

**この資料は日本Mテクノロジー学会員専用です。**

**この資料を学会員以外がコピーしたり、学会員以外に配布することを禁じます。**

**Copy right : M Technology Association – Japan**

日本Mテクノロジー学会事務局

〒260-8677 千葉市中央区亥鼻 1-8-1

千葉大学医学部附属病院企画情報部内 鈴木隆弘

Tel: 043-226-2346

Fax: 043-226-2373

Email: [mta-office@mta.gr.jp](mailto:mta-office@mta.gr.jp)



# MTA2014

## 第41回 日本Mテクノロジー学会大会 講演論文集

2014年8月29日～30日  
志摩地中海村（三重県志摩市）

大会長

プログラム委員長

実行委員長

岐阜大学大学院医学系研究科

長崎大学大学院医歯薬学総合研究科

三重大学大学院工学研究科

紀ノ定 保臣

本多 正幸

川中 普晴

巻頭言

## 第 41 回 日本 M テクノロジー学会大会

メインテーマ

### M Language Returns

M 言語の基礎を正しく理解し、  
ビッグデータ時代のデータ活用に挑戦しよう！

岐阜大学大学院医学系研究科 紀ノ定 保臣



第 41 回日本 M テクノロジー学会大会の大会長を務めさせていただくことを大変光栄に存じます。M 言語は長い歴史と医療分野を中心に数々の実装成果を上げてきた実績があります。その本質は簡易なプログラム記述を可能とする M 言語と balanced tree 構造を基本とするデータベース構造にあります。これらの特徴は健康医療福祉分野のデータ処理に不可欠な基本要素であり、その重要性は今後も変わることはないと考えております。一方、情報通信技術 (ICT) の急速な進展と社会への浸透は、大規模データの発生と蓄積、蓄積されたデータの処理と活用を多様化するとともに、大規模に蓄積された各種データを高速に処理し、社会の望ましい発展に貢献することを求めています。さらに、この新しい社会のニーズに対応するために NoSQL や Hadoop, AWS といった新たなデータの蓄積・処理・活用手段が急速に拡大しています。

M 言語は今の社会に貢献することができる要素を持った技術であり、MTA 会員はそのことを強く意識すべきだと思います。”M に欠けている要素があれば新規に開発すればよい”，”M の良い点を再認識しその特徴をさらに発展させればよい”。そのためには、M 技術のみならず、新興技術も正しく理解し、その成果を M にフィードバックする姿勢が必要です。

本大会が、M をさらに発展させるための議論をする場となれるように、三つの特徴を演出しました。一つ目は大会初日の英語による講演（海外からの招待講演）、二つ目は大会二日目の大会長特別企画（他分野の技術を理解するための「Big Data in Healthcare」）、三つ目は M 言語のチュートリアルとハンズオン・セミナーです。

第 41 回 MTA 大会は、伊勢志摩の地を会場として選びました。会場は、伊勢志摩国立公園のほぼ中心にある英虞湾を望む絶景の地にあります。都会の雑踏を離れ、新鮮な魚介類を堪能しながら会員の交流を深めつつ、自身の研究を見つめなおす大会となれば幸いです。式年遷宮も終わり、いっそう賑わいをみせる伊勢志摩の魅力を楽しんでいただけるような企画も用意する所存です。

最後に、本大会が会員諸氏の研究の一助となれば幸いです。多くの研究者が参加し、活発な議論と明日に向けて新たに挑戦する意欲が生まれることを期待しています。また、会員が積極的に研究成果を海外に発信し、世界で活躍する姿を見せて頂けることを楽しみにしています。

# 第41回 日本Mテクノロジー学会大会 大会プログラム

## 第1日目 (8月29日)

11:30~	受付開始 (学会会場)
12:00~12:50	日本Mテクノロジー学会 評議委員会
12:50~13:00	開会挨拶 (メイン会場) MTA2014大会長 紀ノ定 保臣 (岐阜大学大学院医学系研究科)
13:00~14:20	一般講演 (1) 司会: 鈴木 隆弘 (千葉大学医学部附属病院企画情報部) 放射線治療病歴データベースシステムにおけるデータ生成状況と新規データスキーマの検討 向井 まさみ (放射線医学総合研究所)  分散と統合を繰り返すシステムの事例 山本 和子 (日本医療情報学会 医療情報技師育成部会)  医療情報技師更新制度における資格更新手続きを行う技師とその内容について 伊勢田 司 (日本医療情報学会 医療情報技師育成部会)  院内患者位置情報表示システムと活用 土屋 喬義 (土屋小児病院)
14:20~14:30	休憩
14:30~15:30	特別講演 司会: 紀ノ定 保臣 (岐阜大学大学院医学系研究科) M: past, present and future (M: 現在、過去、未来) Atsushi Inoue 先生 (Eastern Washington University, USA)
15:30~15:40	休憩
15:40~17:00	チュートリアル1 (多目的ルーム) M言語初心者講座 プログラムを通してMを知ろう! 土井 俊祐 先生 (千葉大学医学部附属病院地域医療連携部)  チュートリアル2 (メイン会場) インターシステムズ最新動向 佐藤 比呂志 先生 (インターシステムズジャパン株式会社)
17:00~17:30	休憩
17:30~	懇親会 (会場内のレストラン)

## 第2日目 (8月30日)

9:00~10:00	<b>一般講演 (2) 司会: 山本 皓二 (鈴鹿医療科学大学)</b> 臨床研究のための多施設間診療データ連携の試み 竹村 匡正 (兵庫県立大学)  医事データのOpenEHR・アーキタイプ適用に関する検討 櫻井 理紗 (兵庫県立大学)  Instant Message Service基盤としてのHealthshareの活用事例 桑田 成規 (国立循環器病センター病院)
10:00~10:10	休憩
10:10~11:40	<b>大会長特別企画 司会: 紀ノ定 保臣 (岐阜大学大学院医学系研究科)</b> Big Data in Healthcare 第一部 (10:10-10:50) Visual Mining StudioとBig Data Module を用いた医薬品副作用データベースの分析 五十嵐 健太 (株式会社NTTデータ数理システム)  第二部 (10:50-11:10) Apache Hadoopを使った構築方法とその事例紹介 山崎 重雄 (ちらし屋ドットコム)  第三部 (11:10-11:40) Big Data in Healthcare 紀ノ定保臣 (岐阜大学大学院医学系研究科)
11:40~11:50	休憩
11:50~12:30	<b>一般講演 (3) 司会: 嶋 芳成 (日本ダイナシステム株式会社)</b> 「とねっと」 (埼玉利根保健医療圏地域医療ネットワークシステム) へのデータ転送 伊藤 章 (土屋小児病院)  象徴的記号のシンボル・プロセッサと階層型データ 高橋 亘 (基礎コミュニケーション科学研究所)
12:30~13:30	昼休憩 (日本Mテクノロジー学会 総会)
13:30~16:30	<b>チュートリアル</b> GT.M ハンズオンセミナー ~ GT.M Webアプリケーション開発入門 ~ 澤田 潔 先生 (日本ダイナシステム株式会社, MTA-OSSJP分科会)
16:30	閉会

特別講演  
&  
大会長特別企画

# M: Past, Present and Future

Atsushi Inoue

Eastern Washington University, Spokane, WA 99202, USA

E-mail: inoueatshij@gmail.com

Evolution and features of the M language are outlined in both aspects of information technologies and applications. Specifically, it is focused on those that are to be reflected with EHR technologies and the recent healthcare reform, i.e. Obama Care and HITECH Act, in the US. Selected emerging issues are introduced and discussed while predicting the future of M, especially for matters of Big Data.

Language M (or **MUMPS** – Multi-User and Multi-Programming System) was initially emerged in 1967 at Massachusetts General Hospital (MGH) as a part of their hospital information system project. This is a **general-purpose programming language for developing database applications**. Its main features include a hierarchical key-valued database engine and a high-throughput yet reliable (i.e. ACID – Atomicity, Consistency, Isolation and Durability) transaction processing capacity.

By early 1970s, M was in widespread use mostly among medical communities and was often locally modified for their own needs. This became the **ANSI standard** in 1977 (X11.1-1977). In 1980s, M has been ported to **multiple platforms** and its important implementations came out by various vendors. Examples include DSM, ISM, **GT.M**, DTM, and MGM. In 1990s and 2000s, such technologies adopted **open technologies**, e.g. Open M by **InterSystems** and GT.M under the (A)**GPL** license. De-

rivatives developed in this period include MIIS and MAGIC by MEDITECH, ANSI Standard MUMPS led by Ray Newman, and a notable open source implementation, MUMPS/II, by Kevin O’Kane at University of Northern Iowa.

In 1997, InterSystems released **Caché** as an object-oriented database system with rich, enhanced interfaces – API, O(J)DBC, and embedded SQL. This is based mainly on ISM with influences and adoption of many other implementations, that are mostly acquired as a result of buying out other vendors. Cache has been deployed by many MUMPS developers who write applications in various operating systems. By 2000, InterSystems became a dominant **middleware** vendor in MUMPS market. By mid 2000s, most use of M became in the form of GT.M maintained by Fidelity Information Services or InterSystems Caché. Currently, many major proprietary products and institutes use M, including but not necessarily limited to VistA, Epic, GM Healthcare (IDX + Centricity), EMIS, MEDITECH, Partners Healthcare, and Sunquest Information Systems. M has also been used in financial industries such as AmeriTrade among others.

In 2009, HITECH Act (aka Obama Care) is enacted and sets **meaningful use** of interoperable **EHR** in health care as a critical national goal. Goal of progress on EHR adoption has been set by three

stages and incentivized:

Stage 1. Certified EHR by 2009.

Stage 2. Information exchange by 2012.

Stage 3. **Quality** care by 2015.

This also serves the US as a major economic development driving force as a result of a huge demand on health informatics technologist positions that are necessary to support the entire lifecycle of those adopted EHR systems. Colleges and universities respond to this demand by offering Health Informatics programs at various levels.

As EHR systems have been adopted and interoperated properly and smoothly, health care institutes need to define quality set forth as stage 3 of the meaningful use. As of now, the trend in quality care is twofold: **workflow management** and **population health**. The workflow management focuses on optimality of health care transactions and their flows at health care institutes in terms of care outcomes, duration and costs. In the US, interoperability among different stakeholders such as clinics, hospitals, pharmacies, and insurance agencies is a big challenge due to their diversity and complexity (i.e. a “rich” set of choices). The population health aims to improve health of the entire human population by diligently analyzing broad range of factors throughout their lifetimes to find out which potentially have a measurable impact. This represents a change in focus from individual levels, characteristics of the current mainstream medicine.

For a smooth and effective support of such quality care, Information Technologies need to serve as **platforms** that provide all necessary functionalities and services – distributed processing architecture such as hadoop (MapReduce), web, mobile device supports such as smartphones and tablets, visualiza-

tion, GUI, information security, health information exchange, analytics, diagnostic and patient advising assistance, medical decision support, radiological image analysis, e-prescription, message delivery, consent management, etc. Responding to this, many vendors present various solutions and platforms (e.g. **HealthShare** by InterSystems) that integrate M with many other emerging technologies.

On such platforms, various types and massive amount of data are to be exchanged, stored, searched and analyzed. While efficient and effective methods and technologies of processing each and every type of data up to a certain quantity have been developed, processing such **big data** as a collection of EHR systems from several health institutes in an integrative and comprehensive manner takes us to a new frontier. Seeking breakthrough is urgent, and ‘**thinking outside the box**’ is very likely important.

#### Short Bio.



Atsushi Inoue started his IT career in 1990 as a research scientist of Engineering Information Systems at Hitachi Ltd., Tokyo, Japan. In 1996, he moved to the US for further study and entrepreneurship on Soft Computing. In 1999, he obtained his Ph.D. in Computer Science and Engineering from the University of Cincinnati. He is currently a Professor of Information System and Decision Science at Eastern Washington University.

# Visual Mining Studio と Big Data Module を用いた 医薬品副作用データベースの分析

Analysis of Japanese Adverse Drug Event Report database  
with Visual Mining Studio and Big Data Module

五十嵐 健太<sup>1)</sup>

Kenta Igarashi<sup>1)</sup>

株式会社 NTT データ数理システム<sup>1)</sup>

NTT Data Mathematical Systems Inc.<sup>1)</sup>

キーワード: ビッグデータ, データマイニング, 分析ツール

## 1. はじめに

昨今ビッグデータというキーワードがメディアで盛んに取り上げられている。様々な分野で、従来は目を向けられなかったデータを活用しようという意識が高まっており、データの蓄積、管理を行うための基盤が整備されている。また、オープンデータの推進も行われており、行政機関を中心として、様々なデータがインターネットを通じて容易に手に入るようになってきた。

その一方で、これらのデータの分析への活用はあまり進んでいないように思う。本格的にデータ分析を行うためには、データマイニング、機械学習、統計解析等の技術を使用した複雑な計算が必要となり、敷居が高い部分があるのかもしれない。しかし、多くの場合これらの計算はデータ分析ツールに任せることができ、分析者は結果の解釈と現象の理解に注力することができる。ビッグデータを十分に活用するためには、分析ツールの活用が不可欠である。

本発表では、NTT データ数理システムの分析ツール「Visual Mining Studio」および「Big Data Module」を使用した医薬品副作用データベースの分析を通じて、ビッグデータの分析の様子を紹介する。

## 2. Visual Mining Studio と Big Data Module

Visual Mining Studio は分析の流れをアイコンと矢印によるビジュアルプログラミングで表現し、データ分析を行うツールである。数 100 万行を超えるようなデータに対しても、マウス操作だけで簡単に分析を行うことができる。データマイニングでよく使われる分析のアルゴリズムはすでにアイコンとして提供されており、複雑な分析でもすぐに実行することが可能である。

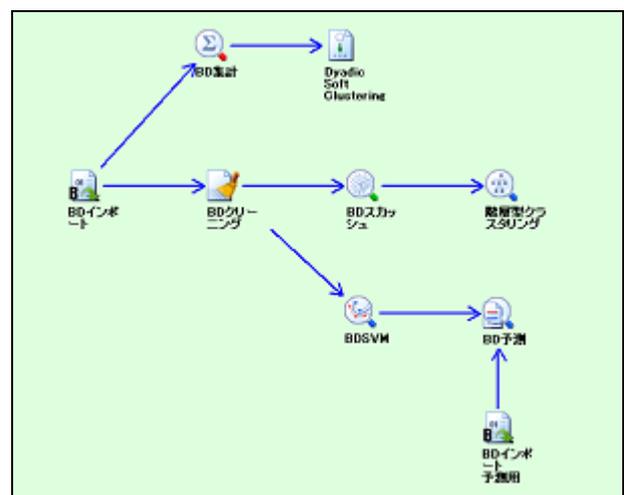


図 1 Visual Mining Studio の分析イメージ

Big Data Module は Visual Mining Studio と同一のプラットフォームの上で動作する分析ツールである。大規模なデータに対してより高速に処理を行うことを意識し、並列処理やオンラインアルゴリズムを使用した高速な分析機能をアイコンとして提供する。1台の市販のPCでも数100GBのデータに対して高度な分析を行うことが可能である。

