

マークシート処理システムの拡張による学力調査データ分析方式[†]森 薫^{*1}・田中 佳代子^{*2}・玉村 雅敏^{*3}・金子 郁容^{*1}慶應義塾大学政策・メディア研究科^{*1}・慶應義塾大学 SFC 研究所^{*2}・慶應義塾大学総合政策学部^{*3}

本稿では、拡張可能なマークシート処理システムを利用して、広域的な学力調査を実現するシステム構成を示す。本システムは、マークシート形式の学力調査における解答情報の電子データ化とデータ分析の機能を有する。宮城県で実施した平成 24 年度みやぎ学力状況調査において、本システムを運用した。この調査では、宮城県内公立高等学校の生徒(約 30,700 人)を対象として、マークシート形式の学力調査を実施した。回収した解答用紙は各学校で電子データ化され、そのデータを本システムに登録した。運用期間におけるデータの登録状況を評価し、本システムの実務での運用可能性について検討する。

キーワード：学力調査，マークシート，データベース，データ分析

1. はじめに

学力調査や学校評価のアンケートなどで、マークシート形式の調査が広く実施されている。学力調査を効果的に活用するためには、教育委員会や学校が多面的な分析を行うだけでなく、その調査結果が早期に提供されることが重要である(文部科学省 2008)。広域的な学力調査を実施し、学校や教育委員会において、役割に応じたデータ分析機能を提供するシステムが開発されてきた(木幡ほか 2008)。このシステムを利用することにより、学力調査の採点と電子データ化を教員が実施し、学校や教育委員会が役割に応じたデータ分析を実施することによって、調査結果の活用場面が拡大した。しかし、学力調査の採点結果を教員が手動でファイルに転記するなど、実務的なコストが課題となっていた。電子データ化を効率化するためには、マークシート形式の調査が有効である。本システムでは、オー

プンソースのマークシート処理システムである SQS(久保ほか 2005)を拡張して、自由度の高いマークシート調査票の作成を実現する。SQSはXML技術を基盤としたマークシート処理システムであり、その拡張もXMLによって可能である。さらに、マークシート処理システムを拡張するためのXMLを利用して、データ分析システムが動作する。データ分析システムはXMLから学力調査の設問構造を認識し、その構造に応じたデータ分析を実行する。

普通紙とイメージスキャナーを利用したマークシート処理方式(Deng et al. 2008, 河野ほか 2011)や、調査票におけるマーク位置の座標情報を設定することによる調査票の作成方式(高橋・松田 2008)が提案されている。本システムの特徴は、XML技術を基盤としたマークシート処理システム(久保ほか 2005)を拡張し、マークシート処理システムとデータ分析システムの連携を実現することにある。学力調査の実施において、マークシート処理システムとデータ分析システムが連携することによって、学校や教育委員会が調査結果を早期に電子データ化して、そのデータを分析することが可能となる。

宮城県で実施した「平成24年度みやぎ学力状況調査」において本システムを運用した。この調査では、宮城県内公立高等学校の生徒(約30,700人)を対象として、マークシート形式の学力調査を実施した。回収した解答用紙はマークシート処理システムによって各学校で電子データ化され、それをデータ分析システムに登録した。システム運用期間におけるデータの登録状況を評

2013年4月2日受理

† Kaoru Mori^{*1}, Kayoko Tanaka^{*2}, Masatoshi Tamamura^{*3} and Ikuyo Kaneko^{*1}: An Analysis Method for Achievement Test with Extensible Optical Mark Recognition System

*1 Graduate School of Media and Governance, Keio University 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan

*2 Keio Research Institute at SFC, Keio University 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan

*3 Faculty of Policy Management, Keio University 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan

備し、本システムの実務での運用可能性について検討する。

2. マークシート処理システムの拡張

本システムは、学力調査に対応したマークシート形式の解答用紙を作成するために、オープンソースのマークシート処理システムであるSQSを拡張する。SQSは普通紙によるマークシート調査を実現するソフトウェアであり、調査票を作成するためのSQS SourceEditorと、イメージスキャナーで読み取った調査票の画像に対してマーク認識を実行するSQS MarkReaderから構成される。SQS SourceEditorにおいては、XMLエディタによって調査票を作成する。

SQS SourceEditorのXMLエディタを利用することによって、設定した質問が自動的にレイアウトされる。SQS SourceEditorのレイアウト形式はアンケート調査を前提としているため、学力調査などへの適用が困難であった。本システムは、調査票のマーク位置を自由に設定可能とする拡張によって、SQS SourceEditorと互換性のあるXMLを生成する。これにより、SQS MarkReaderによってマーク認識可能な調査票を生成する。さらに、実施した調査結果を広域的に集計したデータ分析が可能となる。

2.1. データ構造

SQS SourceEditorで作成した調査票はPDF形式で出力される。そのPDFにはSQSソースファイルとSQSマークファイルが添付されている。これらのファイルはSQS MarkReaderがマーク認識する際に利用する。

2.1.1 SQS ソースファイル

SQSソースファイルは調査票の質問構造を格納するXMLファイルである。このファイルには質問文と選択肢の情報と調査のメタデータが記述されている。質問構造はXHTML、および、XFormsに準拠したXMLで記述されている。

