

大学評価の報告書作成支援システムと大学情報の データウェアハウスについて

森, 雅生
九州大学大学評価情報室

田中, 要江
九州大学大学評価情報室

廣川, 佐千男
九州大学情報基盤研究開発センター

<https://hdl.handle.net/2324/18817>

出版情報 : 日本教育情報学会 第26回年会, 2010-08
バージョン :
権利関係 :

大学評価の報告書作成支援システムと大学情報の データウェアハウスについて

森 雅生¹ 田中 要江² 廣川 佐千男³

<概要> 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的に困難である。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も課題である。本研究では、大学評価の評価書作成を統合的に支援するシステムと、データウェアハウスとの軽量の統合システムを開発した。

<キーワード> デジタルドキュメント、データウェアハウス、大学評価、ウェブサービス

1 はじめに

すべての日本の大学は、教育研究の質を保証するため、認証評価を受審する義務がある。更に、国立大学法人は大学運営の説明責任を果たすために、法人評価を受けることが義務づけられている。これらを大学評価と呼び、大学が自ら教育研究の活動の自己点検・評価を定期的に行い、評価基準に沿って自己評価書を既述する事が前提とされている。評価の根拠として、大学における様々な電子化された文書（以下、文書）や学務情報をはじめとする大学業務データの活用が不可欠である。

しかしながら、大学の教育研究活動は多岐に渡るため、担当者は情報の収集に苦慮している。また、学内の数々の委員会や会議の資料を保管してはいるものの、系統的な管理には至らず、再利用の効果的活用は大きな課題となっている。さらに、学生の基本情報や成績情報、教員の研究資金や財務情報等といった量的データのデータウェアハウス化も課題の一つである。

この研究では、大学評価に対応する報告書作成システムを開発した。このシステムは、文書オーサリングシステムとデータウェアハウスの連携を実現したものである。本研究が取り組む課題は、(1) 互いに矛盾しないような複数の報告書を作成する手法、および (2) データウェアハウスからの量的データと文書データを軽量的に統合する手法、の2点を明らかにする事である。

2 関連研究

近年、情報技術の進歩は「情報爆発」という現象を引き起した。ウェブデータの爆発的増加や、企業および大学などの組織に蓄積されている電子化された情報の急激な増加などはその例である。特に、再利用を目的として組織内に蓄積された大量のデータは、明確な指針を立てにくいことから、管理手法について課題が残されており、それを解決するための手法が研究されている。例えば、Morimotoら [7] は、文書型の知識を再利用する4つの段階として、

1. 文書の収集と蓄積、
2. 文書の検索と表示、
3. 文書の抽出と同定、
4. 新しい文書の作成、

を示した。これに沿って蓄積文書から新しい文書を生成するプロセスについて考察する。

文書の蓄積には、文書の中を検索する機能が必要である。Beyerら [1] は、構造化された文書から、テキストのパターンやルールを発見する事で、効率的に検索インデックスを作る研究を提案している。また、文書の関連性を Linked Data [2] として実現できるのであれば、文書の検索もより効果的になるであろう。しかし、大学の業務文書を管理する上で、まず必要であったのは、文書の所在や作成者、文書の目的などのメタ情報であった。これは、大学が多く自立した部局や事務組織からなる複合組織であるため、文書の属性を同定するために、これらのメタ情報は欠かすことができない。この論文で提案するシ

表1：九州大学法人評価の現況調査表等の状況

Fields	学部学府	報告書の内容		
		節	観点	ページ
教育 研究	31	8	12	959
	20	5	5	311

¹九州大学 大学評価情報室 mori@ir.kyushu-u.ac.jp

²九州大学 大学評価情報室 tanaka@ir.kyushu-u.ac.jp

³九州大学 情報基盤研究開発センター hirokawa@cc.kyushu-u.ac.jp

システムには、報告書に必要な文書のメタ情報を統合的に管理する機能を備えている。

DITA[9]は、構造化された文書の管理を効率的に行うために提案された文書管理アーキテクチャである。これを用いれば、複数の文書で抽出や変更を行うことも可能である[3]。DITAやLinked Dataを採用するには、文書の構造化と、セマンティックウェブの意味での統一的なオントロジーの構築が必要である。しかし、レガシーシステムと現行する業務プロセスに対し、オントロジーを構築し適用する事は大変難しい。例えば、提案するシステムでは、文書の所在を表すためにRDFなどを用いず、具体的な場所を示すこと(例えば、100頁、第5章、第4節のように)にした。