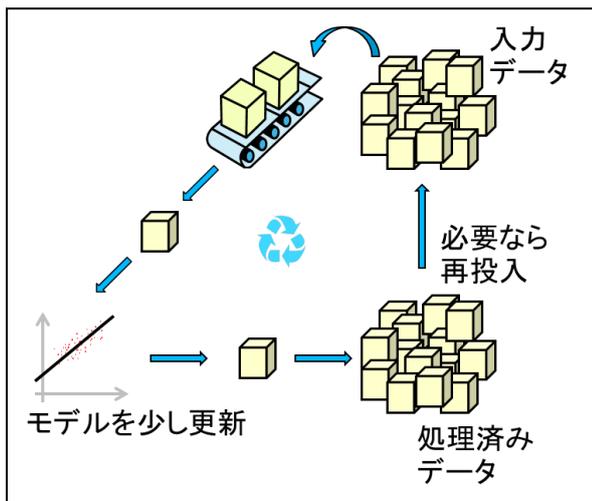


図2 オンラインアルゴリズムのイメージ

### 3. 分析とデータサイズ

ビッグデータの分析を行う仕組みとして真っ先にイメージされるのは、Hadoopをはじめとした大規模分散処理システムではないだろうか。従来のデータベース、分析ツールでは処理することが困難なテラバイト、ペタバイトオーダーのデータにたいして、スケーラブルかつ高速な処理を提供するのがこれらのシステムである。集計、ワードカウントなどの処理には絶大な効果を発揮する。

一方で、データマイニング、機械学習で使用されるようなアルゴリズムの中には分散処理があまり有効でないものも多い。データの操作性なども考慮し、大規模な分散処理システムではなく、手元のPCの上で分析を行う場合も多い。実際、現状ではセンサー、通信などのログデータ以外であれば、十分に情報量が豊富なものでも1台のPC

で処理できるレベルのデータサイズであることが多い。

大規模分散処理システムと手元のPCでの処理には一長一短あり、それぞれ補い合って分析を行うべきではないかと考えている。特に、今後蓄積されるデータおよびその処理時間が増大し続けることを考えると、それぞれの強みを生かし合っていないと現実的な時間で意味のある分析を行うことが難しくなるだろう。

ここで紹介する Visual Mining Studio と Big Data Module は基本的には1台のPCで操作を行うためのツールであるが、Hadoopとの関係するためのGUIを備えている。テラバイトを超えるようなデータの分析が必要となる場合は、Hadoopとそれぞれの長所を組み合わせることで、図3のような分析の幅広い分野をカバーすることが可能である。

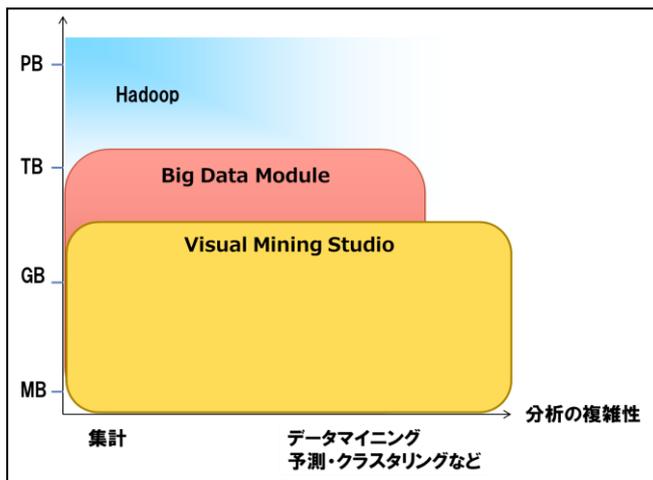


図3 分析ツールとカバー範囲

### 参考文献

1. Bottou, L. and LeCun, Y. (2004). Large scale online learning, *Adv. Neural Inf. Process. Syst.*, **16**, 217-224.

# Apache Hadoop を使った構築方法とその事例紹介

## ビッグデータをカジュアルに着こなす Hadoop の身につけ方

山崎重雄<sup>1)</sup>

YAMAZAKI Shigeo<sup>1)</sup>

株式会社 ちらし屋ドットコム

Chirashiya.com inc.,

キーワード: Hadoop ビッグデータ 分散処理 EMR

### 1. はじめに

Hadoop は、ビッグデータ解析のための分散処理のフレームワークです。Apache Hadoop (以下、Hadoop) 登場以前は、神懸かったエンジニアしか扱えなかった分散処理が、Hadoop の登場により“普通”のエンジニアであっても扱える技術となってきました。つまり今まで持っていた処理できずに捨てていた大量のデータをきちんと活用できるようになり、ビジネスに役立つパターンを発見することが出来るようになったのです。

Hadoop の導入により、データの解析の障壁となっていた処理の本質でない部分はフレームワークである Hadoop が面倒を見てくれるため、ユーザーの本質的な処理である Mapper と Reducer といった 2 つのスク립トを作成する事に集中することが出来ます。また、データ分析に欠かせないデータマイニングについても多くの時間を費やすことが出来、精度の高い解析結果が得られることが出来るようになりました。Hadoop の特徴として、この分散処理フレームワーク以外にサーバーを増やすしてのスケールが可能、様々な言語で処理が書ける等が挙げられ、データを処理する Hadoop のエコシステムにより Hadoop の処理パワーを活用する様々なプロダクトが生まれてきています。また、これら Hadoop エコシステムは、中核システムを含めクラウド環境での提供がされ、ハード

ウェア調達のコスト面においてもスケールを意識したシステムを構築することが出来ます。

図1は、the Digital Universe (地球上で生み出されるデータの数) です。世界のデジタル・ユニバースの分析は未だ1%未満と言われ、オポチュニティあふれたデジタル・ユニバースの発見、分析は豊富なデータとモノとともに大きな価値を生み出します。ビッグデータをカジュアルに着こなす Hadoop を身につけ、ビッグデータを活用するための近道を紹介しましょう。

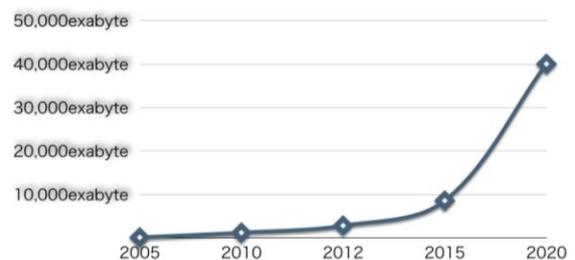


図 1 : the Digital Universe

地球上のすべての海岸にある砂の粒を合計すると

700,500,000,000,000,000 (7 垓 50 京、1 垓=10 の 20 乗) 個。40 ゼットバイトはその 57 倍に相当

### 2. Hadoop2.2 インストール・環境設定

これからHadoopを始める方は、2系(2.2、2.5)をお奨めします。以下にインストールの手順を示します。

Hadoop の動作モードには「スタンドアロンモード」「疑似分散モード」「完全分散モード」の3つのモードがあります。スタンドアロンモードは主に、テストや開発に用いるモードです。疑似分散モードは、デーモン(Hadoop サーバーソフト)を一台のローカルマシンで動かすモードです。完全分散モードは、Hadoop の本来の能力を発揮できるモードです。

- Linux マシンの準備 :

インストールを行うには一台の Linux マシンが必要です。OS のディストリビューションは、「Fedora」や「CentOS」や「Ubuntu」等を使用します。

- Sun JDK のインストール

Hadoop は、Sun JDK (Java SE Development Kit)の利用を推奨しています。Sun の JavaSE のサイトにアクセスし、JDK の最新版(1.6 以上)をダウンロードし、Linux 環境にインストールしてください。

- Linux の環境設定

Hadoop を使用するにあたって、Hadoop を使用するための専用ユーザーグループ、ユーザーを設定し、先の JDK の位置等の環境変数の設定を行います。同時にデーモン(Hadoop サーバープログラム)を動作させるための設定を行います。その後 SSH を使って対象マシンにログインし、Hadoop のデーモンを立ち上げるために SSH のパスフレーズ無しでログインできるように設定しておきます。

- Hadoop のインストール

Apache Hadoop プロジェクトの Web サイトからダウンロードし、所定のディレクトリーにインストールします。また、.bashrc ファイルを編集し所定の環境変数を設定します。

- Hadoop 設定ファイルの編集

Hadoop に関する各種xmlファイルを設定します。(core-site.xml,hdfs-site.xml,mapred-site.xml)

### 3. 疑似分散モードによる円周率計算

Hadoop 内のサンプルプログラムを使い処理時間の計測を行います。

```
hadoop jar hadoop-mapreduce-examples-2.2.0.jar pi 10 1000
```

動作環境:intel i7(1.8G) + Ubuntu14 + 仮想マシン (CPU×1、1M) + Hadoop2.2

上記の組み合わせにおいて、疑似分散モードは2分47秒ほどかかります。同一条件でのスタンドアロンモードにおいては、1分で終了します。

表1は、その動作結果です。

<pre> ※スタンドアロンモード [hduser@yamac-vm:/usr/local/hadoop]\$ hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-exam ples-2.2.0.jar pi 10 1000 Number of Maps = 10 Samples per Map = 1000  途中省略-----  File Output Format Counters   Bytes Written=109 Job Finished in <b>61.348</b> seconds Estimated value of Pi is 3.14080000000000000000 </pre>
<pre> ※疑似分散モード [hduser@yamac-vm:/usr/local/hadoop]\$ hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-exam ples-2.2.0.jar pi 10 1000 Number of Maps = 10 Samples per Map = 1000  途中省略-----  File Output Format Counters   Bytes Written=97 Job Finished in <b>167.88</b> seconds Estimated value of Pi is 3.14080000000000000000 </pre>

表1:動作結果

疑似分散モードの方が処理時間が長いという結果となりました。

この結果ですが、スタンドアロンモードは通常複数台で行う分散処理を1台で行っている、使用するファイルシステム(HDFS とローカルファイルシステム)の違いによるものです。

### 4. Amazon ERM を使用した完全分散モード

Amazon ERM(ElasticMapReduce)は、Hadoop というオープンソースフレームワークを使用したデータの分散処理を Amazon EC2 インスタンスの拡大縮小可能なクラスターで行うことができます。

これによりスケールを自由に設定することが出来ます。

図2は、分散処理の流れを示します。

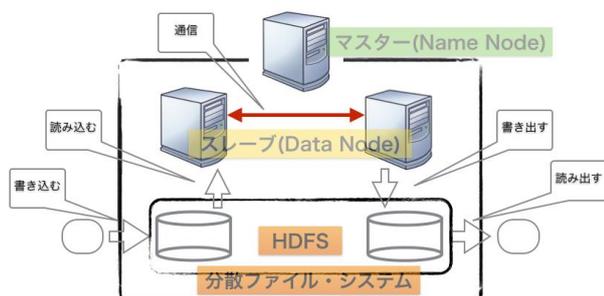


図2:分散処理の流れ

対象となるデータを HDFS に書き込み、Hadoop を起動します。マスターは、HDFS から各スレーブに書き込みデータのマッピングを行います。その後スレーブ間の通信を行いながら HDFS に結果を書き出します。

図3は、AmazonEMR の分散処理の流れを示します。

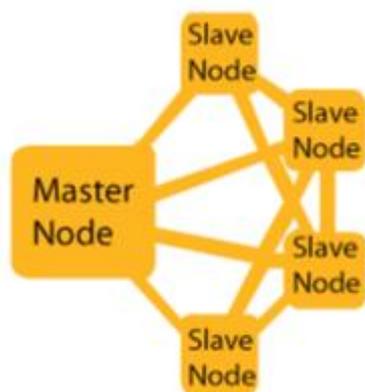
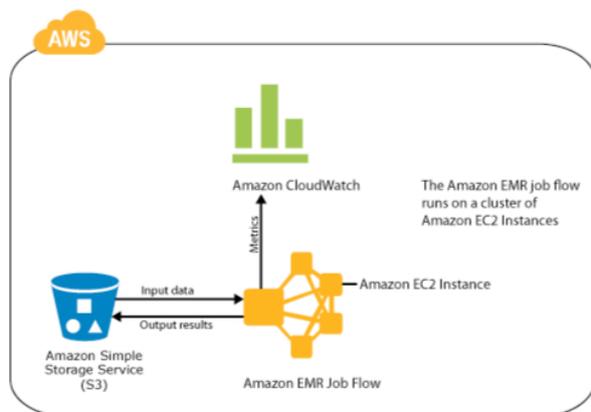


図3:AmazonEMR 分散処理の流れ  
(下段は、Hadoop Cluster のイメージ)

EMR の場合でも原理は同じです HDFS は S3 がその役目を行います。各ノードは EC2 が受け持ちます。

次に文字数をカウントするプログラムを作成し、疑似分散モードと EMR の完全分散モードで処理時間の比較を行います。図4は今回用いた python で書いたスクリプトです。図2の読み込む処理 mapper と書き出す処理 reducer の処理に分けて処理を行います。mapper.py の処理は、与えられたデータより単語を <key><value> に分割して出力し reducer.py に受渡し、並び替えと文字カウントを行い出力します。

```

mapper.py
#!/usr/bin/env python
import sys

for line in sys.stdin:
    line = line.strip()
    words = line.split()
    for word in words:
        print '%s\t%s' % (word, 1)

reducer.py
#!/usr/bin/env python

from operator import itemgetter
import sys

word2count = {}

for line in sys.stdin:
    line = line.strip()

    word, count = line.split('\t', 1)
    try:
        count = int(count)
        word2count[word] = word2count.get(word, 0) + count
    except ValueError:
        pass

sorted_word2count = sorted(word2count.items(), key=itemgetter(0))

for word, count in sorted_word2count:
    print '%s\t%s' % (word, count)

```

図4:Python スクリプト

データは約 90MB の CIA Factbook (2000-2010) を使用します。このデータに対して HDFS 及び S3 に書き込み Hadoop での解析処理を行います。

図5は、疑似分散モードの処理結果です。

動作環境:intel i7(1.8G) + Ubuntu14 + 仮想マシン (CPU×1、1M) + Hadoop2.2

処理時間は、3分ほどです。



- Cluster
- About
- Nodes
- Applications
  - NEW
  - NEW\_SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - REMOVING
  - FINISHING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler

User: hduser  
 Name: streamjob57236970126  
 Application Type: MAPREDUCE  
 State: FINISHED  
 FinalStatus: SUCCEEDED  
 Started: 21-8-2014 18:44:41  
 Elapsed: 3mins, 7sec  
 Tracking URL: History  
 Diagnostics:

ApplicationMaster		Start Time
Attempt Number		
1		21-8-2014 18:44:41

Name	Status	Start time (UTC+9)	Elapsed time
Streaming program	Completed	2014-08-20 22:13	3 minutes
Setup hadoop debugging	Completed	2014-08-20 22:11	28 seconds
Setup hive	Completed	2014-08-20 22:12	24 seconds
Setup pig	Completed	2014-08-20 22:12	18 seconds

Name	Status	Start time (UTC+9)	Elapsed time
Streaming program	Completed	2014-08-21 19:32	2 minutes
Setup pig	Completed	2014-08-21 19:32	16 seconds
Setup hive	Completed	2014-08-21 19:31	23 seconds
Setup hadoop debugging	Completed	2014-08-21 19:30	33 seconds

図5:疑似分散モード処理結果

図6は、EMR による処理結果です。  
 動作環境:AWS クラウド m1.medium 1 台を使用  
 処理時間は、11 分程です。(プログラムの処理時間、3 分ほどです)  
 参考:m1.medium は Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz 相当と思われる。

Cluster: My cluster Terminated Steps completed

Master public DNS: ec2-54-64-21-117.ap-northeast-1.compute.amazonaws.com  
 Tags: --