2.1.2 SQS マークファイル

調査票に描画されるオブジェクトはSQSマークファイルによって記述される。SQSマークファイルには、調査票上のマーク欄や自由記述欄に関する座標情報が記述されている。SQS MarkReaderはSQSマークファイルの情報に基づいてマーク認識処理を実行する。

2.2. システム構成

本システムの構成を図1に示す。本システムは調査票

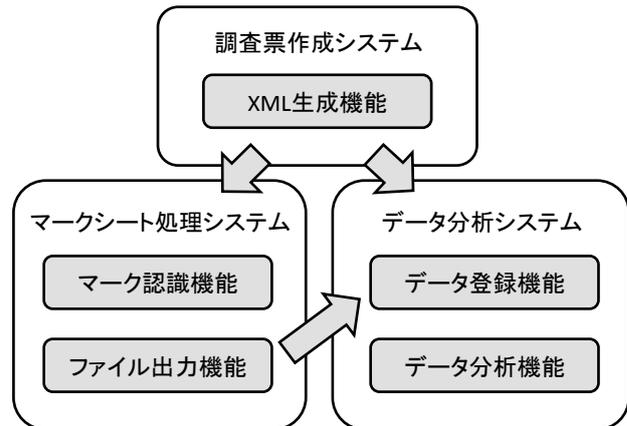


図 1 システム構成

作成システム、マークシート処理システム、データ分析システムから構成される。本システムにおけるマークシート処理システムはSQS MarkReaderである。SQS MarkReaderでマーク認識可能な調査票を作成するために、調査票作成システムにおいてSQSソースファイルとSQSマークファイルの仕様に合致したXMLファイルを生成する。マークシート処理システムはXMLの記述に基づいてマーク認識処理を実行して、その結果をファイルに出力する。

マークシート処理システムから出力されたファイルをデータ分析システムに登録することで、データ分析が可能となる。データ分析システムにおいては、調査票作成システムから生成されたXMLを利用して、データ処理が実行される。すなわち、調査票作成システムから生成されたXMLを介して、マークシート処理システムとデータ分析システムが連携する。

3. データ分析システム

SQS MarkReaderによるマーク認識を実行すると、その結果はエクセル形式とTSV(タブ区切り)形式のファイルに出力される。また、SQS MarkReaderは単体のソフトウェアとして、マーク認識結果を円グラフや棒グラフで表示する機能を有している。しかし、広域的な調査におけるデータ分析を実現するためには、各学校のデータを集約して集計することが必要となる。各学校においてデータを登録し、そのデータを分析可能とするデータ分析システムを実装した。

3.1. データ登録機能

広域的なデータの集約を実現するために、SQS MarkReaderで解答用紙をマーク認識した結果として出力されたファイルを、データ分析システムに登録す

る。データ分析システムにおいては、登録されたデータの妥当性が検証される。調査票作成システムから生成されたXMLを利用して、調査の設問数と各設問に対応した数値範囲を取得する。それらの情報から、登録されたデータが調査票の仕様に合致しているか検証することが可能となる。

解答用紙をスキャンする際の歪みやマーク欄の記入状況が原因となり、マーク認識が正常に実行できなかった場合、マーク認識結果のファイルを編集してデータを修正することも想定される。手動で修正したデータには記入ミスが想定されるため、データ登録時にファイルの妥当性を検証することにより、利用者がファイルを修正する作業を支援する。

3.2. データ分析機能

登録されたデータに対して分析処理を実行する際、分析対象となるデータの規模に応じて計算時間は増大する。利用者からのリクエストに応じて分析処理を起動した場合、多数の利用者からのリクエストの集中によって計算資源が枯渇することが想定される。データ分析システムにおいては、データ分析結果はデータ登録の際に計算してデータベースに蓄積される。結果が事前に計算されているため、対象となるデータの規模が増大したとしても、それによる性能低下を回避することが可能である。この処理はデータ登録をトリガーとして非同期に実行される。

データ分析システムはWebアプリケーションとして構築した。データ分析システムを利用するためには、IDとパスワードによるユーザ認証が必要である。WebアプリケーションはWebブラウザを経由してアクセスするため、クライアント環境への依存性が少ない。広域的な調査を実施する際には、多数のクライアントに専用のソフトウェアを導入して利用環境を構築することは現実的ではないため、Webアプリケーションが適していると考えられる。WebアプリケーションはRuby on Railsで構築し、データベースにはMySQLを利用した。また、データ分析のための非同期処理はDelayedJobで制御した。

利用者にデータ分析の結果を表示するデータ分析機能を実装した。データ分析の種類として、設問別正答率、観点・領域別正答率、正答数ヒストグラム、正答率箱ひげ図、クロス集計がある。これらはSQSで提供されていない機能であり、本システムにおける拡張によって実現された。データ分析のひとつの画面例として、正答数ヒストグラムの分析画面を図2に示す。デー



学力調査の分析 (平成24年度みやぎ学力状況調査)

