MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-310

Analysis of Hospital Processes with Process Mining Techniques

Arturo Orellana García[a], Damián Pérez Alfonso[b], Osvaldo Ulises Larrea Armenteros[c]

[a] *Department of Applications, Center of Medical Informatics, University of Informatics Sciences, Havana, Cuba*
[b] *Professor at the faculty 3, University of Informatics Sciences, Havana, Cuba*
[c] *Engineering student, University of Informatics Sciences, Havana, Cuba*

Abstract

Process mining allows for discovery, monitoring, and improving processes identified in information systems from their event logs. In hospital environments, process analysis has been a crucial factor for cost reduction, control and proper use of resources, better patient care, and achieving service excellence. This paper presents a new component for event logs generation in the Hospital Information System or HIS, developed at University of Informatics Sciences. The event logs obtained are used for analysis of hospital processes with process mining techniques. The proposed solution intends to achieve the generation of event logs in the system with high quality. The performed analyses allowed for redefining functions in the system and proposed proper flow of information. The study exposed the need to incorporate process mining techniques in hospital systems to analyze the processes execution. Moreover, we illustrate its application for making clinical and administrative decisions for the management of hospital activities.

Keywords:

Hospital environment; Process mining; Event logs; Hospital Information System; Process model.

Introduction

Demand for medical services at an international level is increasing. The costs related to search and implementation of strategies and tools to meet the medical needs of populations push for an improvement in the execution of clinical and administrative tasks in health institutions. In order to address this situation, hospitals need to find alternatives that make their processes more efficient. The improvement in quality of processes implies higher productivity [1], so the design of new methods to optimize performance of processes is necessary.

The business process management approach allows for identification and management of processes that are interrelated; in order to analyze organizational performance and make continuous improvements of the results, eliminating errors, and redundant processes in the institution. Process mining allows for understanding how processes are really executed in the system. Its application helps to identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures, and simplify processes [2] for improving business performance. This technology allows getting information about how the process has behaved, the times it takes in the executions, as well as variations between reality execution and the prescribed model. The complexity of this approach is proportional to the amount of data stored; making it invaluable information for decision-making. These data are usually stored in the databases of the systems that support organizational processes. From these databases, certain information could be extracted to create event logs.

An event log is the evidence of events that occur in the computer systems and the networks of an organization [3]. The records are composed of input events. Each entry contains related information to an action or specific operation that was executed on the information system and its corresponding timestamp. These event logs should be formatted as MXML [6] or XES [7].

Currently, to obtain event logs from information systems, it is necessary to use specialized tools like: XESame [4] and Nitro [5]. Even though these software are efficient, their high complexity related to technical configurations and database structures make them inconvenient for inexperienced or non-expert users.

With the aim to computerize, the processes of secondary healthcare level have been developed into a Hospital Information System (HIS) at the University of Informatics Sciences (UCI), in the Center for Medical Informatics (CESIM). This HIS supports the fundamental processes carried out at this level of healthcare and was designed to meet the needs of: storage, processing, gathering, and interpretation of the medical-administrative data generated.

The HIS from UCI does not have a component to generate event logs from the execution traces stored in the system; thus precluding application of process mining techniques for analysis of running processes. Medical and administrative staff have limited access to this source of knowledge for clinical and administrative decision-making.

Due to the above inconveniences, we propose to develop a software component that allows generation of event logs from the Hospital Information System without relying on other external tools. This system component generates event logs with the same characteristics, format, and structure as obtained with XESame tool; but can be used by non-expert users. This solution allows analysis of healthcare process using process mining techniques.

This article has two main sections. The first deals with implementation of a software component to extract traces from the Hospital Information System. The second section focuses on process analysis with process mining techniques, starting from an event log generated using the developed component.

Materials and Methods

Processes mining in a hospital environment

An intelligent approach to healthcare is use of information to create knowledge about patient care and improve performance of the hospital institution. Healthcare organizations increasingly have more data, which involves a large amount of information and few alternatives for analysis; a result of the lack of effective tools or methods that allow for obtaining the decision criteria about execution of the system processes.

The process mining techniques allow extraction of useful and nontrivial information from execution logs stored in information systems [8]. There are three types of process mining techniques: discovery, conformance checking, and enhancement or extension. All these techniques allow us to analyze actual execution of processes. The discovery technique is used to extract process models from an event log. Conformance checking techniques are used to monitor deviations by comparing the model and the event log and improving existing models from its extension [9].

