

システムログイン情報を可視化するためのグループウェアと WebAPIによるダッシュボード構築の検討

Developing a Dashboard for Visualization of Login Information with a Combination of Groupware and WebAPI

久保田真一郎 †, 中野裕司 †, 武藏泰雄 †, 戸田真志 †,
右田雅裕 †, 喜多敏博 †, 松葉龍一 †, 杉谷賢一 †

Shin-Ichiro Kubota†, Hiroshi Nakano†, Yasuo Muasashi†, Masashi Toda†,
Masahiro Migita†, Toshihiro Kita†, Ryuichi Matsuba†, Kenichi Sugitani†

kubota@cc.kumamoto-u.ac.jp

熊本大学 †
Kumamoto University †

概要

組織内で運用される様々な Web システムへのログインを統一的行うために、統合認証システムが運用されている。このログデータを用いて各種システムへのログイン状況を可視化し、WebAPIによりグループウェア上で表示できるように整備する検討を行った。グループウェアによりアクセス可能なユーザや権限を管理することで、情報格付け取扱い手順に従い、関係者のみが閲覧および操作する環境を提供することを目指す。

キーワード

システムログ, 可視化, IR, ダッシュボード

1 はじめに

日頃、学内で利用される様々な Web サービスの多くは、個人ユーザが利用するための認証機能を持っており、認証の一元化を実現する仕組みとして、シングルサインオンによる運用が行われる。このシングルサインオンを実現する仕組みとして、本学では CAS 認証が導入されており、その CAS 認証サーバには、毎秒蓄積される膨大な認証データのログが記録されている。システムへのログイン情報は、そのシステムの運用状況を把握する上での 1 つの指標を構成できる。将来の情報システム利用を考え、意思決定する上で必要な情報の 1 つであろう。しかし、記録されるログデータはテキスト情報でその運用状況を把握するには困難な形式である。将来の情報システムを検討するような意思決定が行われる場

面では、記録されたデータをしばしば可視化して情報共有される。また、特に必要な情報を可視化して 1 か所に集めて表示するインターフェースをダッシュボードと呼んでいる。本稿では、大学 IR の基盤となるデータの収集、クレンジング処理、分析、可視化、表示インターフェースであるダッシュボードの構築について扱い、情報システムの認証データを用いた可視化およびダッシュボードの構成について検討する。

1.1 統合認証システムの認証ログを扱う利点

本節では、統合認証システムの認証ログを扱う利点について整理しておく。統合認証がない場合には複数のサーバに認証ログが記録されるが、統合認証システムを経由して様々な Web システムが利用されるため、統

合認証サーバが記録するログに認証された情報が集約される。それぞれのサーバに記録された認証ログを集めて集計処理する必要がなく、認証ログデータを共有する際の安全面においても有効であると考えている。また、CAS 認証の場合、認証が行われる際に発行されるチケットを追跡することでユーザの利用傾向なども把握できると考えている。

2 システム構築について

通常、ログイン情報の集計結果などは管理運営のために利用される情報であり、不特定多数のユーザが閲覧するわけではなく、特定のユーザによって閲覧され意思決定が行われる。また、ログイン情報の集計結果のように管理運営の意思決定に利用されるような情報は、時々刻々と変化するニーズのために、意思決定を行うユーザがニーズに合わせた情報を求めると考えられる。そのため、専用のシステムにより決められた形式で情報を表現したのでは、時々刻々と変化するニーズに対応できないと考えられる。つまり、管理運営上、意思決定に用いられる情報は、

