

複数にまたがるシステムで学習行動履歴を統合する ログ管理システム

A Log Management System to Integrate Study Behavior History in Multiple Systems

藤井稔也[†] 篠原信夫[†]
Toshiya Fujii[†] Nobuo Shinohara[†]

[†] 国際医療福祉大学大学院 国際福祉学研究所

[†] Graduate School of Health and Welfare, International University of Health and Welfare.

要旨

オープンソースやプロプライエタリなシステムを利用し追加や改良を年々施しながら稼働させている組織は世に多く存在するであろう。大学における学習支援システムも例外ではないと思われる。その場合、学習者の行動ログはそれぞれのシステム内に個々に蓄えられるため、それらログを活かしきれない。統合ログ管理サーバという仕組みが一般的であれば、それを導入すればよいが実際問題として現状で適用できる解決策はない。本研究では、スクリプト言語とリレーショナルデータベースによるラピッド開発手法を用い、このような事例に対処し、より汎用的な解決策があるかを検討した。

1. はじめに

多くの大学、大学院では従来の対面授業に収まらない形態の教育を模索している。その一つの方向性として予め用意した映像、授業資料や確認試験をオンライン上で行う VOD (Video On Demand) 形式や、対面授業を録画編集し、やむを得ない理由による欠席者への学習機会の提供をする方式などがある。国際医療福祉大学大学院（以下、当大学院と表記する）は、社会人学生が多く所属することから、このような方法の需要度が高く、平成19年度より一部の科目から VOD システムの導入を試みてきている。これらのシステムは密に統合されたものというよりは、色々な機器やシステムの併存であり、有効に活用するには運用上の工夫が必要となるのが常態となっている。

学習支援という観点から考えると、学生の学習行動の把握により、個々のより良い学習の形成に繋がるとして、結果として指導する教員の成績処理の手間も省力化されるものが優れたシステムといえる。しかし、統一的に設計されていないシステムにおいてはアクセスログ自体が分散して、データベース形式も各々異なったものであるため、このような用途に答えるためには、システムを統一するログ管理システムが必要となるだろう。

当大学院では、オープンソースの Moodle [1] を利用した学習支援システムを中核とし、授業映像等の供給元として Windows Media ストリーミングサーバを利用している。また、学外の医療関連従事者を対象とするサービスである乃木坂スクール(医療福祉 e チャンネル)も学生向けに一部利用しているという構成になっている。そのため、全体的な学生の学習履歴の把握には、それらのログデータを照合し解析する必要があったが、これまでは未着手であった。本研究では、それらのログデータを蓄積し利用が容易にできること、学生の学習行動の履歴の形式化や一般化について探ることを目的とした。

2. 関連研究

従来、アクセスログはシステムに障害が発生した場合にその原因を探るためといった特定の目的には不可欠であると考えられていた。しかし、個々のシステムだけではなく統合的に管理しようという試みとして工業規格、国際規格に提案されたものはなく、現状として各々のシステムは独自形式のログを生成している。

アクセスログの重要性が注目されだした一因はウェブにある。検索エンジンサービスを提供する Google は創業の比較的早い時期から、ユーザのアクセス履歴に秘められた価値があると気づいていた [2]。その後の Google の躍進から、現在、その認識は広く共有されるようになってきている。また、近年、開発が進められているお薦め商品などを割り出すリコメンド・エンジンは正にそのような履歴データからのデータマイニング技術といえる。

データマイニングの前段階の処理をするものには、Fluentd [3]がログ収集と活用のための枠組みとしてオープンソースで提供されている。データフォーマットとして JSON 形式を採用し、入力、出力のプラグイン構造を持つ柔軟なシステムである。これはビッグデータと呼ばれる大規模な量のデータを扱うことを目的として作られているため数千人利用レベルの比較的小規模のデータ処理には適用しにくい。

統合ログ管理システムの研究としては、神尾らによる不正アクセスへの対策として分散したアクセスノードを統一する集中ログ管理システム [4]や、森山らによる行動履歴追跡機能をもつログ管理システム [5]、敷田らの部品依存関係に基づくログ管理支援法 [6]がある。いずれも大規模なシステムを志向したものであり、当大学院のシステムに適用するのは無理があった。友野らは、異なる種類のログを YAML 形式として統一して保管することを提案している [7]。これは保存法に重点を置いた研究であり、これも利用できるものではなかった。

