International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

2019IJIEMR.Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 6th Mar 2018. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-03

Title: E-HEALTH DATA MINING BY APPLYING ACCESS CONTROL OVER DATA

Volume 08, Issue 03, Pages: 23–26.

Paper Authors

MR.B.PRADEEP KUMAR, K.VIJAYA LAKSHMI

Vignan's Lara Institute of Technology & Science

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

E-HEALTH DATA MINING BY APPLYING ACCESS CONTROL OVER DATA

MR.B.PRADEEP KUMAR[1], K.VIJAYA LAKSHMI[2]

Assistant Professor[1], Department of M.C.A ,Vignan's Lara Institute of Technology & Science

M.C.A Student[2],Department of M.C.A, Vignan's Lara Institute of Technology & Science

Abstract:

Reliable research demands data of known quality. This can be very challenging for electronic health record (EHR) based research where data quality issues can be complex and often unknown. Emerging technologies such as process mining can reveal insights into how to improve care pathways but only if technological advances are matched by strategies and methods to improve data quality. The aim of this work was to develop a care pathway data quality framework (CP-DQF) to identify, manage and mitigate EHR data quality in the context of process mining, using dental EHRs as an example. Objectives: To: 1) Design a framework implementable within our e-health record research environments; 2) Scale it to further dimensions and sources; 3) Run code to mark the data; 4) Mitigate issues and provide an audit trail. Methods: We reviewed the existing literature covering data quality frameworks for process mining and for data mining of EHRs and constructed a unified data quality framework that met the requirements of both. We applied the framework to a practical case study mining primary care dental pathways from an EHR covering 41 dental clinics and 231,760 patients in the Republic of Ireland. Results: Applying the framework helped identify many potential data quality issues and mark-up every data point affected. This enabled systematic assessment of the data quality issues relevant to mining care pathways.

Introduction

The emergence of Infrastructure as a Service framework brings new opportunities, which also accompanies with new challenges in auto scaling, resource allocation, and security. A fundamental challenge underpinning these problems is the continuous tracking and monitoring of resource usage in the system. In this paper, we present ATOM, an efficient and effective framework to automatically track, monitor, and orchestrate resource usage in an Infrastructure as a Service (IaaS) system that is widely used in cloud infrastructure. We use novel tracking method to continuously track important system usage metrics with low overhead, and develop a Principal Component Analysis (PCA) based approach to continuously monitor and automatically find anomalies based on the approximated tracking results. We show how to dynamically set the tracking threshold based on the detection results, and further, how to adjust tracking algorithm to ensure its optimality under dynamic workloads. We demonstrate the extensibility of ATOM through virtual machine (VM) clustering.

Lastly, when potential anomalies are identified, we use introspection tools to perform memory forensics on VMs guided by analyzed results from tracking and monitoring to identify malicious behavior inside a VM. We evaluate the performance of our framework in an open source IaaS system.

Existing system:

Our case-study describes a scenario where the researcher has direct access to the data through SQL Server Management Studio. This access allowed addition of the metadata fields to the research data, database scripting, inclusion of additional clauses in the cohort selection process etc. The current framework design incorporates assumptions based on this scenario. Different research scenarios may require alternative approaches, for example, storing the DQ metadata in distinct and separate tables or locations, or database normalization measures

Proposed system

The proposed database design (Figure 2) fulfills the requirements of the case-study herein. Other scenarios may require redesign. Simpler case-studies may only require the DQIssuesRegister while more complex scenarios may require further normalization of the database to improve data integrity and reduce data redundancy. It is unknown how this would impact the performance of cohort selection queries. Our case-study deals with research data from a single, homogeneous EHR source. Consideration needs to be given to additional DQ matters such as 'Variety' in scenarios with complex, multi-source, multi-institution research projects using heterogeneous data sources - perhaps as approached by Knowlton et al.

Module Implementation

1) Add Metadata to the research data. Mark-up fields are added to the research data allowing us to store DQ information with the data element (usually a row). This information can be used to exclude the data from the dataset as it is extracted for a specific experiment. Suggested fields are: a Boolean called BadRow and a vector string called BadRowCodes. The vector string can hold multiple error codes simultaneously.

