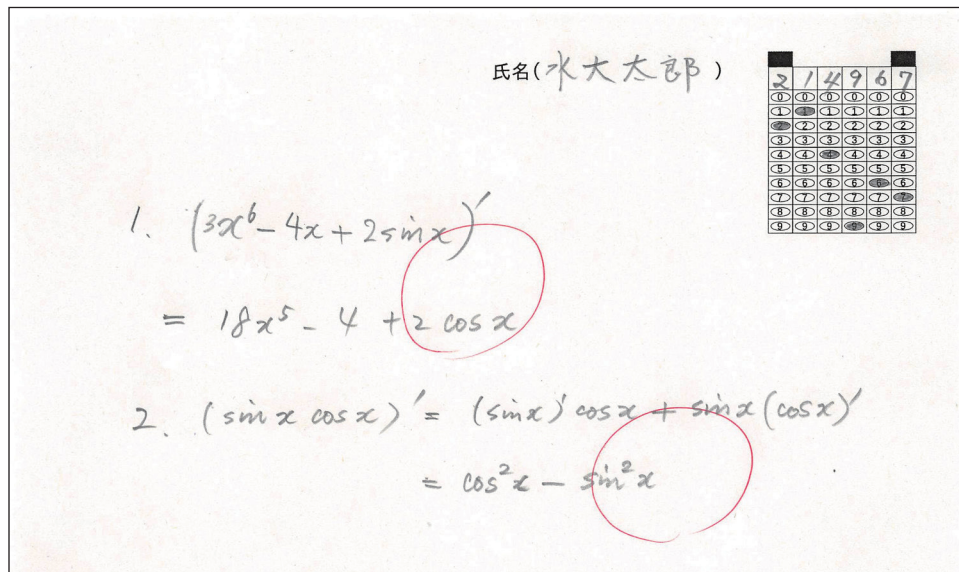


Fig. 1. 答案画像例



Fig. 2. 学籍番号確認画面

実装結果

Linux Centos7 を HP Proliant ml310 g5 上にのせて ms サーバーとし、HTTPD デーモンには apache を使用した。ウェブスクリプトは Perl で記述した。スクリプトの機能は、上の仕様に対応して以下ようになる：

1. 教師ユーザーのログイン認証とインターフェイス表示と教師・試験ごとに分けたファイルの保存
2. アップロードファイルからのページ画像をファイルとして取り出す (PDF と ZIP で異なる)
3. マークシート部分を解析してページ画像に対応する学籍番号を判断する
4. マークシート画像あるいは裏面の画像と学籍番号の対応を教師ユーザーに返す
5. 教師ユーザーから返された情報に基づいて学籍番号ごとにページを PDF ファイルにまとめる
6. 学生ユーザーのログイン認証と、当該学籍番号に対応した PDF 答案ファイルのダウンロードリンクの生成・表示

教師ユーザーに対しては独自のディレクトリを作成して、作成したファイルを教師ごとに区別している。

上の5の「学籍番号ごとにまとめられた PDF ファイル」にはファイル名に教師ユーザがアップロード時に指定したファイル名 (の . 拡張子部分を除いたもの) と末尾 (拡張子の前) にその学籍番号を入れることで試験と学生を区別している。

学生は学籍番号を含んだアカウント名とパスワードで授業などで指定された URL にアクセスしログインする。ログインするとアカウント名に含まれる学籍番号がファイル名の末尾にあるファイルが探し出されダウンロードリンクが表示される。現状その学生の全ての答案のダウンロードリンクが表示される。

答案ファイルの処理には Linux の gs (ghostscript のコマンド) と Perl の CPAN モジュール (Imager, PDF::API2, Archive::Zip) を援用するので、Perl プログラムとして 400 行余りのスクリプトに過ぎないが、以下 Linux や Perl 固有の事項は除き、スクリプトにとって本質的と思われる部分について述べておきたい。