Investigations carried out provided evidence that process mining is a novel and effective technology for analysis of hospital processes. In Gynecology and Oncology [10], process mining techniques helped to optimize, from dotted points, the trajectory of patients through the process of care. Also, discovering the busiest times in the area of Emergency [11] allowed for control and allocation of resources to this hospital sector. Other research in the area of Emergency [12] presented the detection of eventualities (incomplete assignments, missing information, little correspondence between the business process and system) in the process activities using process mining techniques. Its application in the Electronic History Records [13] allowed for enhancement of patient care processes.

Maruster [14] proposed the application of Petri Nets based on Process Mining for modeling hospital information. Moreover, the application of process mining on hospital systems allowed monitoring use of resources such as surgical implements, for nursing and external consultation. Process mining applications in other environments [15-18] demonstrate its effectiveness in detecting malfunctions, such as deviation of resources, extra stay time of a product in a warehouse stock or loss of any product, in such a way that allows for adjustment of product ordering frequency. From the engineering point of view, programming errors in the processes definition and activities that do not run properly or are simplified could be identified. Process mining techniques allows finding bottlenecks in information flow.

Event Log

To formalize the structure of event logs used in process mining, two standards were defined: Mining eXtensible Markup Language (MXML) and eXtensible Event Stream (XES). MXML is a format based on XML for the exchange of event logs. MXML was the first standard that emerged in 2003 and was adopted by the process mining tool ProM [19].

MXML establishes a standard notation to store dates, resources and types of transactions. In 2010, XES replaced MXML as the new process mining format independent of the tool [17]. XES is a standard based on practical experiences of MXML, and is less restrictive and truly extensible. Its main purpose is to provide an interchange format for event logs between tools and application domains [7].

To ensure a successful analysis of process mining, the quality of the storage format of the event log should be ensured. This is defined from three fundamental aspects: reliability, completeness, and security. Reliability stands for that ability to safely assume that the recorded events really occurred and that the attributes of the events are correct. Completeness is the characteristic of not missing any event in a certain context. In addition, any registered event must have well-defined semantics. The event data are considered safe when considerations of privacy and security are taken on, when events are registered [18].

Attributes of Event Log

An event log contains evidence of processes execution stored like traces. This record of elements can also contain attributes. As the registration of elements is created only once, the impact of including many attributes in the registry is minimal [6]. It is important to include relevant information describing the contents and origin of the event log. The following attributes should be taken into consideration for trace recording:

- **Process Name**: The name of the process under which the log will record its execution.
- **Data Source**: A description of the information system from which the event log is extracted from.
- **Source Organization**: The name of the organization providing the data.
- **Description**: A brief description of the contents of the event log.
- **Version**: An identifier to differentiate versions of event logs.
- **Author**: Name and contact details of who defined the conversion.
- **Process Mining Project**: A reference to the Process Mining Project or purpose of the event log.

Development environment

We proceeded with the implementation of the software component from the theoretical study of event logs. The following technologies and tools were used:

Eclipse Ganymede 3.4.2: Eclipse is an IDE (Integrated Development Environment) open and multiplatform code that has achieved a high degree of maturity in the development of what are known as "rich client applications". It has tools to develop console applications and web services with different application servers such as JBoss, Websphere, and Glassfish. It was originally developed by IBM (International Business Machines) and its future is now in the hands of Eclipse Foundation, an independent nonprofit organization that fosters an open source community and a set of complementary products, capabilities and services.

JBoss Application Server 4.2.2: JBoss Application Server is a Java EE (Java Enterprise Edition) free software implemented in pure Java. Being based in Java, it can be used on any operating system that supports it. It provides a full range of services for Java EE 5 and expansion of business services, including clustering, caching, and persistence. JBoss is ideal for Java applications and applications based on the web. It also supports EJB 3.0 (Enterprise Java), making applications development much simpler.

JBoss Seam V2.1.1: JBoss Seam is a powerful framework for developing Web 2.0 applications by unifying and integrating technologies such as AJAX, JSF, EJB, Java Portlets and BPM (Business Process Management). Another important feature is that POJOs (Plain Old Java Objects) can be validated in addition to directly handling the application logic and business sessions from beans.

Java V1.6: Java is an object oriented programming language developed by Sun Microsystems in the early 1990s. The language itself borrows much of its syntax from C and C ++ but has a simpler object model and eliminates tools of low level, which can cause many errors such as direct manipulation of pointers or memory.

Hibernate V3.3: Hibernate is an ORM (Object Relational Mapping) tool for Java platform. It facilitates the mapping of attributes between a traditional relational database and the object model of an application using declarative XML files or annotations, on the beans of entities that allow for establishing these relationships.

