

## ウェブデータウェアハウスと協働する業務報告書 オーサリングシステム

森 雅生<sup>†1</sup> 田中 要江<sup>†1</sup> 廣川 佐千男<sup>†2</sup>

この研究では、大学評価に関する報告書の作成と文書管理を支援するシステムと根拠データを集積管理し分析を支援するデータウェアハウスを開発した。大学評価は自己点検・評価を前提としているが、教育研究活動は多岐にわたり、全てを把握することは一般に困難である。例えば、会議の資料(質的データ)の系統的・機能的な管理を行うことは難しいとされ、また、学生成績情報や教員人事情報といった情報のデータウェアハウス構築(量的データ)も課題とされている。本研究では、大学評価の根拠資料に必要な質的データと量的データの蓄積、および評価書作成を効率的・整合的に支援するシステムを紹介する。

### Enterprise Report Authoring System Cooperated with Web Data Warehouse

MASAO MORI,<sup>†1</sup> TOSHIE TANAKA<sup>†1</sup>  
and SACHIO HIROKAWA<sup>†2</sup>

We present the development of an enterprise report authoring system and a web data warehouse for university evaluation and accreditation. The report authoring system assists to compose self-study reports and to archive related digital documents. The web data warehouse is a web-based repository of institutional information. Though university evaluation and accreditation are conducted premising self-study by universities, it is difficult to grasp variety of educational and research activities of university. For instance, efficient data management of faculty council (qualitative data) is a hard work. It is also a problem to design and build a data warehouse (quantitative data) effectively. Our proposing systems also support cooperation of qualitative data and quantitative data, and consistent composition of self-study reports for the purpose of evidence-based university evaluation.

### 1. 序 文

情報技術の急速な発展により、私たちは”情報爆発”という困難に直面している。膨大な量の情報を利用する観点からは、様々な研究がなされている。しかしながら、情報を提供するという立場からの問題も大きい。この論文は、組織の説明責任、すなわち「信頼のある業務報告書の作成」という観点から問題に取り組んだ。報告書の根拠資料として情報を集積するとき、個々の情報の信憑性だけでなく、情報間の整合性に配慮しなければならない。加えて、業務報告書作成には、包括的な情報を継続的に収集管理し一貫性を持たせて提供されなければならない。このような要件を踏まえて、我々は報告書オーサリングシステムとデータウェアハウスおよび、システムを軽量的な連携機構を構築した。報告書オーサリングシステムは、様々な報告書様式に対応し、関連電子文書を蓄積する機能を有する。データウェアハウスは、ウェブに基づく組織または機関の情報を一元的に集約したりポジトリである。

ケーススタディとして、これらのシステムを大学評価報告書の作成支援に適用した事例を報告する。日本の全ての大学は、卒業生に学位を与える教育機関として国から認定を受けるための「機関別認証評価」を受審する事が義務づけられている。また、国立大学法人は、自らが立てた中期目標計画に沿って大学を運営している事を国民と政府に示す「法人評価」を受けなければならない。これらの評価を大学評価と称する。大学の役員会や教授会の会議資料(質的データ)は、自己評価の際の根拠資料とするために蓄積されているが、機能的な管理をされるに至ってはいない。評価や説明責任を果たす材料として、会議資料の再利用をどう行うかは大きな問題となっている。このように大学評価では、膨大な量の自己評価書と根拠資料の提出が大学に義務づけられている。これは、大学にとって切迫した課題である。本研究は、このような課題に対応する現場の必要性から始めたものである。

このような状況は、大学に限った事ではない。民間企業では内部統制による文書管理、公的機関では情報公開法や公文書管理法の施行により、組織が自らの組織情報の管理を義務化する気運が高まってきている。民間企業および公的機関における組織情報について、恒常的な集積と電子化の努力はなされているが、再利用や活用と言った観点から、本来の意味での構造化はなされていない。このように本質的な構造化がなされていない電子文書に基づい

<sup>†1</sup> 九州大学 大学評価情報室  
Institutional Research Office, Kyushu University

<sup>†2</sup> 九州大学 情報基盤研究開発センター  
Research Institute for Information Technology, Kyushu University