3. 方法

当大学院においては、学習支援システムとして Moodle を採用し学生や教員のユーザ管理を行なっている。VOD 科目においては、Moodle のコース機能を個々の科目に当てはめ、科目内を15のコマに分け、それぞれに授業資料、授業映像、確認問題を設けている。授業資料は、Moodle 内にアップロードされた PDF 書類等を参照しており、学生はクリック操作でダウンロード出来る。授業映像は、別の機器である Windows Media サーバのコンテンツを参照しており、学生がクリック操作を行うとメディアプレーヤアプリケーションが起動する。また、確認問題には、Moodle の小テスト機能を利用している (図1)。

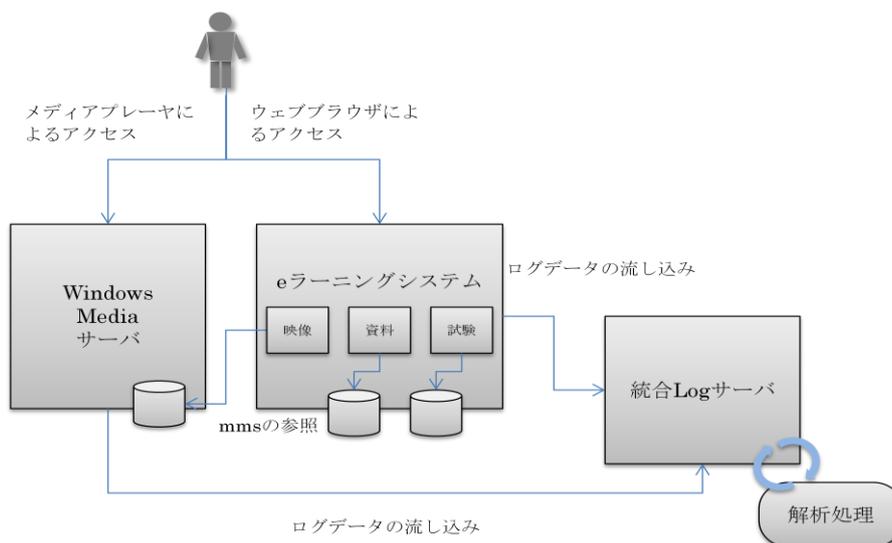


図1 システム全体図

授業資料の閲覧と確認問題の回答は全て Moodle のログとして完結するが、ストリーミング映像である授業映像は、Moodle システムとして視聴開始は取得できるが視聴終了や視聴持続状態の情報は取得できない状況にある。Moodle ログ情報だけを利用し実際に通して見たのか判断しないとすると学生の不正を許してしまう危険性がある。しかし、それらの情報は Windows Media サーバのログに存在しており、

それと統合することによって始めて完全な情報を得ることが出来る。

本研究では、そのような情報の蓄積と利用を容易にし、学生の学習行動の履歴の形式化や一般化が出来ることを第一の目標としてデータ形式の設計を行いリレーショナルデータベースに実装し、ログデータを流し込んで評価を行った。表1はその形式である。

表1 学習行動履歴レコード形式 (一部抜粋)

欄	型	意味
id	serial	主キー
source_name	varchar(100)	ログデータソースシステム名
source_id	integer	ソース ID
source_continuous_number	integer	関連付け番号
access_time	timestamp	アクセス時間
access_ip	varchar(50)	アクセス IP 番号
access_behavior_name	varchar(50)	アクセス種別
access_url	varchar(500)	アクセスされた URL
target_item_name	varchar(100)	対象の名称
target_item_course_name	varchar(50)	対象の科目
target_item_category_name	varchar(100)	対象のカテゴリ名
user_id	varchar(50)	ユーザ ID
user_title_name	varchar(50)	ユーザの種別
user_belonging_name	varchar(50)	ユーザの所属
user_surname	varchar(100)	ユーザの苗字
user_middlename	varchar(100)	ユーザのミドルネーム
user_forename	varchar(100)	ユーザの名前
appendix_information	varchar(1000)	補助情報