- 特定のユーザが閲覧できること
- ユーザの要望に合わせた可視化ができること

の2つがあげられると考えている。

グループウェアを使うことで特定のユーザが閲覧できる Web ページを構成できるため、グループウェア上で必要なユーザのみがアクセスできる Web ページを構成し、意思決定のために求められる情報を可視化して表現することで、これらの仕様を満たす。ただの Web ページで構成した場合、ユーザの要望にあわせてグラフ画像や集計表を記述することで情報提供することは可能であるが、提示情報を一部変更するなどの要望を受けてから、更新するまでの時間が必要であり、意思決定が行われるその場で提示情報を編集して提示することは困難である。提示情報を一部変更するよう求められ、可能な限りその場で対応できるようにするにはインタラクティブな Web ページを構成する必要がある。一般には Web プログラミングを使って実現する。先に示した特定ユーザのみに制限するページを構成すると Web プログラミングによってアクセスを制御する部分まで記述が必要となる。グループウェアのアクセス制御機能を利用しつつ、Web プログラミングにより、意思決定の場での簡易な編集も可能とする仕組みとして WebAPI が利用できる。WebAPI 開発には、相応の労力が必要となるが、集計処理を得意とする R 言語を利用し、その集計結果を WebAPI で提示する R 言語のライブラリを用いることでこれを解決する。

2.1 扱うデータについて

本報告で扱う情報は統合認証システムを使ったログイン情報で、syslog により蓄積されたテキストファイルが対象となる。このような素データを集計処理して可視化するまでには、集計処理を行うためのデータ化の作業（データクレンジング）が必要である。今回は syslog により蓄積されたログ情報をもとに必要な情報を手動で取り出し、R 言語で集計処理するための CSV を生成した上で行う。今回はデータクレンジングの自動化および記録手法については扱わず、理想的にデータにアクセスできる状態を仮定し、ログイン情報を可視化する手法について報告する。本来であれば情報を一元管理するためにデータクレンジングとそれを記録する手順も含めて自動化されなければならない。データクレンジングの自動化および記録手法については別途報告したい。

本学で動作する統合認証システムについては永井ら [1] により詳しく報告されている。本取り組みで扱うログイン情報は CAS 認証で出力されるログのうち Service Ticket に関する記録に着目して収集し、データクレンジングを行った。Service Ticket は認証通過後に利用されるサービスごとに発行されるチケットであるため、ユーザが CAS 認証通過後に使用した Web システムのすべてを確認することができる。

本来であれば、一元管理とアクセスを容易にするために、データクレンジングを行ったデータをデータベースに記録して利用するが、今回は理想的にデータベースへアクセス可能であるとして、可視化処理部分についてのみ扱う。本報告では可視化処理までの流れを確認するため、データベースへのアクセスの代わりに、R 言語で扱われるオブジェクトを保存したファイルを読み込み、その後処理することとした。R 言語ではプログラム内で定義したオブジェクトをそのままファイルとして保存することができ、保存したファイルは他の処理プログラムで読み込んで利用することができる。今回は、ログデータをデータクレンジングした後に、静的に R 言語プログラムで読み込み、1つのオブジェクトとして整形処理し、保存したファイルを利用する。

データクレンジングされたログデータは、日時、サーバ URI、アクセスしたファイル、Service Ticket の番号、認証されたユーザ名を項目として持つが、R 言語による静的な集計処理で、「10分ごとの日時」、「サーバ URI」、「10分間で認証されたユーザの数」の3項目を持つオブジェクトを1時間ごとに生成するようにした。また、毎時生成される度に処理時間の24時間前からのデータを保持するオブジェクトも同時に作成し、その24時間分のデータをもつオブジェクトのファイルを1時間ごとに更新するようにした。本来はデータベースへの問い合わせで、WebAPI サービスにアクセスした時間をもとに、