分野	学部	報告書の内容		
		節の数	観点の数	ページ数
教育	31	8	12	959
研究	20	5	5	311

図 1 九州大学における大学評価報告書の例

て、業務報告書を作成するにあたり、効率的で効果的な支援システムを開発する事が本研究の目的である。

大きな組織において作成される様々なデータや文書を、統一的に管理することは容易なことではない。必要な資料を必要ときに、効率よく利用することが重要である。そのための新たなシステムを構築するとしても、それまで蓄積されてきている資料を有効に活用できなければならない。しかし、全資料の移行は、費用対効果の点から考えると現実的ではない。筆者らは、九州大学の大学評価情報室に所属し、大学評価報告書や認証評価報告書の作成支援や、そのためのデータ蓄積システムの研究開発を行っている。九州大学は、2193 人の教員、2610 人の職員、18588 人の学部生と大学院生からなる総合大学である（2009 年現在）。教育については、29 の学部学府があり、研究については 16 の研究院と 3 つの研究所がある。評価報告書は組織全体として整合性が取れたものでなければならない。そのためは、それぞれの部局の基本資料に基づき作成されなければならない。しかも、単に資料を結合するのではなく、評価の視点に応じた構成が必要で、基礎資料をどのように解釈するかという文書をあらたに書き起す作業も必要である。例えば、2008 年（平成 20 年）に、文部科学省に提出した法人評価における業務実績報告書と、部局による教育・研究の現況調査報告書はそれぞれ、959 ページ、311 ページもの分量となる（図表 1）。

提案するシステムは、

- (1) 根拠資料としての電子文書（質的データ）と、報告書作成目的で作られる文書とを整合的に組み合わせるには、どのような方法が良いか、
  - (2) 組織の諸活動を示す数値データ（量的データ）はデータウェアハウスへ蓄積されるが、これらのデータの共有と活用を見据えた効率的な再利用するには、どのようなデータウェアハウスをどのように構成すれば良いのか、
  - (3) 報告書作成を支援するために、これらの質的データと量的データをどのように統合するか、
- という 3 つの課題の解決を目的とする。

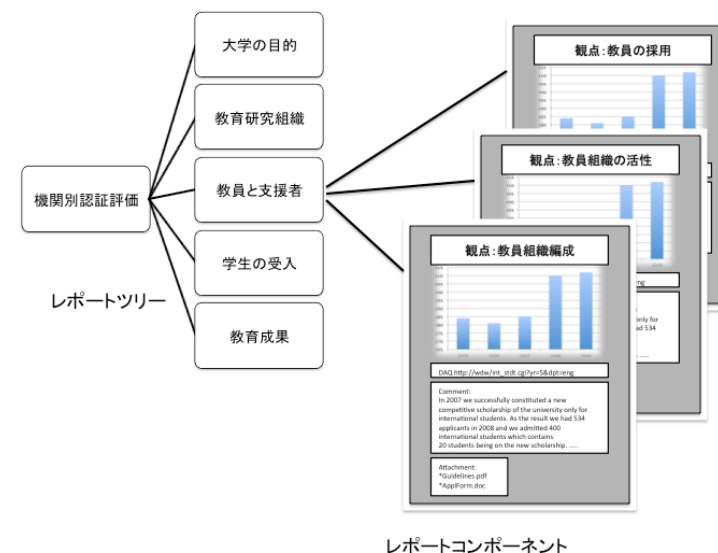


図 2 レポートツリーの概要図

## 2. システムの概要

### 2.1 レポートコンポーネントとレポートツリー

この節では、レポートオーサリングシステム（*report authoring system*、略称 RAS）について述べる（図 3 の左側）。データ構造レポートコンポーネント（図 2）は、RAS の情報単位であり、次の 7 つの要素からなる。

- 識別 ID
- タイトル
- コメント
- データウェアハウスへの問い合わせ（DAQ）
- 表とグラフ
- 関連文書添付
- メタデータ

レポートコンポーネントは、報告書の章節の最小単位に対応している。筆者らの経験から、章節の最小単位は報告の観点や分析項目などの個々に独立した概念である事が多い。レポートコンポーネントにおけるタイトル、コメント、データウェアハウスへの問い合わせ (DAQ) と関連文書添付は、報告書における各報告項目に関する詳細である。表とグラフでは、データウェアハウスから得られたデータ (CSV データ) に基づいて表やグラフなど、可視化した画像を示す。可視化機能はレポートコンポーネントに配備されている。データ分析クエリ ( *data analysis query*、略称 DAQ) は、データウェアハウスへの問い合わせであり、データベースへのクエリを CGI プログラムにしたものである。ゆえに、ウェブ API としての利用を前提としている。詳しくは次節で述べる。メタデータは、レポートコンポーネントの所有者情報や更新情報である。

RAS の中核的機能であるレポートツリーについて述べる。業務報告書は章や節による木構造をなしている。この RAS では、レポートコンポーネントを葉とし、章節を内点とする木構造を与える事により、業務報告書全体を表現するデータ構造を与えている。このデータ構造をレポートツリーと呼ぶ。レポートツリーの根や内点となる点のデータ構造は次の通り。