データベースには Postgresql を、操作のためのスクリプト言語には PHP を用いた。バージョンはそれぞれ 8.4.7 と 5.3.2 である。流し込む元の Moodle のログ構造には、アクセス時間、ユーザ ID、アクセス元の IP 番号、コース ID、モジュール ID などがあり、他方、Windows Media サーバのログ構造には、アクセス時間、アクセス元の IP 番号、操作種別、操作からの持続時間等の情報が存在する。Moodle 上でのユーザ操作を発端として Windows Media サーバのアクセスが発生しても、ログデータ上では、両者の関連性の直接的な情報は存在せず、結びつけには両者のアクセス時間と IP 番号の照合をする必要がある。そのため、PHP を用いて Moodle の授業映像アクセスのログのアクセス IP 番号をキーにし、その時間以降に出現する Windows Media サーバの適合するログレコードを抽出し関連付け処理を行った。

4. 結果と考察

今回は、Moodle システムからはリアルタイムにスクリプト処理でログ管理システムにログデータを流し込んだが、Windows Media サーバのログデータは CSV 形式で取り出し手動で流し込み処理をした。そのため、ログレコードは時系列順には並んでいないが解析・照合処理すること自体に支障はなかったことが確かめられた。この照合処理により Moodle のログだけでは解らない視聴の細かな状況が抽出できたが、Windows Media サーバ固有情報である再生、早送り、巻き戻し等の詳細なユーザ操作の情報解析は行なっていないため、学生が映像を最初から最後まで通して見たかどうかの判定は実際にはできていない。これは今後の実装の課題となる。

今回の実装だけではなく、複数のシステムからログを流し込むシステムにおいては本質的に時系列順にデータが並んでいることは保証できない。これは問い合わせ処理を行う側が、全てのデータを読み込

んで整合性を確かめる必要があることを意味する。それに対しログ管理システム側が、時系列に並べ替える処理をログが追加される度に行うという処理を行うか、ログ追加とは関係なく一定間隔で行うことになる。前者は、ログ書き込み処理の負荷が高くなり、後者は、負荷は低い追加された未処理のデータが混ざる可能性を残しているため、付加的なデータの正当性を保証する仕組みが必要になる。今回は前者の実装を採用したが、後者とのさらなる比較検討が必要である。

今回の設計においては、システム由来の ID 番号情報は無意味であると考え、その ID を一般的な形式、例えば元システムでのユーザ ID を学籍番号や名前に置き換えるなどの処理を施した。しかし ID 情報は、元システムの元データベースを参照する仕組みがあれば、後でログデータベースを更新することも可能である。今回の例で言えば Moodle のログにあるユーザ ID やコース ID のような情報はそのまま記録しておくことにあたる。これはデータ形式の一般化とは反対方向の考え方でもあり設計のバランスが問われるかもしれない。

最後に、ログデータはどのシステムであっても運用していくことにより日々、少しずつ増大していく一方のものである。システムの安定的な運用を考えると、ストレージを逼迫する可能性があるデータは、物理的に分離した別の機器で保存しておくことが望ましい。複数のシステムを運用する環境においては、個々のシステムの安定性は重要であり、その面でも、統合ログ管理サーバはクローズアップされていくと考える。

参考文献

- [1] Moodle, <https://moodle.org>.
- [2] スティーブン・レヴィ, ゲーグル ネット覇者の真実, 阪急コミュニケーションズ, 2011, pp. 75-76.
- [3] Fluentd, <http://fluentd.org/>.
- [4] 神尾政和, 石田常竹, ログの統一管理及び以上検出に関する研究, 情報処理学会研究報告 CSEC, Vol.2004, No.129 2004, pp. 77-82.
- [5] 森山令子, 平井規郎, 郡光則, 行動履歴追跡機能を備えた統合ログ管理システム, Vol.8, No.2, 情報科学技術フォーラム講演論文集 2009, pp. 239-240.
- [6] 敷田幹文, 後藤宏志, 大規模サーバー間の部品依存関係に基づくログ管理支援法, 情報処理学会論文誌, Vol.49, No.3 2008, pp. 1081-1089.
- [7] 友野敬大, 上原稔, 島田裕次, 異なる種類のログの統合的保管に関する提案, 情報科学技術フォーラム講演論文集, Vol.9 No.4, 2010, pp. 303-304.