Summary	Configuration Details
ID: j-WC42DRLD0QDP Creation date: 2014-08-20 22:06 (UTC+9) End date: 2014-08-20 22:18 (UTC+9) Elapsed time: 11 minutes Auto-terminate: Yes Termination protection: On	AMI version: 3.1.1 Hadoop Amazon 2.4.0 distribution: Applications: Hive 0.11.0.2, Pig 0.12.0 Log URI: s3://chirashiya-emr/logs/cia/

Security/Network	Hardware
Availability zone: ap-northeast-1c Subnet ID: subnet-c41e90ac Key name: Chirashiya.com EC2 instance profile: EMR role: -- Visible to all users: None Change	Master: Terminated 1 m1.medium Core: Terminated 1 m1.medium Task: --

図6:EMR による処理結果

結果は、EMR の方が処理時間が長いという結果となりました。EMR の処理時間の中には、サーバーの立ち上げ、ターミネイトの時間が含まれています。処理の時間だけを見れば同じ結果となります。  
 また参考に、EMR を使いノードのインスタンスを複数台(4 台)立ち上げた場合の結果を記載します。  
 図7はインスタンス 1 台(上段)と 4 台(下段)の処理に懸かった時間です

図7:EMR での処理時間

(上:インスタンス 1 台、下:インスタンス4台)

データ容量(90MB)が小さい場合においては、さほど変わらない結果となります。

### 5. 考察

一台で行う疑似分散モードと複数台のノードを使う EMR を使った完全分散モードにおいて、データの容量が少ないという条件でしたが、処理時間に関してはさほど違いが出ない結果となりました。今回の事例において Hadoop のパフォーマンスを十分に引き出すには至りませんでした、Hadoop は大容量(ビッグデータの)解析(重い処理)を複数台のマシンに分散させて行うことにより処理時間を短縮することが出来ます。なお、日常使っている得意な言語により解析処理を記述出来ることも大きな特徴です。

### 6. 最後に

Hadoop を使ったビッグデータの解析は、分析手法(テキストマイニング)とともに有用な情報を見つけ出すための道具として不可欠な技術になりつつあるといえます。

### 参考文献

1. Apache Hadoop <http://hadoop.apache.org/docs/r2.5.0/>  
 インストール <http://metasearch.sourceforge.jp/wiki/index.php>
2. Amazon EMR <http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/emr-what-is-emr.htm>

# ヘルスケア分野におけるビッグデータ処理

大規模医薬品有害事象データベースを活用したビッグデータ処理：抗うつ薬服用患者から見える自殺企図と症候との関係検出

Big Data in Healthcare

## Identification of suicide-related events through network analysis of adverse event reports

紀ノ定保臣<sup>1)</sup>, 一宮尚志<sup>1)</sup>, Alwis NAZIR<sup>2)</sup>

Yasutomi Kinoshita<sup>1)</sup> Takashi Ichinomiya<sup>1)</sup> Auther<sup>2)</sup>

岐阜大学大学院医学系研究科<sup>1)</sup>, 岐阜大学大学院連合創薬医療情報研究科<sup>2)</sup>

Gifu University Graduate School of Medicine<sup>1)</sup>

Gifu University United Graduate School of Drug Discovery and Medical Information Sciences<sup>2)</sup>

キーワード： Data Mining, Big Data, SSRI, Suicide, FAERS

### 1. Introduction

Suicide is a critical symptom of depression. To reduce suicide deaths, it is essential to estimate accurately the suicide risk of patients and to give proper treatment. Researchers have extensively investigated suicide risk factors, such as acute mood episodes, personal history, and family history. However, more reliable medical signs that indicate suicidal risk are required.

The aim of this study was to identify the symptoms strongly related to suicide. To this end, we analyzed adverse event reports on selective serotonin reuptake inhibitors (SSRIs). SSRIs are the most widely used antidepressant worldwide. They are believed to decrease the risk of suicide; however, even if depressive patients take SSRIs properly, their suicide risk is still greater than that of individuals without depression. Among the reported adverse events of SSRIs are suicidal ideation, suicidal attempt, suicidal behavior, and complete suicide. These adverse events related to suicide provide useful information toward estimating suicide risk.

### 2. Materials and Methods

In this study, we attempted to compile a list of the adverse events that are strongly related to suicide. For this purpose, we analyzed the reports in the US Food and Drug Association (FDA) Adverse Event Reporting System (FAERS). FAERS is the spontaneous reporting system of adverse events and is widely used in pharmacovigilance analysis. To analyze FAERS data, we employed the method of community detection in networks.

We obtained the original dataset of FAERS containing the reports received by the FDA from 1997 to the second quarter of 2012. From this dataset, we extracted reports of adverse events in which SSRI involvement was suspected. Each report in FAERS is composed of a unique identification number, a list of drugs that the patient took, and a list of patient reactions. The names of reactions are coded by the preferred term (PT) defined by the Medical Dictionary for Regulatory Activities (MedDRA®). Though the terms defined by MedDRA® depend on the version employed, we neglected such differences in our

analysis: since most PTs did not change between 1997 and 2012, we believed that these differences would not lead to severe errors in our study. However, the names of drugs are not standardized in FAERS: drugs may appear under their generic or proprietary names. For the standardization of drug names, we used American Hospital Formulary Service (AHFS) “Drug Information 2010” and DrugBank. DrugBank was employed to link the generic and proprietary names of each drug; AHFS drug information was used to combine the generic drug name with its function, such as “SSRI.”

### 3. Results

In the present study, we constructed the network of symptoms reported as adverse events in FAERS and extracted the symptoms that are strongly correlated with suicidal symptoms. Compared with other data mining methods, such as association analysis and Bayesian analysis, network analysis presents several advantages and disadvantages. The major advantage with network analysis is that it can analyze indirect connections. Because it addresses the whole network structure, it can capture hidden relations that would fail to be recognized by association analysis or Bayesian analysis, which deal with the correlation among a restricted number of objects. However, the method of network analysis is still under development. There is no established method for analyzing a network, and many new methods are proposed every year. With FAERS data, several studies have applied association analysis or Bayesian analysis, but none have been based on network analysis.

In the present study, we constructed a symptom network of adverse events. We conducted modularity-based community detection to extract the list of suicide-related adverse events that were believed to have a strong connection with suicidal

events: suicidal behavior, suicidal ideation, suicidal attempts, and complete suicide. We investigated the relationship between suicide risk and those events.

The network of adverse events of SSRIs had 3795 nodes. Using the community detection method, we obtained lists of suicide-related adverse events. As noted above, the algorithm for community detection is stochastic, and we obtained a different list for each run of the analysis. A typical list includes 141 events. We conducted network analysis 20 times, and the number of items on the list varied from 141 to 290 events. The lists included both well-known risk factors, such as social problems and stress, and other adverse events, such as cardiorespiratory arrest and gastric ulcer. We would show you the exact results in our presentation.

### 4. Discussion and Conclusion

The method adopted in the present study will be helpful in analyzing other diseases or adverse events. Compared with other analysis techniques, such as association analysis and Bayesian analysis, an advantage of our method lies in its ability to capture the indirect connection between symptoms.

In this study, we constructed lists of suicide-related adverse events from the FAERS using network analysis. We found that the more listed adverse events a person possessed, the greater the risk of suicide. This result suggests that such lists allow a qualitative estimation of the risk of suicide. Though our method presents problems, such as the high level of false-positive results and excessively long symptom lists, such drawbacks may be countered by refining the modularity detection algorithm, modifying the means of network construction, and applying different signal detection methods.

#### [Reference]

1. A. Nazir, T. Ichinomiya, N. Miyamura, Y. Sekiya, Y. Kinoshita. “Identification of suicide-related events through network analysis of adverse event reports” *Drug Safety*, 2014, 37(7)

チュートリアル  
&  
ハンズオンセミナー

# 一般講演

# 放射線治療病歴データベースシステムにおける データ生成状況と新規データスキーマの検討

## The Situation of the data registration for ten-years and Examination for Next Generation Schema on Radiation Oncology Database System

向井まさみ<sup>1)</sup>, 横岡由姫<sup>1)</sup>, 安藤裕<sup>2)</sup>, 奥田保男<sup>1)</sup>, 辻比呂志<sup>3)</sup>

MUKAI Masami<sup>1)</sup> YOKOOKA Yuki<sup>1)</sup> ANDO Yutaka<sup>2)</sup>

OKUDA Yasuo<sup>2)</sup> TSUJI Hiroshi<sup>3)</sup>

放医研・重粒子医科学センター・医療情報室<sup>1)</sup>,

放医研・重粒子医科学センター病院<sup>2)</sup>, 放医研・重粒子医科学センター<sup>3)</sup>

National Institute of Radiological Sciences, Medical Informatics Section, CHIBA<sup>1)</sup>

National Institute of Radiological Sciences,

Research Center for Charged Particle Therapy Hospital<sup>2)</sup>

National Institute of Radiological Sciences,

Research Center for Charged Particle Therapy<sup>2)</sup>

キーワード:放射線治療、データベース、利用状況、システム更新

### 1. はじめに

当院は、ベッド数 100 床、外来患者数は 70～100 人/日の放射線科単科の病院である。診療は放射線治療に特化しており、1961 年に X 線等による放射線治療を開始し、1994 年より炭素イオン線を用いた悪性腫瘍に対する放射線治療を開始し、2014 年 8 月までに約 9,300 例の治療を行っている。また当院は放射線に関する研究機関でもあり、これらの放射線治療に関する疾患情報、治療内容、予後の情報は臨床研究のための重要な情報である。放射線治療に関する情報を長期に渡り一貫して管理し症例報告や治療実績件数抽出などを簡易に行うことを目的に、1999 年に放射線治療病歴データベースシステム(AMIDAS: Advanced Medical Information Database System)を構築し 2000 年より運用を開始した。2009 年に本システムのデータベースに Cache<sup>®</sup>を採用し運用を継続している[1]。

### 2. 目的

2015 年度稼働を目標にシステム更新を行うことになった。過去 10 年間の運用を踏まえ、移行すべき機能の整理および必要なデータ項目の整理を行い、新システムの構成を検討することとした。本研究では、データ構造および項目(以降、第 II 期 AMIDAS 用 Schema)の整備・検討結果について報告する。

### 3. 方法

第 II 期 AMIDAS 用 Schema の検討にあたり次の 5 ステップで検討を行った。Step1)2014 年 10 月までの全てのデータ(テーブル・カラム)について、テストデータ以外の実データが格納されているものを全て洗い出す。Step2)洗い出したデータを基に利用していないテーブルおよびカラム、二重に登録されているカラムを削除した Schema 案を



# 分散と統合を繰り返すシステムの事例

An example of distributed and integrated data base systems

山本和子

Kazuko Yamamoto

一般社団法人日本医療情報学会医療情報技師育成部会

Healthcare Information Technologist Fostering Task Force of JAMI

キーワード: 分散データベース、検定試験、医療情報技師

## 1. はじめに

医療情報の IT 化の進歩はめざましく、それを支える人材を育成するために、2003 年に医療情報技師能力検定試験制度ができて早や 10 年が過ぎている。受検者数は一旦低下したが、その後増加し 2014 年には 5000 人を超えた。医療情報技師の合格者数は図 1 に示したように、2013 年には延べ 13,934 人、上級医療情報技師は 282 人に達している。そろそろビックデータの入り口にさしかかった今、このシステム設計でよかったかどうか、メリット、デメリットを考察し、専門家のご意見を頂戴したい。

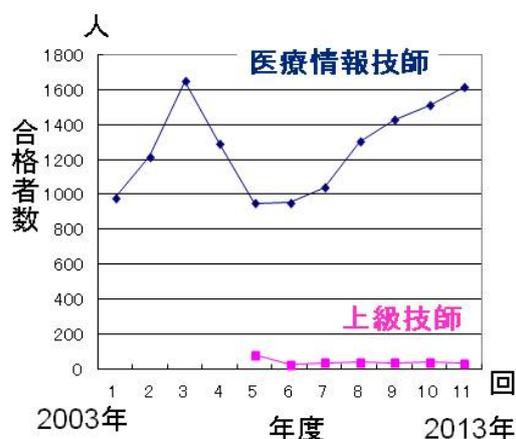


図 1. 医療情報技師合格者数の推移

## 2. システムの概要

当初 (2003 年) は講習会と検定試験申込者のデータ入力为主で Caché の CSP を用いて PC へデータ登録するシステムを開発した<sup>1)</sup>。当時はワークステーションと VB 端末とのサーバ・クライアントシステムが流行していた時代である。Caché の CSP を用いたのはサーバとクライアント間の複雑なデータ通信を必要としなかったからで、以後すべてのシステムはこれで稼働している。

インターネットの発達に伴い、各種申し込みはインターネットからできるように改良し、入力作業が削減された。セキュリティを守るために外部システムと内部システムに物理的に分離している。その後、業務内容が増加して表 1 のようになる。表 1 の中で医療情報技師関連の各種申込クラスは 12 種、試験の成績関連クラスが 21 種、試験問題作成クラスが 6 種、会計・給与関連が 10 種、検定合格者クラスが 4 種、合計 53 種。その他テーブル関連のクラスがあり、1 年単位でクラス名を変更し分散型 (スモールデータベース) のシステムにしていた。クラスは毎年増加し、10 年経過後は 530 種になる。

表 1. 各種申込業務一覧

分類	種類名称	分類	種類名称
F	医療情報技師能力検定	D	講習会DVD(販売完了)
S	上級技師一次試験	V	画像情報専門技師検定試験
U	上級技師二次試験	C	画像情報専門技師認定料
G	基礎知識検定試験	M	上級技師認定料
K	講習会	A	医療情報技師更新
L	生涯研修セミナー	B	上級医療情報技師更新
Q	EL:講習会	H	画像情報専門技師更新
E	EL:生涯研修セミナー	LG	ロゴのダウンロード
R	EL:基礎知識(現在中止)	I	上級技師試験問題販売
T	EL:過去問(現在中止)		基礎知識検定試験会場申込

注) EL : E-Learning

### 3. 分散型を採用した理由と問題点

当時、分散型を採用した理由は<sup>1)</sup>、1) CachéのクラスはRDBの形式になっている。SQLを使用できる。2) 表形式(SQL)でデータが見える。簡単にcsvファイルに取り出せる。3) プログラム開発用のエディタ(スタジオ)が使える。4) インデックスファイルを作成できる。5) 階層構造のクラスを作れる。6) OSに関係しない。7) データ量が少ない方が処理速度が速い。等である。結果として派生した問題点は、M言語の時のようにクラス名を間接法で表現する方法がわからなかったので、クラス数の増加に伴い、毎年プログラム数が増加し、そのメンテナンス作業の負担が大きくなった。

### 4. 業務の機能増加に対する対策

毎年プログラム変更を無くすことを目的に業務を整理した。

#### 4-1. 外部システム

Web上にある各種申込データは一つのデータベース(MySQL)に蓄積し、毎日オフラインで内部データベースに表1の分類名をキーに各クラスに配分している。プログラムは1本である。

#### 4-2. 内部システム

開発時間削減のための改良点は、

- 1) 画面統一。入力・参照画面を統一した。
- 2) プロパティ統一。クラスのプロパティ名を共通にしてコピーだけでクラスを生成できるようにした。
- 3) インディックスは設定しない。必要時にMのグローバルファイルでデータを並び変えている。
- 4) 業務の内容を定型化。プログラムを変更することなく永久的に利用できる方法として業務内容を、掃出型、増殖型、混合型の3種にわけて対応している。