企業では、毎日の業務報告を蓄積することによって、情報開示や報告書に対応している。文書型の知識を活用することも望ましい[7]。しかし、大学の現場においては、日常業務における書類発生に比べて、求められる報告書の作成頻度は低い。高々、年単位、短くても月単位である。OLAPと協同する報告書作成[8]のように、即時性を求められる訳ではない。大学における経営の視点は、長期的なものである。求められる報告書作成も、どの資料や文書を使い、どのようにまとめるかという、高度な判断が求められる。特に、評価書作成の局面では、評価の目的に応じた資料の選択や判断が必要である。求められているのは、こうした選択や判断を支援するシステムであり、自動化する事ではない。本稿では、報告書を取りまとめるために、どの文書または資料をどのように構成するかを対話的に支援するシステムを構築した。

このように構造化されずに蓄積したデータと、データウェアハウスのような構造化されたデータを統合して文書を作成する仕組みの研究は、この10年間に多くなされてきた[2][4][8][10]。これらの研究は、情報検索に基づくものや、構造化された情報のオントロジーを仮定されていることが多く、量的データを分析する手法を探索する事が目的であった。この研究は、多様な報告書に対応することを目的とし、構造化やオントロジーを仮定せずに文書作成を手助けする仕組みを提案している。

表1は、九州大学が第1期法人評価の際に提出した現況調査表等の規模を表したものである。ページ数の良し悪しは別にしても、文書の記述に対し整合性を担保するには、人力では困難である。提案するシステムの一つの目的は、報告書の整合性を担保する事でもある。

3 システムの概要

3.1 文書オーサリングシステム

この節では、本研究で提案するシステムの文書管理を行う部分である文書オーサリングシステム(document authoring system, 略称 DAS)を解説する(図1左)。このシステムにおける情報単位を評価トピックと呼ぶ。評価トピックは以下の5つの情報からなる。

1. トピックタイトル
2. 自己評価記述
3. Data analysis query (URL)
4. 関連文書の添付
5. メタ情報

トピックタイトルは、評価の観点名や基準の項目名である。このトピックについて、自己評価記述とデータウェアハウスから提供されるデータ(CSV形式)のURL、および関連する文書の添付の5要素をひとまとめにしたものが評価トピックである。特に3番目の根拠データのURLは、データウェアハウス内で実行されるデータ分析クエリ(data analysis query, 略称 DAQ)を指す。メタ情報は、評価トピックの所有者や学部名などの管理情報である。

自己評価報告書の文書構造は、評価トピックを葉とし章節を枝とする木構造を与える事によって定義する。これを報告書項目木(report tree)と呼ぶ。報告書項目木は、求められる評価やユーザの個別のニーズによって定義を変える事ができる。

例えば、評価トピック「教員数」は、法人評価と認証評価のいずれにも取り上げられる項目であるが、DASでは同じ評価トピックを参照するので、整合性を担保することができる。報告書に必要な評価トピックを準備し、項目木を定義すると、ユーザはXML形式の報告書を得ることができる。このように各評価トピックごとに自己評価の記述と関連文書および根拠データの対応を管理し、報告書を作成する場合には項目木を定義するだけで、報告書における整合性を実現できる。

ユーザは、各評価トピックの自己評価記述および関連文書の添付を行うが、それぞれ個別なものとして独立に取り扱う。管理者は、報告書項目木の定義を大学評価の種類に応じて設定する。実装には、Ruby on RailsとMySQLを用いた。DASのデモンストレーションは、YouTube⁴で見ることができる。