2) Pre-processing or discussion section? Decide where the DQ issue is to be dealt with, in pre- processing or by way of discussion. This will determine whether we can mark the data with this issue. If not, we will address it in the research discussion.

3) Does an issue disqualify the data from the experiment? If the DQ issue is serious for any specific experiment, the experiment should be marked, and the data excluded from use there.

4) Evaluate the effect of these data disqualifications. Does it require re-execution of previously executed marking or mitigation code? Does it skew results? E.g. Removal of data may violate previously satisfied data integrity constraints.

5) Write/Run the Marking Code from the CP-DQF against the data. Executing the code stored with the DQ issue in the register will mark the research data's metadata with information about its DQ. e.g. Mark orphaned treatments (no client exists) as 'bad' Update PMTreatments set BadRow =1, BadRowCodes=

Concat(BadRowCodes,' 7') where ClientID not in (select PMClientID from PMClients);

6) Write/Run the Mitigation Code against the data. Executing the mitigation code (if exists) will update the research data to improve its quality.

7) Update the DQ issues register with the results. Record the scope of the issue and the scope of the mitigation efforts – primarily for reporting purposes.

8) Write/Run the CohortSelection Code. Cohort/Dataset selection code can now be written incorporating the metadata as a criterion for exclusion/inclusion in the data set. In the implementation below, treatment events are only selected if the metadata, BadRow is NULL

Conclusion:

The design for the CP-DQF data quality framework has been presented. It is implementable as a software tool that can be used to manage the DQ issues of research using EHRs. In our work we have applied the CP-DQF framework to a large dental EHR and the framework proved useful in providing a structured method to identify and document issues following the DQ dimensions established by the existing literature, notably. Our example illustrates how code to mark the data to mitigate DQ can be implemented. Intimacy with the data was helpful in identifying many of the information sources and data issues. The case study also showed DQ issues linked to individual experiments in the research and how this can cause affected data to be excluded if appropriate. The CP-DQF framework has the functionality to be used as an audit trail tool for all data transformations and data cleaning activities. This would satisfy the demands for greater transparency in the pre-processing of EHR-data in preparation for research. By slightly varying the cohort selection criteria, it is also possible to compare research results before and after the exclusion of bad quality data the impact. While the framework was prototyped in the Microsoft SQL Server environment, researchers in other environments could easily replicate this design. The entity design is simple but effective and the dictionaries of sources, dimensions and levels can be tailored to the research. Use of the CP-DQF may help researchers think about the potential DQ issues in their research, log and manage them in a structured environment, create an audit trail for data transformations, assess and mark their data with quality information, mitigate the issue if possible, exclude data from their experiments if appropriate, compare before and after research outputs and finally, report on DQ metrics. This will lead to known and more robust EHR DQ, a secure audit trail of DQ transformations, reproducible research steps and more reliable process mining results

References

[1] N. Weiskopf and C. Weng, "Methods and Dimensions of electronic health record data quality assessment: enabling reuse for clinical research," Journal of the American Medical Informatics Association, vol. 20, no. 1, pp. 144-151, 2013.

[2] M. Kahn, T. Callahan, J. Barnard, A. Bauck, J. Brown, B. Davidson, H. Estiri, C. Goerg, E. Holve, S. Johnson, S. Liaw, M. Hamilton-Lopez, D. Meeker, T. Ong, P.

Ryan, N. Shang, N. Weiskopf, C. Weng, M. Zozus and L. Schilling, "A harmonised data quality assessment terminology and framework for the secondary use of Electronic Health Record Data," eGEMs (Generating Evidence & Methods to improve patient outcomes), vol. 4, no. 1, 2016.

[3] McKinsey Global Institute, "Big Data: The next frontier for innovation, competition and productivity," 2011. [Online]. Available: http://www.mckinsey.com/business-functions/digital- mckinsey/our-insights/big-data-the-next-frontier-for- innovation. [Accessed 19 July 2017].

[4] The Parliamentary Office of Science and Technology, "Big Data and Public Health (POSTNOTE 474)," 19 July 2017. [Online]. Available: http://researchbriefings.parliament.uk/ResearchBriefin g/Summary/POST-PN-474.

[5] M. Wilson, N. Dumontier, I. Jan Aalbersberg, G. Appleton, M. Axton, A. B. N. Baak and B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, pp. 1- 9, 2016.