### 5. 業務内容分類

#### 5-1. 掃出型

各種申込データベース、成績関連データベース、会計関連データベースは、1年が過ぎれば日常業務で利用することはないので、1年後は別のクラスに移している(図2)。たまに参照または、統計解析する程度である。しかし問題点としては、1年毎にクラス数が増加していく。いずれ別のPCに移し変える予定であるが、統合してプライベートクラウドにすべきかどうか検討中である。

#### 5-2. 増殖型

医療情報技師合格者のデータベース(DB)は永久的に増え続ける。日々メールアドレスの更新があり、5年毎に技師資格が更新されていく。このDBから技師のMailing List(以下MLと略す)を作成している。技師のMLは同時に500人しか配信できない。従って全年度の合格者を一つのクラスに統合し、それから500人ずつ分割してML配信プログラムに送り込んでいる。負荷がかからないように18時頃に配信を開始し、翌朝9時には配信が終了していなければならない。技師数が増加し2日間に分割して配

信しなければならなくなった。プログラムの更新を無くすために、クラス名を汎用化し、図3のように運用する予定である。

### 5-3. 混合型

問題作成システムがこれに該当する。試験問題はA型（五択問題）、C型（穴埋め問題）、L型（長文問題）に分類し、どのような形の問題でも登録できるようにプログラムを統一した。

システムの概要を図4に示す。サーバ自体がデスクトップ型PCと3個のノートPCの分散型であり、1年たてばデスクトップ型PCとノート型PCの両方に保管されているデータを統合し、出題した問題には識別係数をつけて、プール問題として保管しておかなければならない。試験問題は文字列であり画像も含まれている。文字列の中に改行の制御記号などが入っているとクラスからデータのエクスポートが難しい。PCそのものが別の場所で使用されているため、同時書き込みも不可能で、どのように統合するか検討中である。

## 6. 考察

### 6-1. 掃出型について

掃出型の場合はプログラムの更新が不要になり、開発費（開発時間）が削減できた。一方、分散型にしたためにクラス数が増え続けることが問題点として残っている。

クラスを1年単位のスモールデータベースにしたのは、当時のExcelが大量のデータを扱えなかったからである。現在のExcelはかなり大量のデータを扱えるように改良されている。超Excelもある。データの後利用に統計解析が目的であるならば、多数のクラスを統合して1つにすると便利かもしれない。

### 6-2. 増殖型について

増殖型についても図3のように運営することでプログラムの更新は不要になった。しかし、クラスをこのままにしておくかどうかは検討課題となっている。

### 6-3. 混合型について

異なるPCに存在するデータベースを統合するためにグローバルファイルで保存する方が便利かどうか、試行錯誤をくり返している。

### 6-4. 今後の動向

M言語は階層構造のデータベースであり、大量のデータ処理に適している。また、データベース名を間接法で表示することで、汎用プログラムも簡単に作成でき便利であった。M言語をもとにしてCachéが生まれた時、広く一般に受け入れられたのは、上述（第3項）したメリットがあったためであると考えられる。しかし、クラス数が増加し続けるのは困る。いずれ統合しなければならない。最近はNonSQL言語が開発されており、今後の動向を見ていきたい。

## 7. おわりに

何分にも一学会の小さなシステムとして、パーソナルコンピュータで細々と自己開発してきたシステムである。業務が増加しても対応できるように改良を続けてきた。蓄積されたデータ量の増加にともない、小さな費用で大きな効果をあげるために、どのようにデータベースを構築していくか、それが今後の課題となっている。

## 参考文献

1. 山本和子, 伊勢田司, 岡田美保子, ユーザ側からみた Caché : CSP のメリット・デメリット-Mと比較して, 第33回日本エム・テクノロジー学会大会論文集, 2006.
2. 山本和子他, 学習評価のための問題収集・提供システムの開発, 医療情報学 2007 ; 27 (Suppl.) ; 975-976



図 2. 掃出型（各種申込・会計・成績等）システムの概要

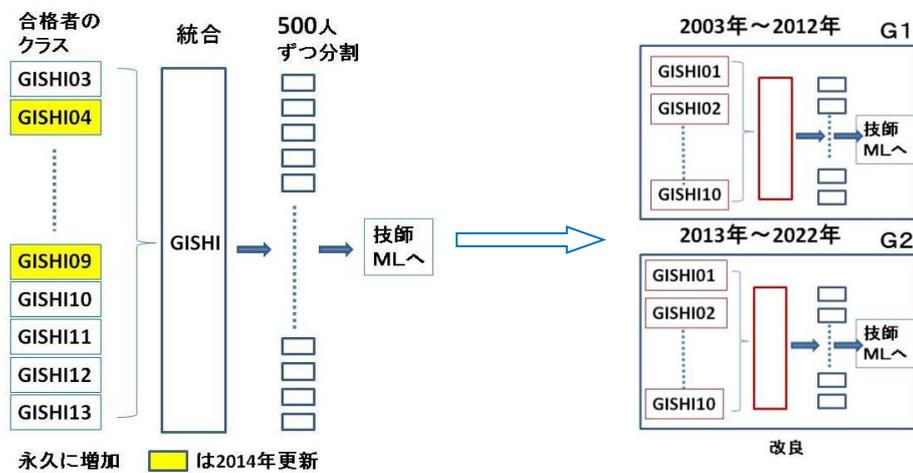


図 3. 増殖型（技師・上級合格者）システムの概要

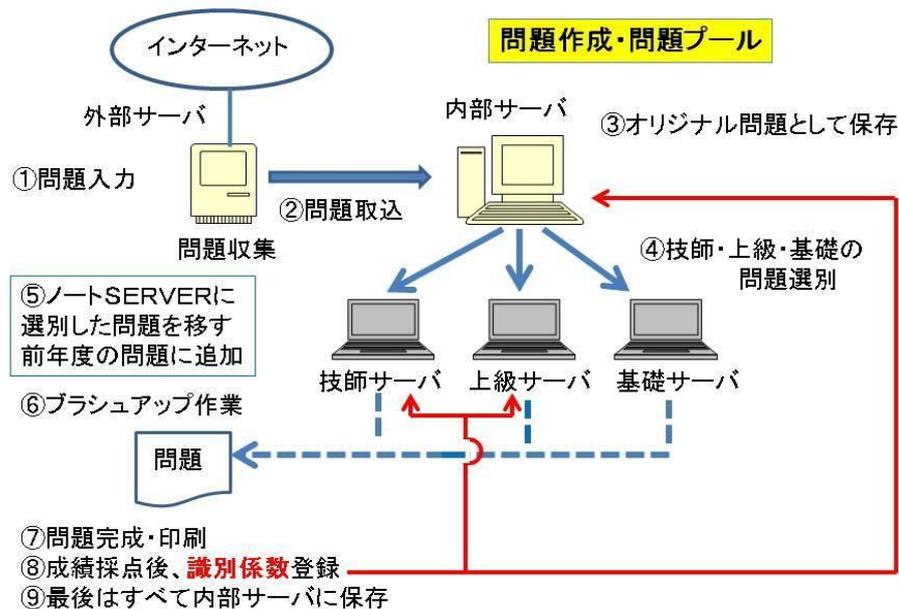


図 4. 混合型（サーバ分散型：問題作成）システムの概要

# 医療情報技師の 10 年目の更新手続きについて

最初の医療情報技師は 10 年経ってどうなったか

About the 10<sup>th</sup> year of Healthcare Information Technologist's updating procedure.

The first Healthcare Information Technologist is a situation of ten years after.

伊勢田 司<sup>1)</sup>, 山本 和子<sup>1)</sup>

Tsukasa Iseda<sup>1)</sup> Kazuko Yamamoto<sup>1)</sup>

一般社団法人日本医療情報学会医療情報技師育成部会<sup>1)</sup>

Japan Association for Medical Informatics<sup>1)</sup>

キーワード: 医療情報技師, 資格更新, 日本医療情報学会

## 1. はじめに

医療情報技師資格認定制度が 2003 年から始まって, 初代医療情報技師は 2014 年 4 月で 2 回目の資格更新手続きを終えた. 医療情報技師取得から 5 年の認定期間を終えて, 継続して医療情報技師であり続ける技師の特性, および資格更新に必要なポイントをどのように収集しているかを調査し, 今後の医療情報技師育成事業の一助とする.

## 2. 対象

表 1 2003 年医療情報技師

認定者種別	認定者	1 回更	2 回更
		新	新
2003 年技師	979 人	530 人	306 人

表 1 とその後の 2008 年までの医療情報技師とその更新手続きデータを調査する.

医療情報技師は受検時のデータを元に医療情報技師育成部会 CACHE サーバで管理している. 医療情報技師更新制度では認定取得時, または更新から 5 年間に医療情報技師として自己研鑽を積み, 勉強会や医療情報技師としての活動により 5 年間に 50 ポイント取得すること, および生涯研修セミナー参加 (e-Learning でも可) が更新の条件である. ポイントのつく学会参加については主催者からの医療情報技師参加者データ

を元に医療情報技師育成部会で管理する. 特に大規模学会 (日本医療情報学会主催学会, その他) の場合は, 認定時に医療情報技師に配布した二次元バーコードデータ付き認定証カードを二次元バーコードリーダーにて読取り, CACHE サーバにて集計したものを対象とする.

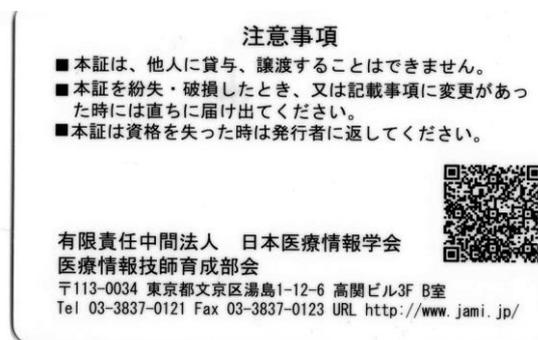


図 1 医療情報技師認定証裏バーコードの例

## 3. 方法

医療情報技師の受検時データから, 更新者数の推移, 平均年齢, 会場ごとの更新率, 2003 年の技師の会場毎の年齢構成, 勤務先ごとの更新率, ポイントのつく学会に参加した回数と獲得ポイントの平均を図表化した. また, 2014 年 4 月時点での更新者のポイントのつく学会参加上位 10 を表にあらわした.

#### 4. 結果

医療情報技師の初回認定者数、1回、2回更新者数を図2に示した。

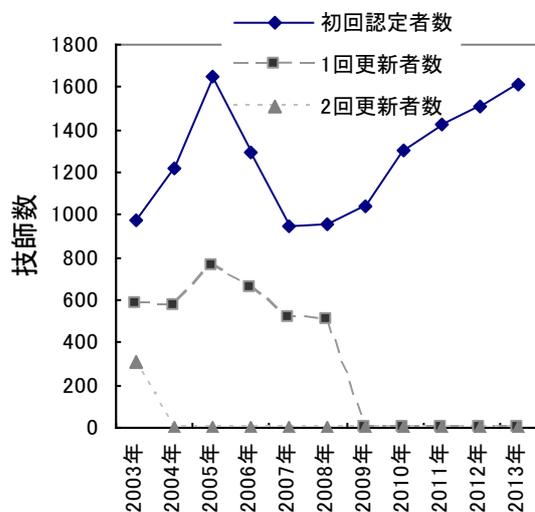
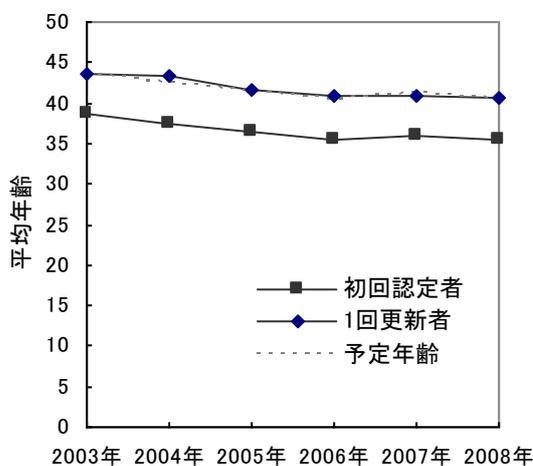


図2 医療情報技師更新者数推移

医療情報技師の認定者数は2003年から2005年を第一のピークとして増加し、その後、2007年までは2003年当初程度まで減少したが、その後、2013年まで増加を続け、2013年は2005年とほぼ同数まで回復した。更新対象者は初回認定者数とおおむね同様のグラフの動きが見られる。しかしながら、2004年と2008年は初回認定者数が前年よりも増加しているにもかかわらず更新者は前年よりも減少している。2回目の更新者は2003年のみであるが、さらに減少している。



2003年 2004年 2005年 2006年 2007年 2008年

図3 医療情報技師更新者平均年齢  
医療情報技師の初回認定者の平均年齢（更新期間に合わせて調整済）は2003年39歳から緩やかに35歳程度まで下降を続けている。

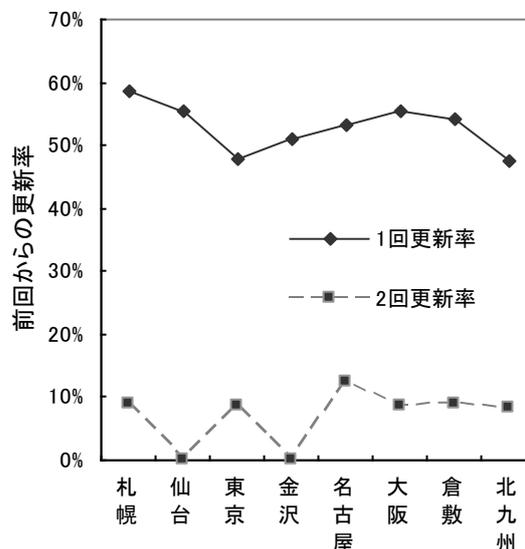


図4 医療情報技師更新者全体\_会場別  
2003年から2008年までの更新者全体を会場別に比較すると、東京と北九州で更新者が低く示された。

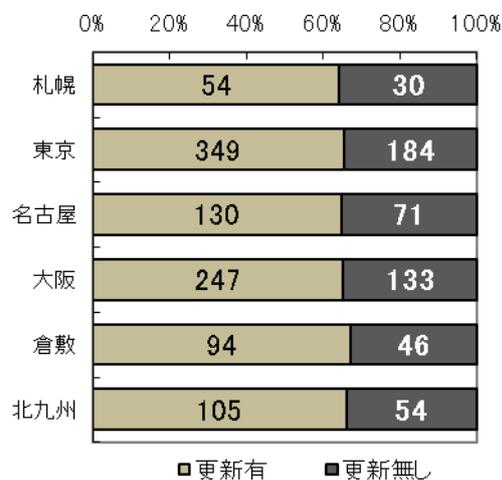


図5 2003年医療情報技師1回目更新率\_会場別

2003年医療情報技師のみの会場別更新率は1回目の更新ではほぼ同数であった。

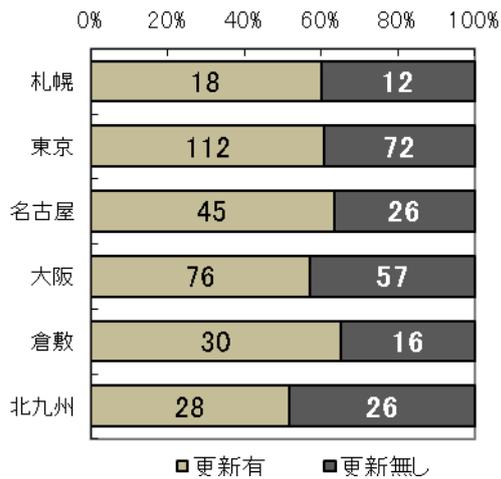


図 6 2003 年医療情報技師 2 回目更新率\_会場別

2 回目の更新は北九州と大阪でやや低く、倉敷では高かった。

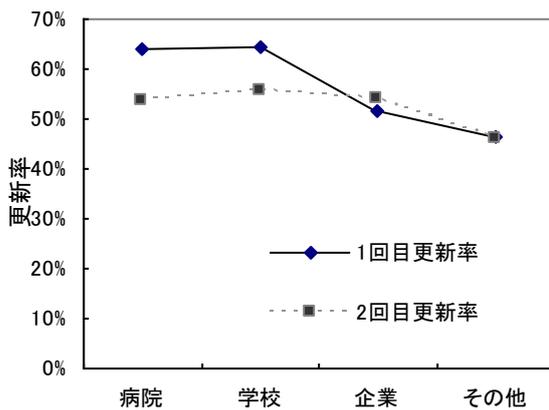


図 7 2003 年医療情報技師更新率\_勤務先

2003 年医療情報技師の 1 回目の更新率は病院、学校で高く、企業で低かった。2 回目は病院が若干企業を下回る更新率であった。学校は元々母数が少ないこともあり、更新率の影響は少ない。