⁴<http://www.youtube.com/watch?v=okAT6aseks8>

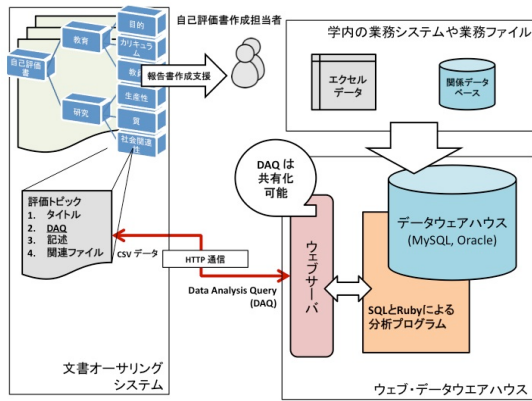


図 1：システム概要

3.2 データウェアハウス

この節では、評価の根拠データとして活用される情報を、一元的に格納し分析や共有を行うためのデータウェアハウス(data warehouse, 略称 DW) について述べる (図 1 右)。

DW の情報源は、学内の業務用システム (学生, 教員, 教育, 研究, 財務) からのバックアップや、スプレッドシートなどのフラットファイルである。手動による整理を経て DW 内の関係データベースに格納される。データの分析処理は、SQL やスクリプト言語によりプログラムが可能である。これらの分析処理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、HTTP を介してアクセスは行う CGI として実装される。また、DW は DAQ によるデータの要求に対して CSV 形式でデータを返す。

例えば「学生数」という評価トピックでは、学部ごと年度ごとに算出しなければならない。このとき、評価トピックが DW から得る CSV データは、DW 内で構成された分析処理プログラムから得ることができる。分析処理プログラムには、CGI として実装されている DAQ を介してアクセスすることができる。例えば、次の DAQ

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

は、過去 5 年分の工学部 (*engineering*) の受け入れ留学生数の推移のデータが得られる DAQ である。パラメータ *yr*(year) や *dpt*(department) の値を変える事により、異なる学部や年度のデータを得られる。このように両システムの連携を、WebAPI の形態を取って疎結合することにより、軽量の統合を実現することができた。DAQ は、この論文で提案している文書オーサリングシステムのみならず、他のサービスでも利用することができる。

現在、九州大学で運用している大学情報データウェアハウスには、主に学校基本調査のバックアップデータを再利用したデータベースを装備している。九州大学大学評価情報室では、このリサイクルデータ [11] を用いて、主に教育に関する活動状況を経年で捉える事のできるデータ集「ファクトブック」を毎年発行している [12]。このデータ集の特徴の一つに、新規の調査を全く行っていない事が挙げられる。大学評価という、新しい業務が義務づけられた大学の現場では、慣れない評価業務からくる「評価疲れ」が指摘されている。ファクトブックの出版は、負担の軽減を図りつつ、ファクトデータに基づく大学経営を推進することを目的とした取り組みである。

4 特徴

この研究の課題は、(1) 互いに矛盾することなく、複数の評価報告書を作成する手法、および (2) 量的データと文書データの軽量の統合方法、の 2 つを明らかにする事であると述べた。この節では、この 2 つの点についてシステムを評価する。

まず、提案するシステムが大学評価における報告書作成支援に関して、文書データの整合的な再利用をどのように実現しているかを考察する。

大学評価における自己点検・評価報告書の規準では、観点や分析項目が完全な木構造によって整理されている。この木構造の葉の部分には、大学評価において普遍的な話題となっている事が多い。そこで、その葉の部分の基本単位とした評価トピックを定め、規準に沿って報告書項目木を与えることで、報告書を生成するアーキテクチャとした。複数の異なる報告書がある場合は、それぞれ規準に沿った項目木を与えることで、評価トピックの整合的な再利用が可能となった。

次に、データウェアハウスとの軽量の統合連携に就いて述べる。

DAQ を CSV 形式のデータを返す CGI とする事で、DW と他のシステムとの連携を軽量的に実現することができた。従来のデータウェアハウスや BI では、分析処理プログラムや、コメント機能などを一つのシステムの中に組み込んでいるが、提案する DW のようにデータ分析機能と他の機能との密な結合を排除した事で、データ分析プログラムの共有や再利用が可能となっている。