マークシートを解析して学籍番号を読む部分はこのスクリプトの主要部分といえる。この部分は2つの独立なルーチンからなっている。マークシート欄を答案画像から取り

出すルーチンと取り出したマークシートから学籍番号を読み取るルーチンである。マークシート欄は欄上部の2つの黒い長方形でその最上部の位置を判断するようにしているが、答案用紙を広く探すよりは用紙の特定部分に欄があるかどうか判断する方がコードも単純になるし、読取精度も上げやすい。よって専用の用紙を使う方がよい。長方形マーク判別は HSV⁶⁾ の明度 (V) の閾値で判定している。この閾値の高低により長方形マークの認識に影響する。試行錯誤の結果 HSV 明度 0.90 以下の条件でマークを認識している。

この点是用紙の裏を処理することとも関連する。裏を読むときはマークシート欄がないことで用紙の裏と認識し表と同じ学籍番号を対応させるのでマークシート欄の位置が限定されている方が認識が容易なのである。

マークシート欄が取り出せたら、学籍番号を読むのは比較的簡単である。マーク部の1～9のセルの明度を判断して一番暗いセルの番号を採ればかなりの認識率を得ることができる。

結果

筆者は、2学期にわたって担当する数科目で ms システムを使用した。用途は答案用紙に記された記述式の期末試験の返却 (採点された答案をオンラインで参照できるようにすること) である。

まず、それらの試験のうち2つの試験で ms システムで答案をまとめたファイルをアップロードしてから学籍番号の確認画面が返るまでのサーバーの応答時間を Table 1 に掲げる。

Table 1 サーバーの応答時間

科目	受験者数	ページ総数	処理時間 (ZIP)	処理時間 (PDF)	誤読
線形代数	20	33	45 秒	97 秒	0
確率統計学	105	158	220 秒	—	2

表中、処理時間 (ZIP)、処理時間 (PDF) はそれぞれ、スキャンした JPEG をまとめた ZIP か PDF かのファイルをアップロードしてから確認画面が返るまでの時間である。誤読は学籍番号を正確にページに対応づけられなかったページ数である。「確率統計学」については PDF の処理時間は測定していない。表中、ページ総数は答案用紙の裏に解答が書かれた場合も1ページと数えた総ページ数である。

アップロードするときのファイルフォーマットについては、PDF ならスキャナ付属のアプリケーションが自動で

スキャン画像をまとめてPDFにしてくれるが、JPEG ファイルをZIPにまとめてアップするのに比べサーバーの処理時間がほぼ2倍となる。ZIPをほどこいて各画像ファイルを取り出すのに比べ、PDFから各ページ画像を取り出すのにずっと多くの時間を要するのが原因である。

表中の誤読2例はどちらも裏に書かれた解答をマークシート欄と誤認識し誤った学籍番号を対応させたものである。スクリプトでは現状マークシート欄をある程度広い範囲で探すようにしているので、位置を狭く決め打ちすることでこのタイプの誤読は減らすことができる。

確認画面 (Fig.2) における確認は誤読が2~3ページであれば数分程度である。実際、Table 1 の確率統計学の答案用紙をスキャナにセットしてから、サーバー上で読み取った学籍番号の確認、修正、提出し、学生がオンラインで確認可となるまでにかかった時間は10分20秒であった (i5 2.6GHz のPCでスキャナIX500とそのアプリケーションを使用)。この時間例では2つの誤読の訂正は20秒以内に済ませている。受験した学生が氏名だけ書いて学籍番号を書かない場合などは学籍番号を調べる必要があるのもっとかかる。しかし定期試験では学籍番号の記入漏れはこれまでのところない。

考 察

マークシートからの学籍番号の誤読はほぼ生じない。答案用紙の裏の処理は、答案用紙としてmsシステム専用用紙を使いマークシートの位置を固定してしまえば誤読はほとんど生じなくなるであろう。誤読をなくすことは難しいが現状の誤読率でも実用に差し支えないので問題とはならない。

上述のようにこのスキャナを使った場合100人150ページ程度の答案であればスキャン開始から作業終了まで10分程度なので時間的にさほど負担ではないと筆者は感じる。

PDFでアップロードする場合の処理時間の短縮にはgs以外にもっと速くjpeg画像をPDFから取り出すプログラムがライブラリがあればよいのであるが現状見つからない。