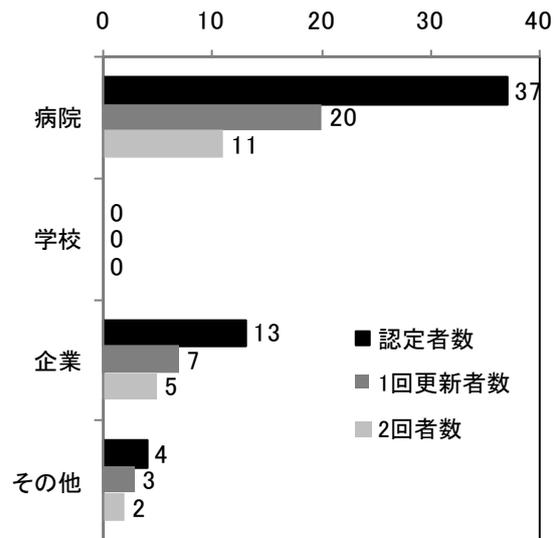


図 8 2003 年医療情報技師勤務先別更新者数推移\_札幌

2003 年技師のうち札幌会場で受検し、認定された技師の勤務先は病院が多く、次いで企業であった。当初 37 人の病院を勤務先とする技師は 2 回の更新で 11 人になった。

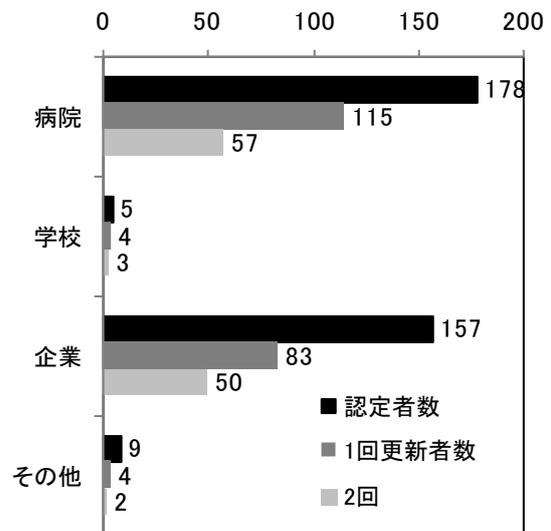


図 9 2003 年医療情報技師勤務先別更新者数推移\_東京

東京会場で受検し、認定された技師の勤務先は病院と企業が多い。当初 178 人であった病院を勤務先とする技師は 2 回の更新で 57 人となり、会場別では一番大きな 121 人の減少となった。企業では 157 人が 50 人と

なった.

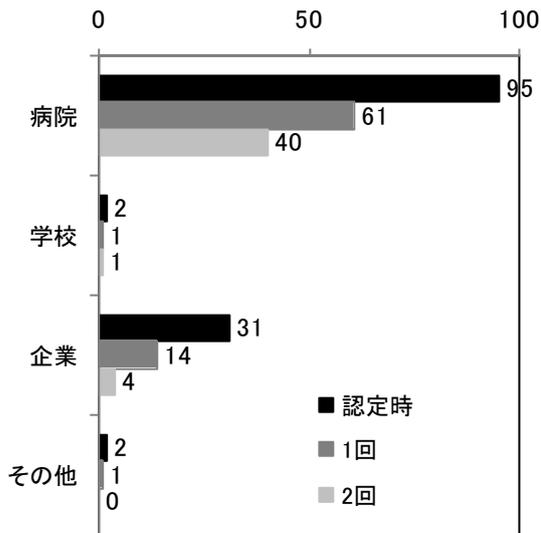


図 10 2003 年医療情報技師勤務先別更新者数推移\_名古屋

名古屋会場の技師の勤務先は病院が多く、病院では 2 回の更新で 95 人が 40 人となった。

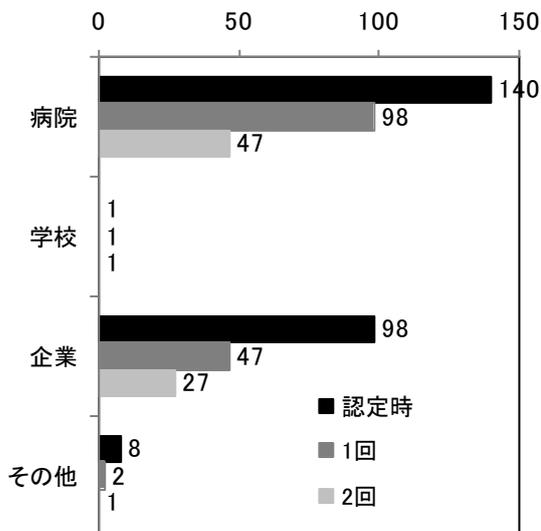


図 11 2003 年医療情報技師勤務先別更新者数推移\_大阪

大阪会場の受検者は東京会場と傾向が良く似ており、病院を勤務先とする技師が 140 人であったが 47 人にまで減少した。

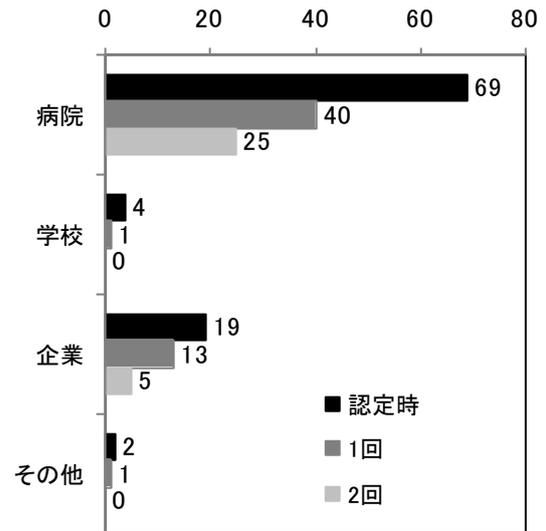


図 12 2003 年医療情報技師勤務先別更新者数推移\_倉敷

倉敷会場は病院を勤務先とする技師が 69 人から 25 人に減少した。企業は元々 19 人しか居ないところ、5 人まで減少した。

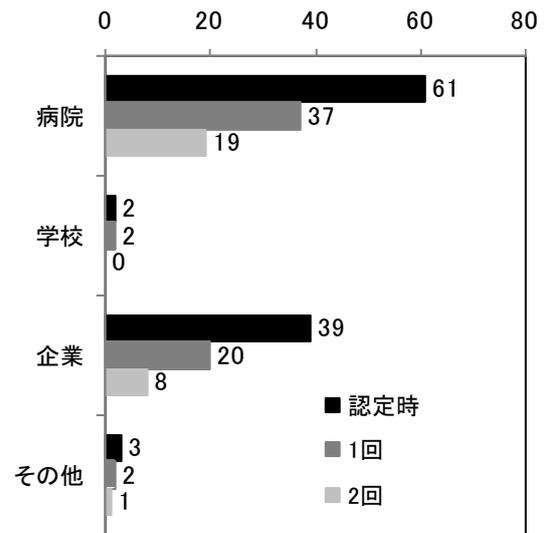


図 13 2003 年医療情報技師勤務先別更新者数推移\_北九州

北九州会場では勤務先を企業とする技師が 2 回の更新で 39 人から 8 人にまで減少した。

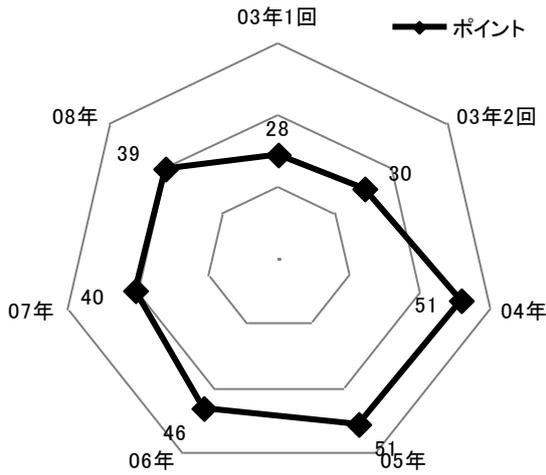


図 14 医療情報技師更新者学会参加ポイント

2003年の1回目と2回目のポイントは低く、2004年と2005年のポイントが高かった。

表 2 2008年更新および2003年2回目更新者学会参加上位10

順位	コード	名称
1	10-059	第30回連合大会(浜松)
2	09-032	第29回連合大会(広島)
3	08-025	e-Learning 医療統計セミナーB
4	08-063	e-Learning 経営セミナー
5	13-057-00	第33回連合大会(神戸)
6	11-004	第15回春シンポジウム(千葉)
7	11-051	第31回連合大会(鹿児島)
8	10-030	e-Learning 医療情報セキュリティ
9	12-070	第32回連合大会(新潟)
10	12-121	e-Learning 病院施設でのネットワーク

2014年4月の時点での医療情報技師の更新者が参加したポイントのつく学会のうち上位10では、e-Learning 以外は日本医療情報学会が主催する学術大会が上位を占めていた。

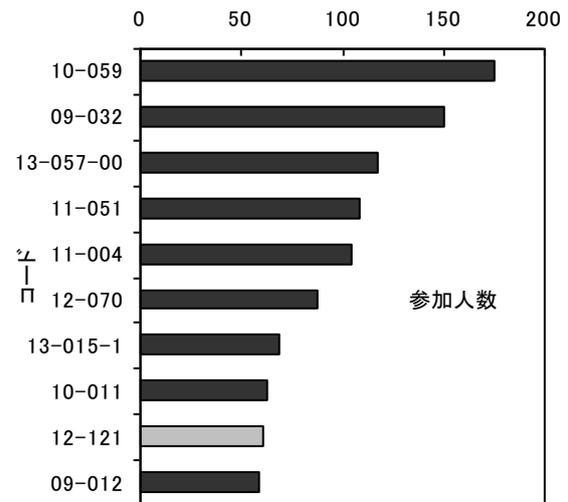


図 15 医療情報技師2003年2回目更新者の学会参加ポイント上位10

2003年の2回目更新者の学会参加ポイントは9位に医療情報技師育成部会が主催するE-Learning 講座が入っている(12-121)が、その他は全て日本医療情報学会が主催する学会である。

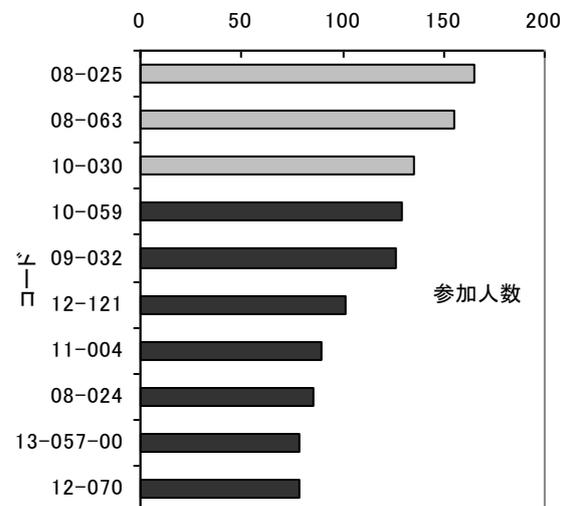


図 16 医療情報技師2008年1回目更新者の学会参加ポイント上位10

2008年の1回目更新者の学会参加ポイントは3位までが医療情報技師育成部会が主催するE-Learning 講座が入っている(08-025, 08-053, 10-030)が、その他は全て日本医療情報学会が主催する学会である。

## 5. 考察

医療情報技師の認定者は若年年齢化が見受けられるが、更新者もほぼ同様であるので、医療情報技師の更新には年齢はあまり関係がないといえる。

10年経過した2003年のみの会場別更新者から、会場別にばらつきが出始めた。2003年の技師の勤務先別更新状況では、当初高率で更新していた病院が10年経過後、更新率が低くなった。東京と大阪で病院を勤務先とする医療情報技師が特に減少している。医療情報システム周りを大手ベンダーに委ねることの多い首都圏の病院では医療情報技師の資格取得時程度の知識のみで充分とする、傾向にあるのではないかと推察される。

医療情報技師更新制度では「医療情報システムに従事していれば5年で25ポイント付与」という項目があるので医療情報技師の更新に必要な50ポイントから学会参加ポイントは25ポイントあれば最低、更新が可能である。それに加えて2003年と2004年の医療情報技師は当初、20,10ポイントと下駄を履かせた状態で更新を迎えたはずである。恩恵を受けたはずの2004年の50ポイント超え、また他の年でも50ポイント近く得点していることは、更新に有利な者が医療情報技師を継続していることを示している。

更新者が参加したポイントのつく学会上位10においても、日本医療情報学会の学会がほぼ上位を占め、これらに参加するものが更新を行っていることが示されている。

図15、図16について2回目更新者はすでにe-Learningを受講済みであるがゆえに2回目更新では、利用しないこと、1回目更新者は「生涯研修セミナー」受講のために

e-Learningを利用することを示している。

## 6. まとめ

医療情報技師誕生から10年が経過し、更新手続きも2回目を迎えた。医療情報技師として活躍し、資格を必要とする人が精査し残っていく過程を分析した。首都圏の病院勤務の医療情報技師は特に2回目の更新を行わない。これは医療情報技師資格認定のための勉強程度でその後、特に更新をしなくても十分にその役割を果たしているのか、または資格取得程度の知識では補えないほど医療情報システムが複雑化・高度化したのか、その他に要因があるのかの判断はできない。しかしながら、企業側の技師は大きく減少していない。2012年医療情報技師を対象としたアンケート調査で、医療情報技師の資格取得者の3割弱はふさわしい仕事に就いていないという結果が出ている。<sup>1</sup>

医療情報技師として10年継続した更新者は日本医療情報学会が主催する学会に参加し、規程の50ポイント以上を学会参加中心に加算することができる環境にいる人であることが判明した。また、地方によりばらつきがあるので、それを補間する地方の勉強会、e-Learningのコンテンツ追加を検討していくことが望まれる。