管理者は、適度に汎用性を持たせたデータ処理プログラムを作成し DAQ とする。ユーザは、DAQ のパラメータを変更する事により、他の用途にも活用できる。実装は、DW の情報源であ

る業務システムにあわせた関係データベース群 (Oracle, MySQL, MS-Access) を用いた。



図 2 : 画面イメージ

5 まとめと今後の課題

この論文では、大学評価の評価書作成を総合的に支援するシステムと、データウェアハウスとの軽量の統合システムを開発した。データウェアハウスへのアクセス方法として、HTTP通信で実現するDAQを提案した。また、多様な報告書への対応について、再構成可能な評価トピックを葉とする木構造で報告書を生成する事で、評価トピックの総合的な再利用を実現している。

評価トピックは独立して編集するため、報告書の文脈による書きぶりの変更や、DAQにより得られるデータ加工をユーザが自由に行う事はできない。このような、評価トピックへの柔軟性は今後の課題であるが、DAQで得られるデータの加工については、筆者らが提案しているウェブ・マッシュアップ的手法[6][5]をユーザインターフェイスに導入する事は有効であると考える。

参考文献

- [1] Kevin Beyer, Vuk Ercegovic, and Rajasekar Krishnamurthy et al. Towards a scalable enterprise content analytics platform. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2009.
- [2] Christian Bizer, Tom Health, Kingsley Idehen, and Tim Berners-Lee. Linked data on the web

(ldow2008). In *Proceeding of the 17th international conference on World Wide Web*, pages 1265–1266, 2008.

- [3] Oscar Diaz, Felipe I. Anfurrutia, and Jon Kortabitarte. Using dita for documenting software product lines. In *Proceedings of the 9th ACM symposium on Document engineering*, pages 231–240. Association for Computing Machinery, 2009.
- [4] Antonio Ferrández and Jesús Peral. The benefits of the interaction between data warehouses and question answering. In *Proceedings of the 2010 EDBT/ICDT Workshops*, volume 426 of *ACM International Conference Proceeding Series*, 2010.
- [5] Masao Mori, Tetsuya Nakatoh, and Sachio Hirokawa. Functional composition of web databases. In *Proceedings of International Conference Asian Digital Libraries 2006*, Lecture Note in Computer Science 4312, pages 439–448. Springer Verlag, 2006.
- [6] Masao Mori, Tetsuya Nakatoh, and Sachio Hirokawa. Links and cycles of web databases. In *The 4th Italian Workshop on Semantic Web Applications and Perspectives*, pages 21–30, 2007.
- [7] Yukiko Morimoto, Hisao Mase, and Hiroshi Tsuji. Perspectives on reuse process support systems for document-type knowledge. In *Human-Computer Interaction, Part IV, HCII 2007*, Lecture Note in Computer Science 4553, pages 682–691, 2007.
- [8] Torsten Priebe and Günter Pernul. Ontology-based integration of olap and information retrieval. In *Proceedings of the 14th International Workshop on Database and Expert Systems Applications*, page 610, 2003.
- [9] Michael Priestley. Dita xml: A reuse by reference architecture for technical documentation. In *Proceedings of SIGDOC'01*, pages 152–156, October 2001. Santa Fe, New Mexico, USA.
- [10] B. Riger, A. Kleber, and E. von Maur. Metadatabased integration of qualitative and quantitative information resources approaching knowledge management. In *ECIS 2000 Proceedings*, 2000.
- [11] 森 雅生. リサイクルデータを用いた大学情報のデータベース化について. *大学探求*, 1(2):31–38, 2009.
- [12] 佐藤 仁, 森 雅生, 高田 英一, and 小湊 卓夫. 大学情報の組織内共有と活用—九州大学大学評価情報室の取り組みから—. *大学探求*, 1(2):1–11, 2009.