- 識別 ID
- 章節タイトル
- 子の識別 ID リスト
- 親の識別 ID

図 2 の右側にある灰色の箱は、レポートコンポーネントの例を示している。棒グラフは、DAQ によってデータウェアハウスから得られた CSV から生成される。次節で述べるウェブデータウェアハウスからではない、他の情報源から得られる静的な CSV データを、グラフの元データとする事もできる。

## 2.2 データ分析クエリ

この節では、レポートオーサリングシステムと同時に開発したウェブデータウェアハウス (*web data warehouse*、略称 WDW) について紹介する。システム概要図 (図 3) では、右側にあたる。

WDW の情報源は、業務システムのバックアップデータである。大学評価の例を取ると、学籍や学生の成績を管理する教育システムや、シラバスおよび授業の時間割を管理するデータベース、教員の研究活動を蓄積する教員データベースや、備品や設備に関する情報、出張の記録、財務システムなどである。これらは、必ずしもリレーショナルデータベースに蓄積

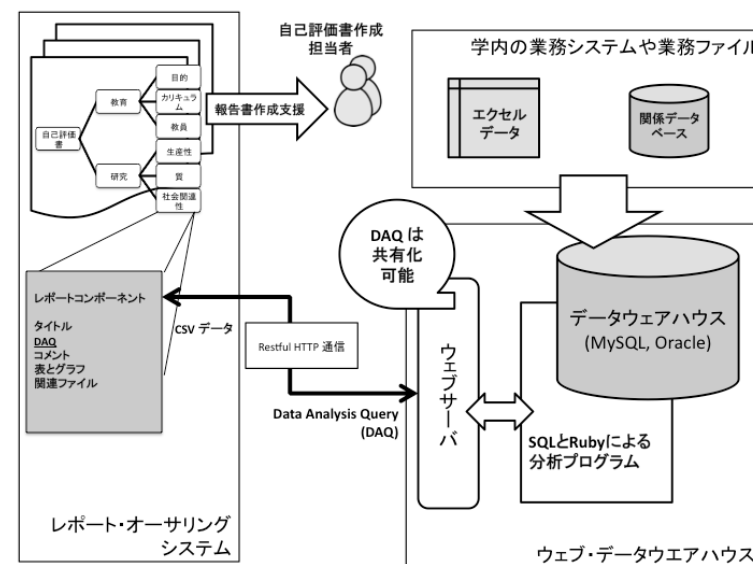


図 3 システムの概要図

されているとは限らず、スプレッドシートやファイルのまま保存してあるものも含まれる。

WDW の管理者は、WDW に蓄積されるデータを分析するためのプログラムを供給する。このプログラムをデータ分析クエリ (*data analysis query*、略称 DAQ) と呼ぶ。DAQ は CGI プログラムとして実装され、restful な HTTP 通信によりアクセスが可能である。DAQ は、データ分析の結果を CSV などの汎用的なファイル形式にして返す。

ここで、RAS と WDW におけるデータ分析 (*data analysis*) について定義を述べておく。データ分析とは、次の 3 段階の作業を指す。

- (1) リレーショナルデータベースからテーブルの結合やフィルタリングによって、目的のデータが示される表を取り出す。
- (2) 取り出された表のクロス集計や並べ替えを行う。
- (3) 表としてそのまま表示するか、グラフを生成して可視化を行う。

途中までの段階でもデータ分析と呼ぶ。これらのデータ処理の結果を踏まえた、解釈や仮説は含まない。DAQ は第 1 段階から第 2 段階までの処理を行い、第 3 段階の可視化を RAS

分析項目	観点	根拠資料	具伴項目
▼ 1. 教育の実施体制	1. 基本的組織の編成	(1)学部・学科・専攻の構成・教育目的 更新日：平成22年04月12日 (2)教育研究上の責任部署 (3)専任教員の配置状況(専攻ごと、種.. 更新日：平成22年04月12日 (4)担当教員の配置状況 更新日：平成22年04月12日	*個別別の配置状況 更新日 平成22年04月12日 *学生一人当たり教員数 更新日 平成22年04月12日
		(5)専任教員の配置状況 (6)担当教員の配置状況 更新日：平成22年04月12日 (7)男女別の教員数	*女性教員の割合 *男女共同参画の観点への配慮
		(8)外国人教員数	*外国人教員数 *外国人教員の雇用の方針・工夫
		(9)学生数	*学生定員 *現員 更新日 平成22年04月12日 *定員充足率
		(10)定員充足の適正化に向けた取組	*各種入試の実施状況 *GP *FDの実施状況
		(11)留学生の受け入れ状況	*現員数 *国際入数
		(12)留学生受け入れの増加のための取..	*大学間協定の締結 *留学生確保の工夫の具伴例 *留学生支援の方策の具伴例
		(13)社会人の受け入れ状況	*現員数 *出身別入数 *入数の取組
		(14)女子学生の割合	*入数
	2. 教育内容、教育方法..	(1)教育内容、教育方法の改善に向けた.. (2)具体的な取組とそれに基づき改善の..	*編成図 *部局独自に策定した事業改善計画 *取組の具伴例(教員会議・学生意見.. *改善の具伴例