## 参考文献

1. 五味悠一郎, 成清哲也, 内藤道夫, 医療情報技師育成事業10年目における医療情報技師の現況調査, 医療情報学; pp1183-1187

# 院内患者位置情報表示システムと活用

## Utilization of patient position information display system

土屋喬義 1,2) 堀井路子 1) 伊藤章 1) 田中千恵子 1)

Takayoshi Tsuchiya<sup>1,2)</sup> Michiko Horii<sup>1)</sup> Chieko Tanaka<sup>1)</sup>

土屋小児病院<sup>1)</sup>, 獨協医科大学小児科学教室<sup>2)</sup>

Tsuchiya Children's Hospital<sup>1)</sup>

Department of Pediatrics, Dokkyo Medical University<sup>2)</sup>

キーワード: Caché, WebLink, MUMPS, 患者位置情報

診療現場で患者の位置情報を把握する事は、診療のボトルネックを発見出来、待ち時間の解消につながる事が出来る。またハイリスク感染患者との接触者の絞込や、医療機関にとって効率的な人員配置の検討など数多くの利点が考えられる。当院では患者位置情報を収集するために受付機、レセコンの入力受付時刻、会計時刻などで発生するイベントを使用している。またバーコードを使用してチェックポイントを追加している。今回 M 言語で独自開発した患者位置情報システムと活用事例を紹介する。

はじめに

2012 年に発表した診療順番待ち表示システム<sup>3)</sup>を作成し外来患者への情報提供に役立てている。

今回の位置情報を更に活用するために個々の外来患者の待ち時間のリアルタイム表示やチェックポイントでの滞留情報の表示を行い、診療と病院運営効率化を目指した。

方法

位置情報を可能な限りローコストにまた入力負荷をかけずに得るために、通過時間を副次的に得られる、又は推定できる(機器外来受付機、会計レジ、レセコン入力時間)より収集した。またバーコードを利用して、診察呼び出し、診察終了、検査開始、処置開

始の各時間を M (Caché) のグローバルに保存記録している。以下に通過時間記録場所(図 1)と時刻出力のタイミング、入力デバイスの種類を示す(表 1)



図 1 位置情報測定ポイント

番号	出カタイミング	入力デバイス
①	一般外来受付	再来受付機
②	非感染外来受付	再来受付機

③～⑪	診察室	バーコードリーダー
⑫	患者呼び出し	バーコードリーダー
⑬	検査処置開始	バーコードリーダー
⑭	検査結果出力	検査システム
⑮	医事入力受付	バーコードリーダー
⑯	医事入力終了	医事システム
⑰	投薬呼び出し	バーコードリーダー
⑱	会計終了	POS レジ

表1 各測定ポイントの機能と入力デバイス



写真3 会計受付用バーコードスキャナ

使用している入力デバイスの写真を示す。



写真1 診療室バーコードスキャナ

プログラムの主な構成は、サーバーサイドで Caché を使用し通信は WebLink、クライアントサイドは web ブラウザを使用し、jQuery (Ajax) を利用してユーザーインターフェイスを作成した。<sup>3)</sup>

作成した患者呼び込みシステムの動作を示す。診察室前に呼び込む患者のリストを作るため患者 ID カードのバーコードと診察室の番号のバーコードを読み込む。なお使用しているバーコードスキャナは HID プロファイルで動作するためソフト上ではキーボードとして取り扱える。



写真2 検査・処置開始記録用バーコードスキャナ



写真4 患者呼び出し用バーコードスキャナ



冬、当院で外来受診していた患児の母親が結核を発症し結核集団感染が発生した。結核の感染リスクは、発端者の排菌量、接触時間、BCG接種の有無、年齢などによって評価する。発端者は子供の診療のため頻回に来院していたもので、発端者が患児を連れて来院していた同時刻に当院に受診して接触が疑われた子供の数は1092名に達した。発端者との同室時間、感染防御の状況（マスク着用の有無と時間）を個々の患者について検討する事により感染のリスクを評価した。

当時より当院ではM（MUMPS）で動作するシステム（レセコン）を中心にMで作成した病院情報システムを運用していたため、Mのグローバルを参照する事により来院時刻、レセプト入力時刻（診療終了時刻）、会計時刻を得る事が出来た。また手書きではあったが処置開始時間のデータが残っていた。小さな病院で患者の居場所も限定されていたため、わずか4ポイントの位置情報であったが、感染リスクの高いグループ（抗結核薬の予防内服）長時間接触者8名と頻回接触者2名まで絞り込んだ。以後新たな結核の発症者を見ることなく経過し、データの解析により効率よくリスクを評価する事が出来た。

院内感染を防ぐため感染リスクの高い疾患の患者様に対しては受診前より予め病院に連絡し不要な接触により感染の拡大に努めている。しかし診断前より感染力のある麻疹患者の来院などでの接触者の調査に威力を発揮している。この様な場合接触者には適切な予防処置が必要となるが、非接触者に予防処置が行われた場合精神的負担は大きい。患者位置データの保存は医療機関

の安全性と対応力を高める重要な要素と考えている。Mのグローバルデータは非常にコンパクトなため。当院では全来院者の位置データをシステム稼働以来、11年分以上保存してある。

位置情報は診療の効率化、適正化にも大きな期待がかけられている。本来ドアなどの通過情報なども位置情報も有効に活用していきたいところであるが、デバイスの価格、サイズなどの課題があり採用に至っていない。更に有効な位置情報の活用法を考えていく予定である。また現在のJavaScriptの急速な発展を取り込み、Bigデータを取り扱えるM言語を活用して行うと考えている。

#### 参考文献

1. 鈴木智行、土屋喬義、田中千恵子、栢森猛、木村一元、“中待ち呼び込み番号および待ち時間表示システムの開発”、第37回日本MTA大会論文集、p.27、2010
2. 土屋喬義、鈴木智行、田中千恵子、木村一元：“診療文書作成プログラムの試作（Cache、jQueryを使用して）”、第38回日本MTA大会論文集、p.8-1、2011
3. 藤生崇則、伊藤章、田中千恵子、土屋喬義：“診療順待ち表示システムのための患者位置情報収集システムの作成”、第38回日本MTA大会論文集、p.48、2012

# 医事データの openEHR・アーキタイプ適用に関する検討

Investigation of Applicability of openEHR/Archetypes for Medical Accounting Data

櫻井理紗<sup>1)</sup>, 竹村匡正<sup>1)</sup>, 朝井隆裕<sup>1)</sup>, 渋谷和芳<sup>2)</sup>, 三澤一樹<sup>2)</sup>, 桑直人<sup>3)</sup>, 岡本和也<sup>4)</sup>,  
黒田知宏<sup>4)</sup>

Risa Sakurai<sup>1)</sup> Tadamasu Takemura<sup>1)</sup> Takahiro Asai<sup>1)</sup> Kazuyoshi Shibuta<sup>2)</sup>

Kazuki Misawa<sup>2)</sup> Naoto Kume<sup>3)</sup> Kazuya Okamoto<sup>4)</sup> Tomohiro Kuroda<sup>4)</sup>

兵庫県立大学大学院応用情報科学研究科<sup>1)</sup>, 株式会社ダンテック<sup>2)</sup>,

京都大学大学院情報学研究科 EHR 共同研究ユニット<sup>3)</sup>,

京都大学医学部附属病院 医療情報企画部<sup>4)</sup>

Graduate School of Applied Informatics, University of Hyogo<sup>1)</sup>

Dan-Tec. Co.,Ltd.<sup>2)</sup>

The EHR Research Unit, Department of Social Informatics, Graduate School of  
Informatics, Kyoto University<sup>3)</sup>

Division of Medical Information Technology and Administration Planning, Kyoto  
University Hospital<sup>1)</sup>

キーワード: 医事データ、openEHR/Archetype 臨床研究

## 1. はじめに

医療機関連携の必要性から医療データの標準化の必要性が言われるようになって久しく、相互運用性を考慮した診療データの交換に伴う規約として HL7 (Health Level 7) や MML (Medical Markup Language)、また、運用・蓄積方法という概念を含み、実装に関する機能を含んだ SS-MIX2 (Standardized Structured Medical record Information eXchange) 標準化ストレージ等の方法論が提示されている。しかし、医療機関連携における診療情報の交換の決め手となる標準化形式は未だ存在せず、ましてや病院情報システムにおいては種々のデータ蓄積の形式および方法を採用しており、標準化というには程遠い状況である。

一方、診療データの利用は診療にかかるシチュエーションばかりではなく、医師主導型の臨床研究に見られるような、研究活動において利用され

つつある。この中でも大規模な診療データばかりではなく、ゲノムデータ、また普段の生活のライフログデータやバイオマーカーのデータ等、様々なデータを統合的に扱う環境が必要になりつつある。しかし、これまでのアプローチでは基本的に先にデータの構造が決定し、マスタ等でデータの内容が標準化されることが前提であり、データの粒度等どのレベルでの標準を求めるかについても、どの医療機関でも抽出しうるデータで標準化を試みる、すなわち「ミニマムデータセット」を決めることが試みられてきた。しかし、先述した臨床研究にこれらのデータを利用する場合等は、ミニマムデータセットでは全く再利用は困難であり、限界があるように思われる。

そのため、我々の研究グループでは、今後の診療ばかりではなく健康のデータをも扱う PHR (Personal Health Record) 時代の到来、および多

様な臨床研究基盤の構築を目的として「マキシマムデータセット」のあり方を検討している。これはすなわち今後発生するデータの意味的な定義そのものを標準化するのではなく、あるコントロールされたオントロジーにおいて意味的な関係をもってデータの属性を規定できるようにしておけば、後々にそのデータを利用しようというコンセプトであり、これこそが健康・医療データの存在意義と考えられるからである。言い換えれば、「どのデータを取得すべきか」ではなく、取得したデータが「何者なのか」を後で同定できることである。そのため、今回我々は意味論的な統一構造下でデータの構造化を図ることを目指している openEHR を対象として、日本の医療の実データである医事データの適用を試みることで、openEHR のマキシマムデータセット基盤となり得るのかの検討を行った。

## 2. 方法

### 2-1. openEHR と Archetype

openEHR は、先述したコンセプトの元、臨床データそのもののアーキタイプという医学的に整合性のとれた意味的な構造で表現し、これをリファレンスモデルというデータ構造のモデルで表現するという形式をとっている。そのため、今回対象とする医事データでいうとアーキタイプによってどれだけ表現し得るかが重要である。

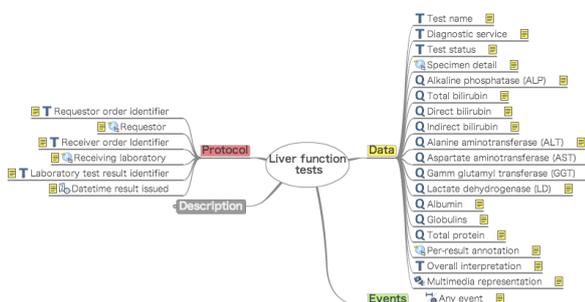


図1 アーキタイプの例 (openEHR.org より)

### 2-2. 医事データ

今回は対象とする医事データとして、DPC/PDPS における F ファイルデータとした。

## 3. 分析

K 大学病院一年間の F ファイルデータ 584701 レコード (レセ電マスタで 2704 種類) に対して、アーキタイプによってどれだけ表現可能かについて検討を行い、その上で Intersystems Cache'によってアーキタイプ構造でデータを蓄積した。結果は、F ファイルデータ 584701 に対してアーキタイプで表現できると考えられたレコードが 344252 レコード (3258 種類) であった。F ファイルのレコード上、加算等の診療行為と直接関係がなく、請求行為のみに依存するようなものは 235371 レコード (301 種類) であった。また、F ファイルデータ上医療行為を考えられるがアーキタイプに定義されていないものについては、5078 レコード (118 種類) みられた。

## 4. 考察

医事データの中で診療データとして必ずしも必要がないデータがある一方で、診療データとして扱えると思われながら、アーキタイプとして表現するのが困難なデータや、そもそも定義がないものも見られた。例えば画像診断における診断部位については、アーキタイプ上は部位が定義されているが、医事データ上はそのような分類はなされていない。また、手術中に使用した薬剤等は手術薬剤として医事データ上は情報を保持しているが、手術アーキタイプ上では薬剤のスロットは存在せず、別セクションで記載する等の検討が必要だと考えられた。

また、アーキタイプで表現可能なデータについては、臨床医学的に意味的な構造化がなされており、このデータが「何者なのか」については明確になった。アーキタイプ自体はまだ診療データや健康データを全て扱えるほど網羅性はないが、今後の充実によってマキシマムデータセットへの展開が期待できると思われる。

## 参考文献

1. Sebastian Garde et.al, Expressing clinical data sets with openEHR archetypes: A solid basis for ubiquitous computing, international journal of medical informatics; s 76S (2007) S334-S34



# 臨床研究のための多施設間診療データ連携の試み

## A Development of Multicenter Clinical Trial System at Kobe COI Project

竹村匡正<sup>1)</sup>, 櫻井理紗<sup>1)</sup>, 渋田和芳<sup>2)</sup>, 三澤一樹<sup>2)</sup>, 桑直人<sup>3)</sup>, 岡本和也<sup>4)</sup>, 黒田知宏<sup>4)</sup>

Tadamasa Takemura<sup>1)</sup> Risa Sakurai<sup>1)</sup> Kazuyoshi Shibuta<sup>2)</sup> Kazuki Misawa<sup>2)</sup>

Naoto Kume<sup>3)</sup> Kazuya Okamoto<sup>4)</sup> Tomohiro Kuroda<sup>4)</sup>

兵庫県立大学大学院応用情報科学研究科<sup>1)</sup>, 株式会社ダンテック<sup>2)</sup>,

京都大学大学院情報学研究科 EHR 共同研究ユニット<sup>3)</sup>,

京都大学医学部附属病院 医療情報企画部<sup>4)</sup>

Graduate School of Applied Informatics, University of Hyogo<sup>1)</sup>

Dan-Tec. Co.,Ltd.<sup>2)</sup>

The EHR Research Unit, Department of Social Informatics, Graduate School of Informatics, Kyoto University<sup>3)</sup>

Division of Medical Information Technology and Administration Planning, Kyoto University Hospital<sup>1)</sup>

キーワード: 他施設連携、臨床研究、臨床データ

### 1. はじめに

コホート研究を含む臨床研究においては、医療機関に限らず多くの施設やシステム間でのデータ連携が求められつつある。一方で、これらのデータはそのデータを管理する施設や、そのデータの元である患者の同意のレベル等によって、多様な管理が求められる。

神戸では、ポートアイランドを中心とした医療産業都市構想において、公益財団法人先端医療振興財団先端医療センターを中心に、「先制医療」のコンセプトのもと、様々な臨床研究が行われている。先制医療とは、医療が健康な時から各個人の将来の健康状態を予測して先手を打つことを目的とした医療であり、この先制医療に寄与するデータとして、ゲノム、バイオマーカー、ライフコース（ライフログ）データを分析し、予測することで、個人の病気への罹患を抑制するものであ

る。これらの推進に向けての活動のひとつに、文部科学省の Center of Innovation Stream (COI Stream)の枠組みにおいて、コホート研究および情報基盤の整備に向けた提案を行っている<sup>1)</sup>。COI ストリームは3つのビジョン（1.少子高齢化先進国としての持続性確保：Smart Life Care、Ageless Society、2.豊かな生活環境の構築（繁栄し、尊敬される国へ）：Smart Japan、3.活気ある持続可能な（Active Sustainability）社会の構築）に対して、2013年の公募において190の地域や拠点が応募し、12の本拠点および本拠点のビジョンを補完する11のサテライト、および将来の拠点候補（トライアル:COI-T）として14拠点（サテライト5拠点）の採用があった。神戸ではその中でCOI-Tの採択となり、現在本採用に向けた準備を行っている。