学年	高校2年
科目	国語
分析手法	ヒストグラム
比較対象	なし

分析

正答数ヒストグラム

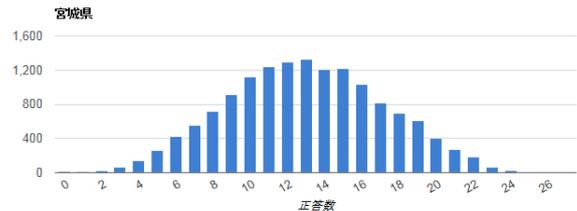


図 2 データ分析画面例

タ分析においては、分析するデータの範囲とデータ分析の種類を指定することによって、データ分析の結果が表示される。各学校がデータ登録した時点でデータ分析が可能であり、学校は分析データを早期に活用することが可能である。

4. システム運用

4.1. 学力調査の実施とデータ登録

平成 24 年度みやぎ学力状況調査において本システムを運用した。調査は平成 24 年 7 月 5 日から 7 月 13 日の間に宮城県内公立高等学校の生徒(約 30,700 人)を対象として実施した。回収したマークシート形式の解答用紙は各学校でスキャンされ、マークシート処理システムから出力されたファイルをデータ分析システムに登録した。

マーク認識の際に発生したエラーについては、学校が結果ファイルを編集して修正することとした。各学校がデータ分析システムを利用し、7月27日までに82.4%、8月3日までに96.4%のデータ登録が完了した。

4.2. 結果と考察

スキャンした解答用紙の画像をマークシート処理システムでマーク認識する際には、イメージスキャナーの設定に起因するエラーが発生した。具体的には、スキャンした画像をグレースケールとして処理した場合、画像の色濃度が低いためにSQS MarkReaderがマークを認識できないケースがあった。この問題に対処するため、SQS MarkReaderではモノクロ2階調の画像を利用することが推奨されている。また、イメージスキ

ショートレター



図 3 マーク認識エラーの発生した解答用紙例

ャナーの画質の設定によっては、生成された画像にノイズが発生しやすくなる。システム運用において実際に生成された画像を図3に示す。ただし、画像には個人情報が含まれているため、画像の一部を編集した。このような画像をSQS MarkReaderで処理すると、解答用紙の四隅に設置した基準点の座標認識に失敗し、マーク認識処理が実行不可能となった。

イメージキャナーの設定を修正することで障害が解決できたケースが大半であり、学校におけるイメージキャナーの設定が本システムの利用性を左右したと考えられる。学校で解決することが不可能だったエラーについては、宮城県教育委員会と宮城県教育研修センターがサポートして、最終的には全データをデータ分析システムに登録した。

5. 結論

本稿では、拡張可能なマークシート処理システムを利用して、広域的な学力調査を実現するシステム構成を示した。本システムではマークシート処理システムとデータ分析システムが連携することによって、マークシート形式の広域的な学力調査を早期にデータ分析

することが可能となった。

宮城県内公立高等学校の生徒を対象として実施した平成24年度みやぎ学力状況調査において本システムを運用した。その全データがデータ分析システムに登録され、学校と教育委員会にデータ分析機能を提供した。

今後の課題として、データ分析機能の拡充とマーク認識エラーの修正インターフェースの実現が挙げられる。より詳細なデータ分析をするために、複数の設問間で誤答パターンを分析する方式を実現する予定である。また、回収した解答用紙の状態に起因するマーク認識が不可能なケースは、今後も継続的に発生することが想定される。適切なデータ修正が簡単に実施できるインターフェースについて実現する予定である。

謝辞

本研究の一部は文部科学省復興教育支援事業の委託を受けたものである。システム運用には宮城県教育庁高校教育課後藤和也氏と宮城県教育研修センター太田克佳氏の協力があつた。ここに感謝申し上げる。

参考文献

Deng, H., Wang, F. and Liang, B. (2008) A Low-Cost OMR Solution for Educational Applications, *Proceedings of International Symposium on Parallel and Distributed Processing with Applications*, 967-970

河野広貴, 船曳信生, 中西透, 天野憲樹 (2011) Web を用いたマークシート方式テスト支援機能の実装, *電子情報通信学会技術研究報告 教育工学(ET)*, **110**(453):193-198

木幡敬史, 関子泰三, 森薫, 玉村雅敏, 金子郁容 (2008) 都道府県レベルのための学力テスト分析システム: デザイン・開発・実施運用, *日本教育工学会論文誌*, **31**(Suppl.):169-172

久保裕也, 玉村雅敏, 木幡敬史, 金子郁容 (2005) カスタマイズ可能な調査スキーマの共有による学校評価支援, *情報処理学会論文誌*, **46**(1): 172-186

文部科学省(2008) 平成21年度全国学力・学習状況調査の実施方法等の改善について
http://www.mext.go.jp/b_menu/shingi/chousa/shotou/045/toushin/1217156.htm (参照日 2013.03.29)

高橋等, 松田稔樹 (2008) マーク位置を任意に設定可能なマークシート読み取りソフトウェアの開発と簡易図形判定への適用, *日本教育工学会論文誌*, **31**(Suppl.):97-100