図 4 レポートオーサリングシステムの外観 (1)

で行う。

### 3. 特 徴

序文において、我々が課題として次の3つの事、(1)報告書における根拠資料として蓄積された電子文書など、質的データの整合的な再利用、(2)データウェアハウスなどに蓄積される量的データの効率的な共有と活用、および(3)これらのシステムの統合、を挙げた。この節では、提案するシステムがこれらの課題にどのように対処しているのかを検証する。

#### 3.1 質的データの整合的な再利用

提案するシステムが、質的データの整合的な再利用に対してどのように貢献しているかを見てみる。

一般に、業務報告書の章節構成は木構造である。多くの業務報告書では、章節の最小単位となっている分析項目や評価の観点を葉として、その他の内点は章と節となっている。そこで本論文では、システムの情報単位であるレポートコンポーネントを葉とした。レポートツリーの構成は可変であり、RAS は既存のレポートコンポーネントに基づく複数の報告書を柔軟に生成することができる。異なる報告書に出現するレポートコンポーネントは、分析項目や観点が本質的に同じであれば同じレポートコンポーネントを転用するので、異なる報告書でも細部における相互の整合性が担保される。

#### 3.2 データ分析の再利用

次に、具体的な大学評価報告書から例を挙げ、量的データの効率的な再利用の観点から、提案するシステムを評価する。

例として、受入留学生の入学状況について、その変化を分析することを考える。各年度各学科の留学生数を算出し、経年変化を示すのであるが、これを行うためには、各年度各学科の留学生数を集計結果を CSV で返す DAQ を、この分析に必要なユーザ (各学科) 全てに作成しなければならない。それには、各ユーザに対して個別に DAQ を作るのではなく、一つのプログラムで複数のユーザに対応できるようにすれば良い。DAQ は restful な HTTP 通信でアクセスされるので、例えば次のような DAQ

`http://wdw/int_stdtd.cgi?yr=5&dpt=eng`

は、工学部 (dpt=eng) の受入留学生数について過去 5 年 (yr=5) に渡る変化を示したデータを CSV 形式で出力する。パラメータの値を変えると、他の学部 of データを出力する。DAQ のパラメータ化を適切に行う事により、同じ観点でのデータ分析プログラムを複数の学科で共有できる。これは、量的データの効率的な再利用を可能にしているといえる。

#### 3.3 疎結合による連携とその効果

最後に、両システムの疎結合による連携について述べる。

RAS における質的データ (文書) と、WDW における量的データ (データウェアハウスのデータ分析) は、DAQ によって対応づけられている。DAQ は restful な HTTP 通信によって実行される。これはウェブサービスの 1 つの形態である。この方式を取る事により、RAS と WDW を互いに密に依存する事のないように設計することができた。図 3 を見てもわかるように、RAS と WDW は DAQ 以外に関連を持たない。WDW から提供される

データ分析は、RAS 以外にも活用され得る。また、WDW 以外のデータ分析も、HTTP でアクセス可能であれば、RAS から参照する事ができる。

### 3.4 「構造化」されていないXMLの氾濫

民間企業では内部統制報告書の管理、公的機関では情報公開法や公文書管理法により、民間企業および公的機関における組織情報は、恒常的に集積されている。このことは組織の大小に関係なく、例えば、教育機関であれば大規模大学から小中学校に至るまで、情報の集積は行われてはいる。これらの情報の大半はワープロファイルやスプレッドシートなど電子化されているが、再利用や活用と言った観点からの本来の意味での構造化はなされていない。

上記の制度や法律に従うためには、集積した情報を効率的に管理し適切な構造化を行い、活用には目的に即して適宜に再構造化を行う、といったことが必要であり、単に情報を集積するだけでは解決できない問題となっている。レポートツリーの構成を、対応すべき報告書の内容に沿って組み替えれば、既存のレポートコンポーネントを参照しつつ、それ以前に作成した報告書の内容と矛盾する事のない報告書を構成することができるのが、このシステムにおける特徴の一つである。