Facelets V1.1: Facelets is a simplified presentation framework where it is possible to freely design a web page and then associate specific JSF components. It provides more freedom to the designer and improves error reporting, having JSF. It allows for definition of disposition of pages based on a template, the composition of components, creating custom labels, and facilitates friendly development for the graphic designer and creation of component libraries.

Results

Figure 1 shows the user interface for the component developed to generate event logs in XES format. This component requires a process name as input, specified by the user to identify the process. It is needed to select the process to be analyzed and delimits start and end date for the events that will be inputted in the event log.

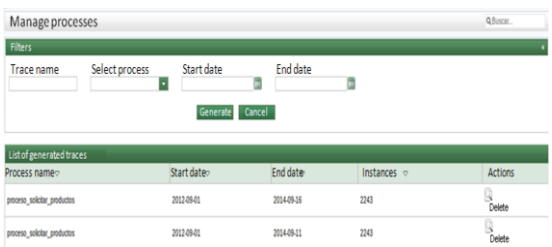

Figure 1 - Component to generate an event log in the HIS.

Below is an excerpt of the event log (see Figure. 2) obtained from the component developed. The extensions and attributes mentioned above are reflected in this image.

```
<trace>
        <string key="concept:name" value="4364"/>
        <string key="description" value="Simulated process instance"/>
        <event>
                <string key="org:resource" value="cirujano"/>
                <date key="time:timestamp" value="2012-01-15T16:53:24.491.000+01:00"/>
                <string key="concept:name" value="ver_detalles_sol_bq"/>
                <string key="lifecycle:transition" value="complete"/>
        </event>
        <event>
                <string key="org:resource" value="administrador"/>
                <date key="time:timestamp" value="2012-01-15T16:53:35.813.000+01:00"/>
                <string key="concept:name" value="despacho_sol_bq"/>
                <string key="lifecycle:transition" value="complete"/>
        </event>
</trace>
```

Figure 2 – Event log excerpt.

An event log related to the process: Request Product of HIS Warehouse module, was obtained from the developed component. The Warehouse module, among other activities, manages the flow of information on the various movements that a product can have in a warehouse. There are three types of product requests: tendering request, application warehouse, and request for surgical block (see Figure 3). To show the usefulness of the developed component, the Request Product process from Warehouse Module of the Hospital Information System created at UCI, was selected. An event log related to this process was extracted and analyzed with the ProM tool as shown below.

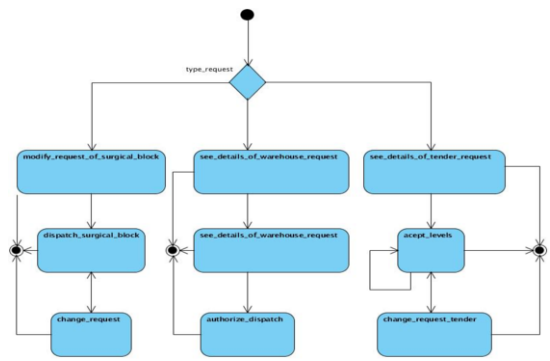

Figure 3 – Business Process Diagram of Request Product.

This event log was loaded on the ProM tool in order to conduct the process mining analysis. The process model showed on Figure 4 was obtained by applying the technique “Mine Net using Heuristic Miner”. This process model is a representation of the actual process behavior recorded in the event log during its executions on the HIS.

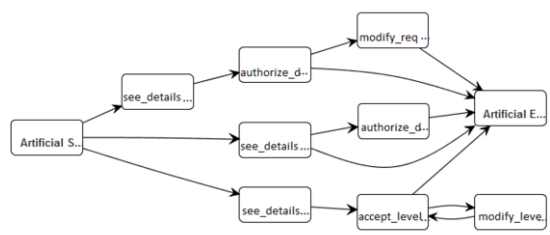

Figure 4 – Model obtained by applying the technique Mine for a Heuristic Net using Heuristic Miner.

This technique shows the similarities between the formal definition of the process and the event log obtained from the traces of their execution. In this model, there exists a difference with respect to the reference model; the process never ends directly after see_details_tender_request activity and whether it should end.

In comparing Figures 3 and 4, the applied technique shows that the process execution in the system is in concordance with the formal definition of the business process. The system has been implemented in the way it its business was conceived.