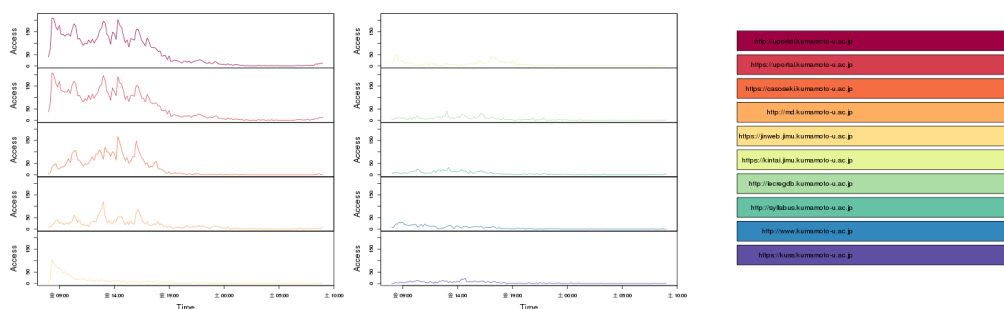


図- 1: グループウェアでの WebAPI 表示

その時刻から 24 時間前までの情報を取り出し、集計処理することで動的に情報が更新されるグラフや表を構成可能となる。

2.2 システム構成

先に述べたグループウェアと R 言語による WebAPI により集計処理結果を制限された Web ページで表示するシステムを構成する。集計処理して表示する情報は不特定多数の人が閲覧できないように、セキュリティに配慮したシステム構成が必要となる。

主なシステム構成要素は次の 3 つとなる。グループウェアはいくつかの条件を含む。

- R 言語処理するためのシステム
- グループウェア
 - － 認証によるログイン
 - － ユーザ単位で利用制限可能なページを構成可能
 - － WebAPI を読み込み表示可能
- ログデータを保持するデータベース

これらのシステム構成要素がネットワークで接続され連携して利用できることが求められる。そのため、R 言語処理するためのシステムとして RStudio Server [2] を利用することとした。グループウェアについては現在学内で運用中である Confluence を利用することとした。本報告では、可視化処理部分についてのみ扱うため、データベースへのアクセスの代わりに、R 言語で扱われるオブジェクトを保存したファイルを読み込み処理することとした。そのオブジェクトのファイルを読み込み、RStudio Server 上に置き、そのオブジェクトファイルを読み込み、集計処理することとした。

今回実装して確かめた処理の流れは次のようになる。

1. R 言語オブジェクトのファイルを読み込む（データベースへのアクセスの代わり）
2. RStudio Server で集計処理する
3. RStudio Server で WebAPI サービスを提供する
4. グループウェア Confluence の特定ページで表示する

今回利用する RStudio Server はすでに学内に蓄積するデータへのアクセスを許可されたサーバ上に構築されており、特定のネットワークからのみアクセス可能なサーバとなっている。学外からはまったくアクセスできず、学内であっても特定のネットワークからのアクセスに制限されている。

RStudio Server での集計処理は、オブジェクトを読み込み、24 時間前からの認証した人数を 10 分ごとにプロットしたグラフを生成する処理を行っている。集計処理した結果、過去にログインされているサーバが 81 個あったため、上位 10 個のサーバに限ってプロットし表示することにした。WebAPI サービスの提供には、R 言語のライブラリ Plumber [3] を用いた。記述方法などはマニュアル [4] を参考に開発した。

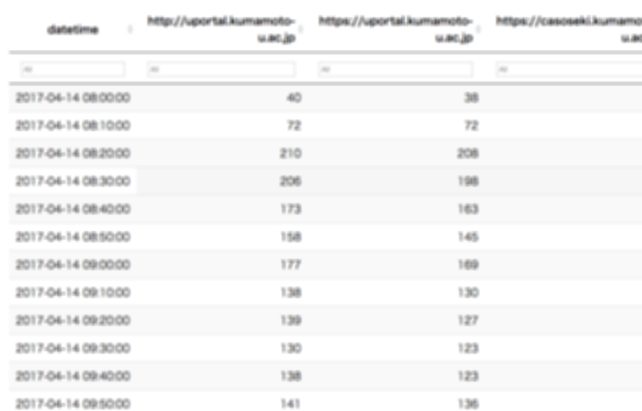
先に述べたように、WebAPI でサービスされる情報に不特定多数の人がアクセスできないように、RStudio Server ではグループウェアのサーバからのみアクセスできるようにネットワークのアクセス制限などの設定が必要となる。また、クロスドメインの制約のためにグループウェア等で Web ページを構成する場合、その Web ページ内に他のドメインの Web ページを埋め込み表示することを制限する設定が行われる。そのため WebAPI による表示をグループウェア内に埋め込む場合は WebAPI のドメインからの表示を許可する設定が必要となる。今回扱うグループウェア Confluence においても RStudio Server が動作するドメイン名の埋め込みを許可するよう設定を行った。