## 4. 関連研究

近年の情報技術の発展は、情報爆発という現象を引き起こしている。特に、大量のHTMLやXMLデータ、企業や機関における電子文書の急速な増加などは顕著である。後々の評価報告書や業務報告書などに再利用する目的で集められる電子文書群について、管理技術や管理方針に関しては課題が散見される。

四ッ谷ほかによる研究<sup>7)</sup>では企業における文書再利用のためのランキング方式を提案している。DITA<sup>5)2)6)</sup>は、半構造データの形式を持つ電子文書を管理するためのアーキテクチャである。特に、プログラミング・ライブラリの電子ドキュメントやヘルプ情報など、既に構造化されている技術文書の効率的なアップデートを実現している。Beyerらが提案する「業務文書文脈解析プラットフォーム」(enterprise content analytics platform)<sup>1)</sup>は、構造化された文書からルールやパターンを発見することで、構造化文書に対する効率的なインデックス作成と検索の手法を示している。これらの先行研究は、既に構造化された文書の検索についての研究といえる。

一方、本論文は、電子化されているが本質的な構造化がなされていない情報を取り扱い、それらの資料の再利用に主眼がある。本論文では、BI (Business Intelligence) の典型



図5 レポートオーサリングシステムの外観(2)

的場面として、業務報告書を取りあげている。蓄積される電子文書は、根拠資料という位置づけで扱われるため、部分的引用はあっても、中身が変更される事はない。構造化されているかどうかを問わず、電子文書はそのまま根拠資料として引用重視で活用しつつ、これらが複数の報告書で引用されても、整合性を担保して利用できるというのが本システムの特徴であり、先行研究とは異なる。

## 5. まとめと今後の課題

本研究では、質的データを管理して業務報告書の作成を支援するレポートオーサリングシステムと、組織情報を効率よく共有するデータウェアハウスを構築した。また、ウェブサービスの枠組みを利用したシステムの結合（疎結合）をにより、連携を実現している（図3）。また、実装は現在、九州大学の評価業務において基幹データベースとして運用されている（図4と図5）。

今後の課題として、次の2つがある。まず、レポートコンポーネントにおける柔軟性の強化である。現在提案しているRASのアーキテクチャでは、レポートコンポーネントのコメントは、固定的に扱われる。報告書によっては、レポートコンポーネント間の文脈や、書きぶりをそろえるなどの調整の必要性がある。

次に、DAQにより得られるデータ分析の柔軟性の強化である。クロス集計など、データ分析の機能は、WDWの内部機能としており、レポートコンポーネントでは、データの可視化しか行わない。よって、ユーザが自由にクロス集計をすることができない。本論文の著者らの研究であるマッシュアップ的手法<sup>4)3)</sup>は、この課題に対して有効な手段でないかと考える。

謝辞 九州大学の各部局へのヒアリングを行い、データ収集の組織的枠組を作っていたいた関隆宏氏（現新潟大学企画戦略本部評価センター准教授）に感謝します。なお、本研究の一部は、科学研究費補助金課題（基盤研究（C））「大学経営における総合的問題の分析と解決を促進する情報基盤の構築（課題番号22530921）」によるものです。

## 参 考 文 献

- 1) Beyer, K., Ercegovic, V. and et al., R.K.: Towards a Scalable Enterprise Content Analytics Platform, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, pp.28-35 (2009).
- 2) Diaz, O., Anfurutia, F.I. and Kortabitarte, J.: Using DITA for documenting software product lines, *Proceedings of the 9th ACM symposium on Document engineering*, Association for Computing Machinery, pp.231-240 (2009).
- 3) Mori, M., Nakatoh, T. and Hirokawa, S.: Functional Composition of Web Databases, *Proceedings of International Conference Asian Digital Libraries 2006*, Lecture note in Computer Science, Vol.4312, Springer Verlag, pp.1439-448 (2006).
- 4) Mori, M., Nakatoh, T. and Hirokawa, S.: Links and Cycles of Web Databases, *The 4th Italian Workshop on Semantic Web Applications and Perspectives*, pp.21-30

(2007).

- 5) Priestley, M.: DITA XML: A Reuse by Reference Architecture for Technical Documentation, *Proceedings of SIGDOC'01*, pp.152-156 (2001). Santa Fe, New Mexico, USA.
- 6) 加藤哲義：DITAによるコンテンツの構造化・再利用のメリット，デジタルドキュメントシンポジウム2009 (2009).
- 7) 四ツ谷雅輝，松林忠孝，弥生隆明，野田十悟，吉田 豊：ディレクトリ構造に着目した企業内文書向けランキング方式，*DBSJ Letters*, Vol.4, No.2, pp.1-4 (2005).