その中で、コホート連携のための情報基盤の構築において、多施設間および一般市民からの大規

模なデータを統合的に扱うシステムの構築を行ってきた。本稿ではこれらの試みについて報告する。

## 2. 方法

神戸 COI では、医学的な小児発達障害のコホート研究の推進とともに、医療のサンプリング・バイアスのない大規模な時系列データを機械学習、特にディープラーニング・状態空間モデルを用いて予測を行わせるというものである。これに合わせて大規模なパラメータを瀬底することでシミュレーションを行い、モデルを構築することで精度の高い予測モデルを構築する。大規模なシミュレーションは理化学研究所・計算科学研究機構の京コンピュータの利用が企図されている。

我々はこのプロジェクトにおいて、時系列データの引き渡し、およびライフコースデータと診療データの統合的に扱える環境の構築を行う必要がある。そのため、今回は Intersystems Cache'を用いた多施設間診療データ連携のシステムの構築を行う。

本システムは、診療機関と研究機関、およびライフコースデータを収集する機関との多施設間連携が必要である。よって、セキュアな通信環境および今後施設が増えることを考慮して、DMVPN (Dynamic Multipoint VPN)を利用して、古メッシュ型の IPsec VPN 通信環境とする。当然、これらの多施設間のデータについても統合的に取り扱う必要がある。また、先制医療のコンセプトを考えると、仮説検証型のこれまでのコホートデータ取得の考え方以外にも、データそのものから意味を抽出する必要がある。そのため、健康・医療データはとりあえずは同一の意味体系（オントロジー）において獲得し、その後の多角的な分析に対応させる必要がある。この同一意味体系に基づいた健康・医療データ蓄積方法については、別稿にて報告している。図1に今回構築するシス

テム概要を示す。

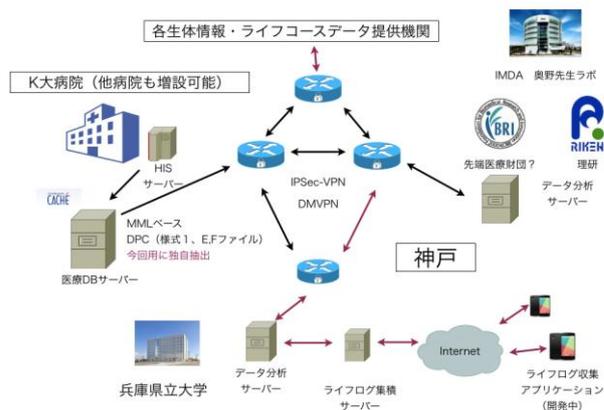


図1 システム概要

## 3. 結果

CISCO 1941 ISR を用いて DMVPM を構築し、多施設間データ連携基盤および Cache'による多施設連動データベース環境の構築を行った。

## 4. 今後

今後、大学病院より数年分の医事データや検査結果等を同一の意味体系において蓄積し、予測モデル及び大規模シミュレーションに利用する予定である。また医療・健康データは各医療機関や行政のポリシーによって、公開可能な範囲が複雑に変更されることが予測されるため、オンデマンドに各施設のデータを許可分だけ取得する環境が重要となる。そのため分散コンピューティング環境である ECP (Intersystems Enterprise Cache Protocol)を用いた多施設データ統合環境を構築する予定である。

## 参考文献

1. 文部科学省、革新的イノベーション推進プログラム (COI Stream) [http://www.mext.go.jp/a\\_menu/kagaku/coi/index.htm](http://www.mext.go.jp/a_menu/kagaku/coi/index.htm) (2014年8月20日アクセス)

# Instant Message Service 基盤としての HealthShare の活用事例

Leveraging HealthShare in a Hospital as an Infrastructure for Instant Message Service

桑田 成規<sup>1)</sup>, 山田 ひとみ<sup>1)2)</sup>, 奈良崎 大士<sup>1)2)</sup>, 朝井 隆裕<sup>2)</sup>, 竹村 匡正<sup>2)</sup>  
Shigeki Kuwata<sup>1)</sup>, Hitomi Yamada<sup>1)2)</sup>, Hiroshi Narazaki<sup>1)2)</sup>, Takahiro Asai<sup>2)</sup>,  
Tadamasa Takemura<sup>2)</sup>

国立循環器病研究センター<sup>1)</sup>, 兵庫県立大学大学院応用情報科学研究科<sup>2)</sup>  
National Cerebral and Cardiovascular Center<sup>1)</sup>  
Graduate School of Applied Informatics, University of Hyogo<sup>2)</sup>

キーワード: Service Oriented Architecture (SOA), システム間連携, アラートシステム

## 1. はじめに

比較的規模の大きな医療機関において、マルチベンダによる病院情報システムの導入は一般的なものとなっている。しかし、システムのマルチベンダ化によって、ユーザは自らの求めるシステムを手に入れることができているのであろうか。これまでは、電子カルテやオーダーリングシステムといった基幹システムの導入企業が System Integrator (SIer) の役割を果たしつつ幹事企業として立ち振る舞い、そのデータ連携仕様のみならず、部門システムのベンダ選定にまで強い影響力を及ぼしてきたといえる。

このような場合、稼働開始後に発生するデータ連携方法の新設・改変に際して“SIer”の仕様や意向が制約条件となることが多く、ユーザの要求に十分対応できない。すなわち、この種のシステムはハブとなる基幹システムに依存せざるをえず、システム間データ連携という観点からみれば、その実態は幹事企業によるシングルベンダのシステムである。したがって、利用環境の変化に伴い新たなシステム間連携を必要とするユーザ側の自由度は低く、またその費用は高止まりとなる。このような医療機関のユーザあるいは経営者は、

マルチベンダ化の恩恵を十分に被っているとはいえない。

## 2. SOA の活用

上述の“不自由な”スキームに由来する課題に対し、より柔軟なシステム間データ連携の実現を可能とする一つのシステム設計概念が Service Oriented Architecture (SOA) である。これは簡単に言えば、連携のハブ機能を基幹システム側から切り離し、特定システムに依存しない形でデータ連携機能を提供するしくみである。医療機関における SOA の導入はすでにいくつかの報告がなされており、システム間データ連携やデータ二次利用に活用されている<sup>1)2)3)</sup>。

一方、国立循環器病研究センターにおいては、2018 年度に予定されているセンター移転に向けて、専門医療連携システム（いわゆる地域医療連携システム）の充実を図っている。その計画において、2013 年度には、SOA を実装するにあたり IHE プロファイルへの適用が容易な Intersystems 社の HealthShare<sup>4)</sup>を導入した。HealthShare の適用範囲は医療連携だけでなく、病院内 Enterprise Service Bus (ESB) としての検査部門システム一

基幹システム間連携、ETL ツールとしての列指向型 DWH の構築においても活用されている。

さらに 2014 年度には SOA の適用範囲を拡大し、システムから発報されるメッセージをリアルタイムでユーザに届けるための基盤システムを構築した。これは、医療安全の観点から、システム側で何らかの重要なデータ変更（例えば、注射速度指示の変更、アレルギー情報の追加など）がなされた場合に、適切なタイミングで適切な医療従事者に対してそれを知らせること（アラート）がきわめて重要であることに端を発する。従来の SOA 不在のシステムであれば、アラート機能は基幹（あるいは部門）システム側のアプリケーションに組み込むしか方法はない。しかし、アラートに必要なロジック（発報のおよび通知先の判断論理）はユーザ側のニーズに応じて高度に customizable でなければならず、従来のスキームでは費用対効果の面から、その実現は限定的な範囲に留まるのが実情である。そこで、SOA の特長を活かし、基幹システムだけでなく多数の部門システムからアラート発報に必要な材料（データ）をリアルタイムで収集し、かつそれを広範囲に適用可能な Instant Message Service 基盤システムの構築に至った。

### 3. システム設計

図 1 にその実装例として、Instant Message Board System (以下、IMBS) の概要を示す（実際の実装形態を簡略化して表示している）。医師のオーダー変更を看護師に伝達する場合のフローを例示すると、①医師によるオーダー変更操作、②電子カルテ参照系 DB へのリアルタイムレプリケーション（基幹システム側機能）、③HealthShare による DB 変更履歴の取得および所与のロジックによるアラート文生成、④⑤IMBS サーバによるアラート文取得および DB への書き込み、⑥⑦端末側 service daemon へのプッシュ、⑧端末側アプリケーションによるユーザへの画面表示、となる。端

末側アプリケーションは、電子カルテのログインおよび患者切り替えのタイミングで、電子カルテ側からのトリガーにより、どのユーザが当該端末を利用し、どの患者のカルテを開いているかが判別できるしくみである。これにより、HealthShare 側で生成されるメッセージを、ロジックにより定められた宛先に対して適切なタイミングで送信することができる。

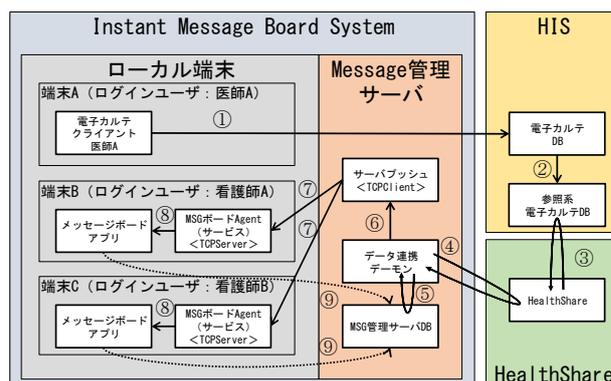


図 1 Instant Message Service 基盤の概要

### 4. おわりに

本システムの本質は、医療安全管理分野での活用を想定したアラート機能である。従来、このような機能は、禁忌薬チェックや医事算定チェックなどの限定的な局面での利用に限られてきたが、今後は SOA の活用により、より多くの領域でアラートシステムが活用される可能性がある。とりわけ、この機能の中核をなすメッセージ生成にかかる判定ロジックを、エンドユーザ自身がメンテナンスできる環境が整備されれば、さらなる発展が期待できると考えられる。

### 参考文献

1. 飯田征昌, 診療情報集約のための SOA を活用した院内システム間連携基盤の構築とその実効性, 新医療 40(6): 92-96, 2013
2. 沼澤功太郎ほか, 名古屋大学医学部附属病院電子カルテシステムでのキャシュ・Ensemble・FileMaker 活用, 第 40 回 M テクノロジー学会大会論文集, 2013
3. 下川忠弘, 当院におけるメッセージ・ウェアハウスの構築, 医療情報学連合大会論文集, 33: 420-422, 2013
4. 上中進太郎, Ensemble と HealthShare の医療向け機能, InterSystems Symposia 2014 講演資料 (B3), <http://www.intersystems.co.jp/events/2014/2014-symposia/pdf/b3.pdf>, 2014

# Mで作成した院内情報システムより

## 「とねっと」へのデータ転送

(埼玉利根保健医療圏地域医療ネットワークシステム)

### HIS Data transfer to Tonet(region healthcare record system) by MUMPS

○伊藤 章<sup>1)</sup>, 田中 千恵子<sup>1)</sup>, 堀井 路子<sup>1)</sup>, 土屋 喬義<sup>1)2)</sup>

Akira Ito<sup>1)</sup>, Chieko Tanaka<sup>1)</sup>, Michiko Horii<sup>1)</sup>, Takayoshi Tsuchiya<sup>1)2)</sup>

土屋小児病院<sup>1)</sup>, 獨協医科大学小児科学教室<sup>2)</sup>

Tsuchiya Children's Hospital<sup>1)</sup>

Dokkyo Medical University - Pediatrics<sup>2)</sup>

#### SS-MIX HL7 とねっと 地域医療情報ネットワーク

##### 【はじめに】

「とねっと」とは、平成24年7月1日から稼働した地域医療情報連携システム(図1)です。利根保健医療圏(埼玉県北東部)においてITを活用し、患者・かかりつけ医・中核医療機関・タブレット端末を搭載した救急車の4つを連携させて、地域完結型医療をめざしています。さらに本人の健康自己管理にも利用しているシステムです。

◆ 利根保健医療圏における地域医療ネットワーク(イメージ)

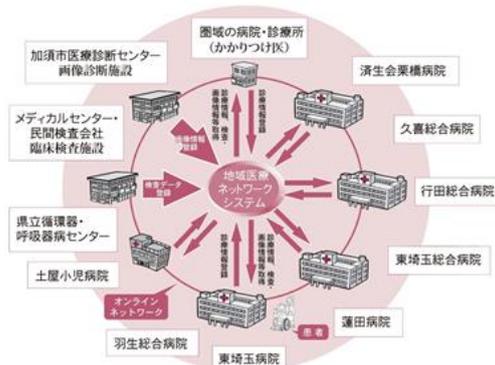


図1 地域医療情報連携システム

平成26年8月20日現在で、7市2町推計人口66万人中、21,587人の方がとねっとに参加しています。データの発生す

る病院・検診施設・検査会社は16施設、閲覧できる参加医療機関は103施設あります。さらにインターネットに接続した加入者のPCから、参加医療機関で受けた検査結果やご自身で入力した検査結果を閲覧でき、ご自身の健康記録を登録することもできます。

##### 【目的】

中核医療機関として地域医療情報の連携や共有を行うため、当院を受診したとねっと加入患者のデータを、Mで構築された当院の医療情報システムから、必要情報を切り出し、指定手順に従ってとねっとサーバーへデータ転送することを目的としています。システム構成を図2に示します。

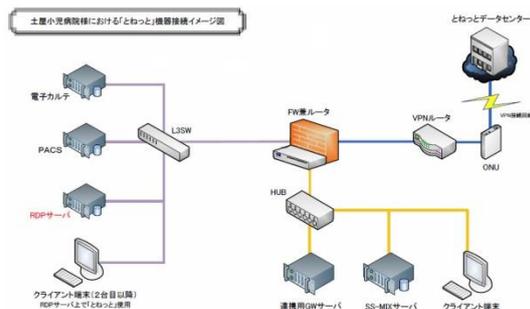


図2 システム構成

【方法】

診療情報のグローバル（CAS）から処方データを切り出し、セグメントで構築されたメッセージファイル（図3）を転送します。

4.2.1 セグメント構成

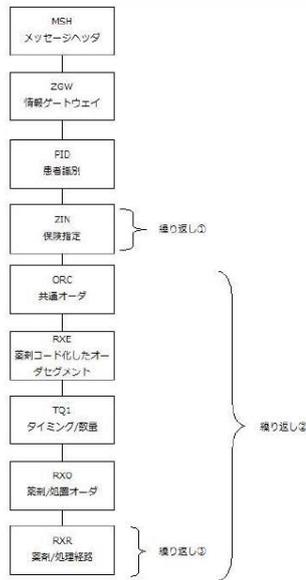


図3 フォーマットフロー

とねっと側サーバーはSS-MIX (Standardized Structured Medical Information Exchange)標準化ストレージを採用していますので（図4）、当院からはHL7 Ver2.5ファイル形式で格納します。