Figure 5 illustrates the observed result of the technique: "Analyze using Dotted Chart". The Dotted Chart shows an overview of the process where the X-axis shows the time, the Y-

axis shows the process execution instances and each point represents an event; the colors of the points refer to the different activities.

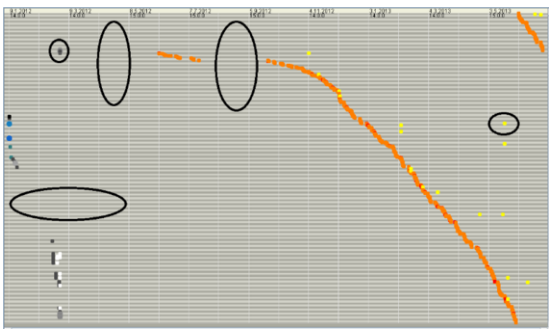

Figure 5 – Aplication of technique "Analize using Dotted Chart" whit black circles representing noise.

The runtime process was extended over seventeen months and three distinct moments with different behaviors could be distinguished. The first moment corresponds to system testing stage, where see_details_of_warehouse_request and authorize_dispatch activities were not executed. Afterwards, the system does not record any execution of the process for three months. The second and third phase were extended a month and a half to nine and a half months, respectively, with a downtime of two months between them; only see_details_of_warehouse_request and authorize_dispatch activities were running. Because of these characteristics, this process shows an anomalous behavior as 6 of the 8 activities making up the final version of the process are left to run.

The next model (Figure 6) represent the technique Replay a log on Petri net for Performance/Conformance, that allows some cases to define in a simple way, for example, the more common activities flow that runs in the system process under analysis.

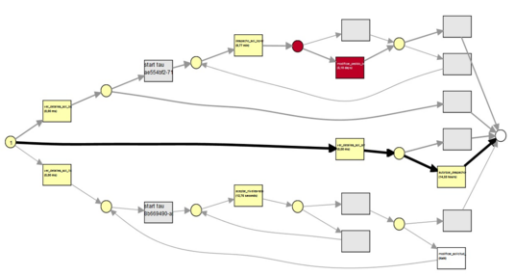

Figure 6 – Aplication of technique "Replay a log on Petri net for Performance/Conformance".

This technique establishes that while darker and thicker lines in the figure indicate a transition, a greater amount of times have been carried out for the activity represented in the flow.

This model allows identification of the flow of the more executed activities in the processes under analysis. Therefore, in the process Product Request, the flow with more executed activities can be defined as composed by the activities see_details_of_warehouse_request and authorize_dispatch.

In comparing the difference of maximum and minimum waiting times between modify_request_of_surgical_block and authorize_dispatch activities, we see that:

- The minimum waiting time of the first activity (3.29 seconds) is less than the second one (4.43 minutes).
- The timeout of the first activity (5 months) is greater than the second activity (7.45 days).
- The average time of the first activity (14.38 hours) is less than the second activity (3.31 days).

These analyses allowed receipt and dispatch of products from the warehouse, as well as detection of a bottleneck in the activity authorize_dispatch to be controlled. The understanding and correct interpretation of the process models obtained from process mining techniques have been discussed in the literature, as challenges [19] to analyze the behavior of computer systems by non-experts in the field of knowledge are the same.

Conclusions and future works

Process mining is presented as an effective alternative to show a radiograph of the actual process execution. However its models are far from favorable to non-experts.

The application of process mining techniques in the hospital environment is novel because there are only a few cases of its study around the world. However there is support for its use in this key social sector.

We developed a component for extracting traces of the HIS, which overcomes the problems raised above. From the analysis of the traces generated by the developed component using process mining techniques, we concluded that they contain valuable information useful for process analysis.

The solution developed guarantees that data persist permanently in the trace processes records. As a part of the developed component, it ensures transformation of stored records in the HIS to processes event logs. It also allows to export the to XES format.

After applying multiple modeling techniques to the event log obtained, was stated that the obtained models are not suitable for understanding by health personnel. The incorporation of process mining techniques to the HIS where models are generated, allowing for analysis by non-experts, is plotted as a goal of this research. Towards this goal, we conduct research on what elements are essential to the analysis and understanding of process execution.

The component developed is a starting point for incorporation of process mining techniques and algorithms of the HIS, in order to support clinical and administrative decision-making. This component becomes the first of its type, designed to analyze processes in the systems HIS.

Understanding the process model (in terms of accuracy and efficiency of understanding) is a function of the characteristics of the model and the characteristics of the user who interprets the model [20]. The visual design of the model is a primary factor analysis. Recker has conducted research on how process models can be designed to maximize their understanding [21].