ログインした学生の全ての答案のダウンロードリンクが表示される点は、現実の必要に鑑みてよく検討する必要がある。試験の区別はアップロード時のファイル名による現行の方法が簡便ではあるが、ファイル名のコンフリクトの問題もあり、試験ごとにディレクトリを切りなおかつ簡便

な方法を検討中である。それらの検討結果をmsに反映させる予定である。

本システムの問題ではないが、スキャナで読み取ったものを返却する場合の問題点として以下のようなことがある。解答者が色の濃い鉛筆を使用した場合スキャナの紙送りのローラーに鉛筆の粉がつき以後の用紙に筋がつくことがある。これはスキャンを直す必要が生じた場合にはスキャン画像にも筋がつく。これを防ぐことは困難である。

誤読率のさらなる削減、さらに可能なら処理時間の短縮、A4縦以外の専用用紙への対応など、望まれる改善点はいろいろあるが、現状の400行程度のコードでも十分実用的であることがわかった。スキャン画像を目で読んで学籍番号を対応させサーバーにアップするのに比べれば効率面でも効果は大きい。また答案を手渡して返す時間が省け、また採点后数十分内にオンラインで確認できる点も効率面での成果である。

本システムの開発の手間については、コードは1ヶ月以内に現在の実用のレベルに達した。「実装結果」に書いたような各種ライブラリやwebサーバーの助けを借りれば、短い期間と少ないコード量で実用システムの実現が可能である。

参考文献に挙げた答案返却システム「飛ぶノート」は機能的には本システムに似ていると思われるが、公開された情報は少ない。また「授業支援ボックス」は用紙の多様性など機能は高いが高価な商用システムである。本論は、安価なコストで比較的簡単に実現できることと実際に使用してみたの情報を示したことで意義があるのではないかと思う。

学生が答案をオンラインで確認するにはログインを必要としているが、ユーザー名とパスワードを忘れる(なくす)ケースが多々ある。アカウントは学校のコンピュータ教室で渡す初期アカウントを用いているが、初期パスワードは教室の方では変えることができるため初期パスワードがわからなくなるのである。これは学内で統一アカウントを使えるようにしない限り解決できない。

まとめに代えて

データを紙に書くことはなくならなから、これらの紙上データを電子化して管理する需要も存在し続けるであろう。スキャナも便利なものがあるし、マークシートの読取りも画像処理ライブラリを利用すれば比較的簡単といえる。紙上の手書きデータの電子化とその管理はe

ポートフォリオのごく一部である。実際現存する e ポートフォリオシステムあるいは LMS の中には他の紙上手書きデータ電子化システムを使えるようにしているものがある。

現状の ms システムはまだ試行段階にあるため多くの人に利用してもらうには改善すべき点がある。本文で言及した他にも、ms システムを e ポートフォリオシステムにまで発展させることも視野に入りたい。各種 LMS との連携を図ることで利用性が高まることも考えられる。認証については独自の認証を用いているが LDAP 認証との連携を考えたい。

その場合、ユーザーの立場で分かり易いシステムにすることを第一に考えたい。

参考文献

- 1) 東京学芸大学 森本康彦研究室 教育システムチーム, 教育分野における e ポートフォリオとは, <http://draco.u-gakugei.ac.jp/eportfolio/> (2013)
- 2) 加藤 他, デジタルペン、タブレット PC、PC および紙と鉛筆の4種類のメディアを用いた試験に関する比較分析, 日本教育工学会第26回全国大会講演論文集 2010, 681-682(2010)
- 3) PC Exam - University of Copenhagen, <http://pc-exam.ku.dk/>
- 4) 飛ぶノート, <http://www.carrier-port.jp/note/>
- 5) 授業支援ボックス, http://www.fujixerox.co.jp/product/mf_etc/class_box/
- 6) HSV, <https://ja.wikipedia.org/wiki/HSV> 色空間