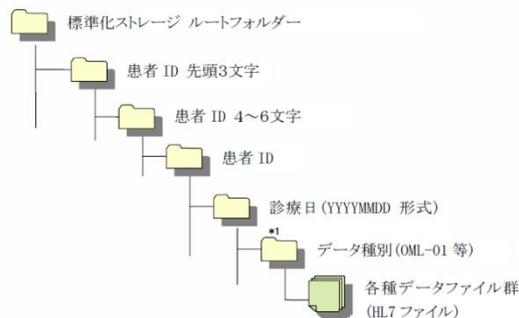


図4 標準化ストレージの構造

以下に手順を示します。

- 1) 新たな会計情報の発生を確認  
`I $D(@%CIX@(%YMD,ISQ))=1`  
`D LP1 S ^RPT(%DD)=^RPT(%DD)+1`
- 2) 調剤料を取っているものを取得  
`F I=1:1`  
`S SIN=$O(@%CAS@(IDN,SYM,NG,0,SDD,DSQ,SIN))`  
`Q:(SIN="")!(SIN>23.99) D LP2`

- 2) グローバルから各セグメントに必要な項目の抜き出し  
 患者番号(KID)、氏名(KNM, KFR)、性別(SEX)、診療日(YMD)、発生日時(YMD HMS)、RP番号(RPCNT)、薬剤コード(DRNM)、薬剤名(DRMD)、投与量(DRVL)、用法(IKAI, YHBF)等

- 4) 格納位置・ファイル名の作成  
`S PATH="T:¥"_$E(KID,1,3)_"¥"_$E(KID,4,6)_"¥"_KID_"¥"_%DAY_"¥OMP-01"`  
`S FNAME=PATH_"¥"_KID_"_"_FYMD_"_OMP-01_"_ISQ_"_"_MSH7_"0_-1"`

- 5) メッセージの作成  
 メッセージは定型化されているので、必要なフィールドだけ変数として値を代入しています  
 以下にMSHセグメントの例を示しますが、7番目のフィールドMSH7だけ変数としています  
`S MSH="MSH|^~¥&|||||",MSH7,"||OMP^O09|1|P|2.5|||||~ISOIR87||ISO 2002-1994||"`

6) 連携サーバーの指定フォルダにHL7形式のファイルを作成

O FNAME:("NWS") U FNAME

D MSH; メッセージヘッダ

D ZGW; 情報ゲートウェイ

D PID; 患者識別

D ZIN; 保険指定

C FNAME U 0

----- 以下繰り返し処理 -----

O FNAME:("WSA") U FNAME

D ORC; 共通オーダ

D RXE; 薬剤コード化したセグメント

D TQ1; タイミング/数量

D RXO; 薬剤/処置オーダ

D RXR; 薬剤/処理経路

C FNAME U 0

### 【成果・考察】

システムワーキンググループ会議で配布されていた「データ連携仕様書」を基に、プログラムを作成しました。アプリケーションといってもユーザーインタフェース部分がないので設計時間は短く、2~3週間のボリュームと見ていましたが、サンプルが提供されていたので短時間で作成することができました。

作成されたファイルは以下(図5)の内容になり、連携GWサーバーの指定フォルダ(図6)に格納しました。



図5 HL7データ

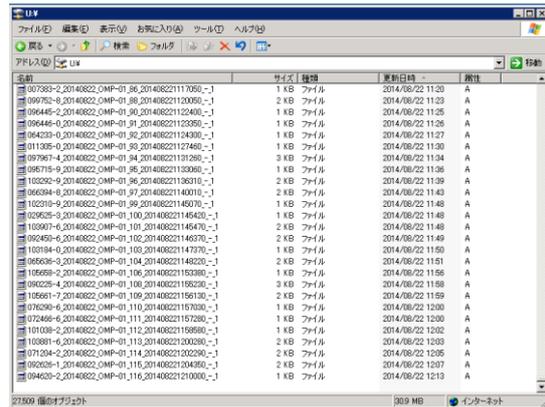


図6 指定フォルダに格納

とねっとサーバーは、定期的に連携GWサーバーのデータファイルを確認し、とねっとデータベースに展開します。

とねっと健康記録から見た結果(図7)を示します。

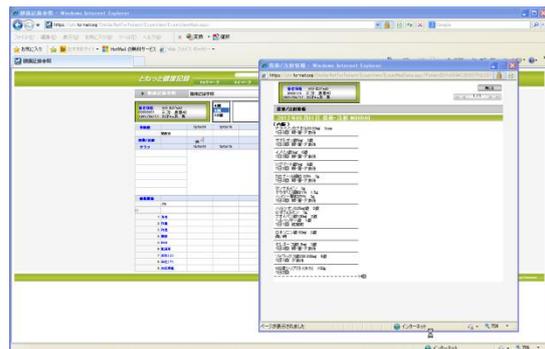


図7 とねっと健康記録

手始めに投薬情報を作成しましたが、これを参考に患者情報・検査情報も同様に作成し、地域医療情報の連携と共有に寄与してまいりたいと考えています。

### 参考文献

1. 地域医療ネットワークとねっとホームページ  
<https://sites.google.com/site/tonetsince2012/home>
2. SS-MIX 標準化ストレージ仕様書 SS-MIX 普及推進コンソーシアム

# 象徴的記号のシンボル・プロセッサと M 言語の階層型データの演算子的アプローチ

Symbol Processor for the Symbolic Sign Codes  
and Some Operational Approach with Hierarchical Data of M Language

高橋 亘

Wataru Takahasi

基礎コミュニケーション科学研究所

Institute for Fundamental Communication Science

キーワード: 象徴的記号, シンボル・プロセッサ, 知的障害, 自閉症, 階層型データ

## 1. はじめに

言葉に障害があったり、言葉に興味を持たなかったりする人・児童の意思表示に、絵記号を活用することが考えられる。これまで、こうした試みの多くは PIC によるものが多かったが、我々は、人が言語 (記号) 機能とは独立に保持している象徴機能 (ピアジェ) を活用した象徴的記号による言語形成の理論を提唱している。我々の基礎にあるツールは象徴的記号のシンボル・プロセッサである。

我々のシンボル・プロセッサは二語文、三語文の機能的な構成を可能にするが、こうした機能には M 言語の階層型データが活用されている。絵記号を言語記号として連合的かつ連辞的に配列していくには、階層型データを機能的に活用することが有効であるが、今回の発表では、こうした技術を総合的に報告するとともに新展開についても述べたい。

## 2. 象徴的記号の立場

我々の、絵記号を用いた意思表示システムの研究<sup>1)</sup>は 2007 年に始まるが、当初の

我々の立場<sup>2-8)</sup>は、カナダの S. C. Maharaj によって開発された PIC (Pictogram Ideogram Communication) の拡張によって絵記号による言語化を試みるものであったが、この立場には多くの制約や限界性があることが判明した<sup>2-8)</sup>。我々は、2012 年ころから、こうした制約からの脱却を図るため、発達心理学と言語学及び認知科学の基盤を明確にすることによって言語化に最適な絵記号の体系を構築することに目標を定めた。

我々の発達心理学的基盤が認知発達心理学者 J. ピアジェの提唱する象徴機能<sup>9)</sup>であることを 2012 年の学会では公表した<sup>10)</sup>。ピアジェはそれまで混乱のあった人間の心象を表現する記号機能と象徴機能を明確に区別した。象徴機能とは心の中にある概念を具体的な事物 (象徴) を用いて表現してみせる機能であり、記号 (言語) が表象性を持たないのと対照的である。言語機能が発達する前段階の児童が、身振り手振りで自分の要求を主張する姿はよく見受けられるものである。我々はこの象徴機能に、言語機能の代用として概念を表現する働きを期待するのである。この意味で我々が使用す

る絵記号は概念を簡明に表現できるもの、心象をよく表象するものが好ましいと言える。

### 3. 絵記号を言語記号のように使用する

我々が目標とするのは、言語学的には象徴的な絵記号を言語記号(言葉)のように使うことである。これは、発達心理学的には象徴機能を活用して言語機能を形成することである。こうした目標を明確にするため我々の絵記号の体系を象徴的記号(SSC; Symbolic Sign Codes)と呼ぶことにした。

我々は、絵記号によって表現されたメッセージが音声言語と自然な対応関係を持つことを重要視した。音声言語は、近代言語学の祖、F. ソシュールが主張するように線条性(一次元に配列)を保持する。これに対して絵記号は本来的に二次元の対象であるから、線条性という条件は意図的に課されなければならない。象徴的記号では記号の線条性という条件によって絵記号の表現法・使用法に制限を加える。2012年の発表<sup>1)</sup>では、このほかに、絵記号が言語記号と自然な対応を持つための条件を4つ挙げている。

F. ソシュールは記号が線条的に配列される方向に連辞関係(文法)を観ると同時に、文法的役割を担っている個々の単語が、その単語が所属する範疇の中から選択された存在であることを観て取っている。彼はこれを記号の連合関係と呼んだ。この記号が、それが属する同質のカテゴリの群的関係を切断しているという意味で、数学的な群の切断(cross section)に対比されるものである。連辞関係で繋がる線条的多様体とこれと直交する連合関係との直積は幾何学的

な用語でファイバー束(バンドル)と呼ばれる多様体である。言語記号がこのようなファイバー束の要素として記号範疇と記号対象によって分類されるのであるから、象徴的記号もまた文法的範疇と記号名称によって分類されなければならない。こうして象徴的記号はM言語の2階層の大域変数 $\wedge$ SSCTreeでカテゴリ分類が定義されることになる。

$\wedge$ SSCTree(記号範疇, 記号名称)

=シンボルのパス $\wedge$ 日本語 $\wedge$ 語活用

この大域変数を使用して現実に記号選択を行うシンボル・プロセッサの画面は図1の通りである。



図1 プロセッサの記号選択

図1では、上部に記号範疇として{人物, 植物, 動物, 道具, 乗物, 衣類, 食物, 飲物, おやつ, 属性, 感情, 行為, 場所, 時間, 気象, 数量, 指示, 疑問}の18カテゴリが用意されていることがわかる。図では、このカテゴリの中から植物が選択され、クリック毎に、20を単位に、このカテゴリに属する象徴的記号が中段にリストされる

様子が示されている。象徴的記号のクリックともなつて下段に選択された記号が一行に並んでいく仕組みである。

#### 4.2 記号結合の認知科学的背景と文法構造の生成

幼児の言語習得には、片言しゃべりに引き続く、二語期と呼ばれる時期があつて、これが文法習得の最初であると考えられている。我々のシンボル・プロセッサでは、当初から文法習得における二語期を重要視し、二語結合の編集に特別なインターフェイスを用意してきた。(図 2)

このような二語結合のインターフェイスは、後に、象徴記号の記号列に如何にして助詞の効果を表現させるのかという考察に端を発して、二語結合について認知科学的

な強度の差という視点を導入するに至つた。少人数でのコミュニケーションで主語が屢々省略されるのに対して他動詞は目的語を省略したのでは意味が通じない。このような思考実験から目的語と他動詞の結合は主語と動詞の結合より密接であると言える。

近年の認知科学的成果、例えばネーデルとモスコヴィッチの多重痕跡理論などによれば「意味記憶はエピソード記憶の痕跡から抽出される」ということが言われている。この原理に従えば目的語は他動詞の前に置かれるのが自然である。二語の結合強度と意味記憶の生成原理を押さえれば、始原言語の語順や基本的文法が推定される。我々は、2011年から2012にかけて、こうした方法論を用いて、知的障害児などに理解しやすい他動詞文の形を提案した<sup>7,8,10,12)</sup>。



図 2 プロセッサの二語文脈の編集画面

以上のような言語発達上の理由および認知科学的な記号形成の議論からくる理由の2つの視点から、二語文を特別扱いして編集することを可能にしている我々のシンボル・プロセッサは、象徴的記号による言

語形成に重要なヒントを保持していると考えられる。図 2 では「スポンジケーキにどんぐりを載せる」という、童話の一場面を説明する文を表現しているところであるが、[スポンジケーキ]の選択に引き続いて、[ど

んぐり][載せる]の部分で二語文脈編集機能によって構成している局面である。[載せる]が選択されると、[どんぐり][載せる]の二記号の間隔が通常より狭く打ち出され、対応する日本語文は“スポンジケーキにどんぐりを載せる”のように助詞を補って表示される。かつては、二語期を支援するために導入された二語文編集画面が、今日では、潜在意識下にある記号結合の強度を表現するものとして、象徴的記号による本格的な構文形成のために使用されている。

二語文の編集を可能にしている M 言語の大域変数は `^ConnectAmp` であり、この大域変数は、先行する象徴的記号の記号範疇と記号名称、後続する象徴的記号の記号範疇と記号名称の合計 4 つの引数を保持している。この 4 階層データを演算子的に活用しつつ、三語文・四語文を構成することで、象徴的記号で直接表現することの困難な日本語の助詞や活用語尾を制御しているのである。

## 5. まとめと展望

象徴的記号のシンボル・プロセッサの近年の立脚点について総合報告した。我々は絵記号による言語形成に取り組む必要性から、旧来の PIC にとって変わる象徴的記号を基礎とする立場をとることになった。

象徴的記号は認知発達心理学者 J. ピアジェの提唱する象徴機能を活用し、象徴的に使用される絵記号に言語形成のための合理的な制約を課することによって認知科学的により認識しやすい記号体系を目指すものである。

我々のプロセッサは、M 言語の階層型データを記号分類と記号結合の二つの観

点から演算的に活用して、三語文・四語文を構成している。今後は、さらに高次の構文規則に挑戦することが期待される。

## 参考文献

1. 高橋 亘, 柳内英二, “PIC シンボルによる知的障害者のコミュニケーション支援システムの M 言語による実現”, 『Proceedings 2007 M Technology Association of Japan』, 19~23 (2007).
2. 高橋 亘, 柳内英二, “PIC シンボルによる知的障害者のコミュニケーション支援システム”, 『関西福祉科学大学紀要』, Vol. 11, 49~54 (2008).
3. 柳内英二, 高橋 亘, “PIC シンボルによる知的障害者の意思表示システムの M 言語による実現”, 『Mumps』, Vol. 24, 41~47 (2008).
4. 高橋 亘, 柳内英二, “PIC シンボルによる知的障害者の意思表示システム”, 『関西福祉科学大学紀要』, Vol. 12, 41~48 (2009).
5. 高橋 亘, 柳内英二, 池田茉莉子, “PIC シンボル・プロセッサの実現と視覚記号の言語学”『総合福祉科学研究』, Vol. 1, 101~114 (2010).
6. 池田茉莉子, 柳内英二, 高橋 亘, “M 言語による PIC シンボル・プロセッサの実現と視覚記号の言語学”, 『Mumps』, Vol. 25, 35~50 (2010).
7. 高橋 亘, “始原言語の形成と PIC シンボルの言語学”, 『総合福祉科学研究』, Vol. 2, 15~24 (2011).
8. 高橋 亘, 柳内英二, “PIC シンボル・プロセッサの三語文脈判断と電子 LL ブックの構成”, 『Mumps』, Vol. 26, 27~37 (2011).
9. J. ピアジェ著, 中垣啓訳, 『ピアジェに学ぶ認知発達の科学』, 北大路書房 (京都, 2007).
10. 高橋 亘, “象徴的記号による言語形成”, 『関西福祉科学大学紀要』, Vol. 16, 1~16(2012).
11. 高橋 亘, “象徴的記号による言語形成とシンボル・プロセッサ”, 『MTA2012 第 39 回 M テクノロジー学会大会論文集』, 15~16 (2012).