References

[1] Deming WE. Calidad, productividad y competitividad: La salida de la crisis. España: Díaz de Santos S.A. 1989.

[2] van der Aalst WMP. Process Mining. Discovery, Conformance and Enhancement of Business Processes. London New York: Springer, 2011.

[3] Verbeek HMW, Buijs JCAM, van Dongen BF, van der Aalst WMP. XES, XESame, and ProM 6. In: Soffer P, Proper E, editors. CAiSE Forum 2010, LNBIP, volume 72; Berlin, Heidelberg. Springer-Verlag; 2010. pp. 60-72.

[4] XESame. (n.d.). Extraído el 20 de abril de 2014 desde http://www.processmining.org. 2014.

[5] Nitro. (n.d.). Extraído el 24 de abril de 2014 desde http://www.fluxicon.org/. 2014

Buijs JCAM. Mapping data sources to XES in a generic way [online]. Master Thesis. Eindhoven, The Netherlands: Technische Universiteit Eindhoven. 2010. [Accedido 16 marzo 2012]. Disponible desde: http://www.processmining.org/_media/xesame/xesma_thesis_final.pdf

[6] Gunter CW. XES Extensible Event Stream standard definition. S.l.: s.n. 2009.

[7] Yzquierdo R. Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso. Tesis de Doctorado, Universidad de las Ciencias Informáticas, Cuba. 2012.

[8] van der Aalst WMP, Weijters T, Maruster L. Workflow mining: discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering; IEEE. 2004;16(9):1128-1142.

[9] Mans RS, Schonenberg MH, Song M, van der Aalst WMP, Bakker PJM. Application of process mining in healthcare: a case study in a dutch hospital. Eindhoven University of Technology, The Netherlands. Berlin, Heidelberg. Springer; 2009. pp. 425-438.

[10]Mans RS, Schonenberg MH, Song M, van der Aalst, WMP, Bakker PJM. Process mining in healthcare: a case study. Eindhoven University of Technology, Netherlands. 2008.

[11]Orellana A, Sánchez Y. Minería de Procesos en salud. Caso de Estudio: modelado de los procesos del área de Emergencia. Twelfth LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014) "Excellence in Engineering To Enhance a Country's Productivity".PRF#254. 2014.

[12]Webster C. EHR BPM: from process mining to process improvement to process usability. Trabajo presentado en EHR Workflow Inc, 2011.

[13]Maruster L, van der Aalst WMP, Weijters T, van den Bosch A, Daelemans W. Automated discovery of workflow models from hospital data. Trabajo publicado en The 13th Belgium-Netherlands Conference on Artificial Intelligence; 2001. pp. 183-190.

[14]Reijers HA, van der Aalst WMP, Weijters AJMM, van Dongen BF, De Medeiros AKA, Song M, Verbeek HMW. Business process mining: an industrial application. Information Systems. 2007;32(5):713-732.

[15]Hernández PM. Aplicación de técnicas de minería de proceso para el control y mejoramiento del proceso de compras nacionales e internacionales de bienes y servicios para proyectos de investigación de la Pontificia Universidad Javeriana. Tesis de Maestría. Pontificia Universidad Javeriana, Facultad de Ingeniería Industrial, Bogotá, 2012.

[16]van der Heijden THC. Process mining project methodology: developing a general approach to apply process mining in practice [online]. Master of Science in Operations Management and Logistics. Netherlands: TUE. School of Industrial Engineering. 2012. [Accedido 20 octubre 2012]. Disponible desde: http://alexandria.tue.nl/extra2/afstversl/tm/Van_der_Heijden_2012.pdf

[17]Rozinat A, van der Aalst WMP. Conformance testing: measuring the fit and appropriateness of event logs and process models. In: Business Process Management Workshops. Berlin, Heidelberg. Springer; 2012. pp. 163-176.

[18]IEEE TASK FORCE ON PROCESS MINING. Manifiesto sobre Minería de proceso [online]. S.l.: s.n. 2011. Disponible desde: http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:pmm-spanish-v1.pdf

[19]Mendling J, Strembeck M, Recker J. Factors of process model comprehension: findings from a series of experiments. Decision Support Systems. 2012;53(1):195–206.

[20]Recker J, Reijers HA, van de Wouw SG. Process model comprehension: the effects of cognitive abilities, learning style, and strategy. Commun Assoc Inf Syst. 2014;34(9):199-222.

Address for correspondence

Ing. Arturo Orellana García, email aorellana@uci.cu, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370.