3 実装の確認

前節のシステムを実際に動作させ、その表示をグループウェア Confluence で確認した (図 1)。Confluence で生成されるページは、グループを限定して構成可能であり、特定ユーザのみがアクセスできるページを構成している。図 1 で示す左の 10 個のグラフは RStudio Server 側で 1 つの画像として生成されており、1 つの WebAPI で提供される。右側の凡例は別の WebAPI として提供される。このように必要となる複数のグラフを生成して、複数の WebAPI で提供することが可能であり、1 つのページ内に必要なグラフや表を表示して意思決定のためのダッシュボードとして利用できる。

今回実装には至らなかったが、ライブラリ Plumber では、エンドポイントとして HTML Widget による提供が可能で、R 言語のライブラリ DataTable [5] により図 2 のように動的な表を提供可能となる。DataTable によって表示された表は、特定の項目についてソーティングやフィルタ、そして検索が可能で、意思決定を行うような場面で直接操作してデータを確認できる。しかし、これには問題点もある。ソースファイルを確認すると、その表示に用いられるすべての情報がそのまま格納されている。SSL/TLS プロトコルにより安全な状態を構築できると考えられるが、表示した情報をブラウザが保持することを考えると必ずしも安全であるとは考えられないため、運用には注意が必要である。

今回利用した RStudio Server は学内でも特定のネットワークからのみアクセス可能な環境で動作している。集計処理を行うシステムは、様々な重要な情報資源へのアクセスが許可されているので、十分に安全面を考慮して運用が必要である。一方で、アクセスや認証により安全面を確保することで、意思決定が行われる場所での操作を難しくするという難点を持つことにもなる。



datetime	http://portal.kumamoto-u.ac.jp	https://portal.kumamoto-u.ac.jp	https://casoeki.kumamoto-u.ac.jp
2017-04-14 08:00:00	40	38	
2017-04-14 08:10:00	72	72	
2017-04-14 08:20:00	210	206	
2017-04-14 08:30:00	206	198	
2017-04-14 08:40:00	173	163	
2017-04-14 08:50:00	158	145	
2017-04-14 09:00:00	177	169	
2017-04-14 09:10:00	138	130	
2017-04-14 09:20:00	139	127	
2017-04-14 09:30:00	130	123	
2017-04-14 09:40:00	138	123	
2017-04-14 09:50:00	141	136	

図- 2: R 言語のライブラリ DataTable を用いたデータの表示

4 まとめ

本報告では、グループウェアと WebAPI を利用したダッシュボードの構成について検討を行った。動作するグループウェア Confluence と WebAPI を提供可能な RStudio Server を用いて、特定のユーザのみが意思決定に必要な情報へアクセスできる環境を構築することができた。RStudio Server の機能を考えるとより動的な情報を提供可能であることも分かっているが、ブラウザがそのデータを保持する点など、安全面に注意が必要となる。また、RStudio Server のように集計処理を行うサーバは多くの重要な情報源へのアクセスを許可している可能性があり、その運用に注意が必要となる。本報告では扱わなかったデータベース構築について早々に検討を行い、ログ情報の一元管理とアクセスの利便性向上を実現したいと考えている。今回の検討によって、R 言語のオブジェクトをファイルとして扱うなど、事前の集計処理なども取り入れることで、WebAPI による表示処理が困難無くできると考えている。すべてを動的処理にする場合には、集計処理を行うサーバのコンピュータ資源が潤沢にするなど工夫が必要と考えられる。

参考文献

- [1] 永井孝幸, 杉谷賢一, 河津秀利, 中野裕司: 学認対応認証基盤とユーザ ID 体系移行用 CAS ゲートウェイの構築, 情報処理学会研究報告教育学習支援情報システム (CLE) , vol. 2013-CLE-11, no. 20, pp. 1-10 (2013).
- [2] <https://www.rstudio.com/products/rstudio/> (2018 年 8 月 24 日確認)
- [3] <https://www.rplumber.io/> (2018 年 8 月 24 日確認)
- [4] <https://www.rplumber.io/docs/index.html> (2018 年 8 月 24 日確認)
- [5] <https://rstudio.github.io/DT/> (2018 年 8 月 24 日確認)