Measuring and Monitoring the Rehabilitation of Patients on Monitoring Stations via the Analyses of poly-structured Data

Mirco Josefiok

Data Management and Data Analysis
OFFIS Institute for Information Technology
Escherweg 2
26121 Oldenburg
mirco.josefiok@offis.de

Supervisor: Prof. Dr. Dr. h.c. H.-Jürgen Appelrath

Abstract: In this paper an approach for measuring an monitoring the rehabilitation of patients on monitoring stations via the combined analyses of structured and unstructured data is presented. Here, the focus is on stroke patients in different rehabilitation phases. For clarification interviews with IT managers from hospitals and with parts of the medical staff were carried. Based on the information from the interviews a first theoretical and a prototypical approach is presented. An outlook towards the possible evaluation and further work is given.

1 Introduction

In this paper a method with an accompanying software application for measuring and monitoring a patients rehabilitation on monitoring stations will be presented. On monitoring stations lots of documentation is handwritten. For a retrospective assessment, an analyses in combination with structured data is necassary. Whilst the analyses of structured data is handled well and also analyzing unstructured data is possible a combination is not yet possible to the extent necessary [BFG$^+$14]. Other approaches do not combine structured and unstructured data in a unified analyses process, but extend the structured database with information gathered from unstructured data. Although approaches for handling huge amounts of unstructured and structured exist in other domains none is is readily applicable to the health care sector [CCS12].

The stroke is one of the most common diseases in Germany. About six billion Euros are spent each year for the treatment of the most frequent types[1]. Most acute stroke patients are initially treated in so called stroke units which are specialized wards for taking care of stroke patients. Patients who suffer from a severe stroke and fall into a coma are treated in

---

[1]see diagnosis code I60, I61, I63, I65 at `www.gbe-bund.de` - last visited 2014-04-14

different phases, labeled from A to F [USG10]. Phase A describes the acute phase, phase B the early stages of rehabilitation, phase C the early mobilization, phase D the rehabilitation, phase E the outpatient follow-up and phase F the integration and participation. It is possible for each phase to be carried out by a different institution. Just in one phase a huge amount of unstructured date is created and collected for individual patients. For example, a patient can stay in phase B for a very long time depending on the process of rehabilitation. Whilst in this phase, very fine grained information about the patients condition is documented by different members of the medical staff, e.g. doctors, therapists and the nursing staff, which is normally done handwriting. No standards exist yet for this documentation. Therefore each institution handles this process using its own standards [HBB+06]. Documentation is either persisted as a more or less structured text document or the handwritten form is scanned. Structured data, which are generally present are routine records, accounting data and sensor data. Routine records and accounting data are usually used as a foundation for analysis and reports. But the less structured documentation offers valuable insight about the rehabilitation of patients [STM+11].

This scenario raises the question of how to enable a continuous, combined analysis of poly-structured data for measuring and monitoring the progress of patients rehabilitation.

2 Determine the available Information and subsequent Questions

Expert interviews were conducted in order to determine the available information, the available structured and unstructured data and to discuss possible approaches. In the first round, IT managers from six hospitals were interviewed. Interviews with head and senior physicians were conducted in a second round. The interviews lasted between one and two hours and were carried out by one interviewer. Following the interviews, one hospital offered to provide sample data for one patient from a monitoring station. The sample data consists of structured and unstructured data. The structured data is extracted from the clinical information system whilst the unstructured data is available partly as MS Word documents and partly as scanned images.

IT managers were chosen as interview partners for a first round of interviews because they have excellent knowledge about the domain and a technical background. They can hopefully provide immediate feedback about the possibility and feasibility of the presented approach here. Topics discussed during the interviews with the IT managers included:

- General information about the organization (e.g. technical equipment, number of employees)
- Systems and data (e.g. software used for reports and analysis, collected and specially prepared data, collection for unstructured data)
- Questions (e.g. typical questions for analysis, boundaries of present solutions)
- Requirements (e.g. requirements for the processing of unstructured data, analysis which should be possible)

One of the results of the interviews is that the IT landscapes of most hospitals are comparable. There is always a settlement system and a clinical information system. All other

systems are usually connected to those two, either directly or via a communication server. This is a result of the continuous efforts to integrate clinical information system more tightly [HW08]. All hospitals provide a software solution to conduct analysis and reports. These differ in terms of their approach of gathering the necessary data. One approach is to gather all relevant data directly from all systems. An alternative approach is to gather all data from the settlement and clinical information systems. Typical questions regarding method and software solutions for a combined analyses of structured and unstructured data included:

- How does one improve the process and persistance of unstructured data?
- Is it possible to offer a simple software solution for searching unstructured data?
- For which key user is it necessary to provide the possibility to work with unstructured data?

More detailed interviews with the medical staff are currently being prepared and conducted. In a first round questions regarding the actual documentation process, shortcomings of present mechanisms, scientific research questions and requirements that a method should meet were discussed, with a focus monitoring stations. As it turned out, there exists a multitude of unstructured data in monitoring stations. Questions which rose during the interviews included:

- How would it be possible to directly monitor and analyze the success of a patients treatment?
- How can the factors for a success of a patient's treatment be identified?
- How can patients be clustered by different and changing characteristics?

An continuous, combined analyses of structured and unstructured data for improving the care situation of individual patients was considered highly desirable by all interview partners. For a first assessment of the discussed possibilities, one hospital provided sample data. The data provided is typical for patients from monitoring stations. It consists of a patients core data like name, address, birthday etc. and all case based information. Different cases as well as the patient are identified by a unique number. The structured data is gathered from the settlement and the clinical information system and is prepared for being persisted in a MS SQL database with MS Analysis Services. The data can be accessed with most Data Warehouse solutions. The MS Word and image files are structured on the file system. File names contain the patient and case numbers. The individual documents are organized in sub folders for different aspects of the treatment (e.g. medical history, diagnosis, documentation, misc etc.). A corresponding XML file exists which contains meta data (e.g. core data, case data, patient information etc.). Whilst most Word documents have a repeating structure depending on their use case, the images contain scanned, hand written documents.

3 Developing an Approach for Measuring and Monitoring the Rehabilitation of Patients on Monitoring Stations

For a recurring analysis of great amounts of data with an increasing complexity of the analysis Data Warehouses in combination with Online Analytical Processing (OLAP) have been established. Analytical Information Systems (AIS) are, in differentiation to operative transaction-oriented systems, systems which support analytical and planning activities. It is not possible to integrate unstructured data into a DWH without a great deal of effort and a considerable chance of failure. In addition, it is not possible with the given technical, organizational, economic and legal circumstances [CG10].

Most institutions run their own systems for to conduct analysis and to create reports. But those systems are limited in their flexibility and extendibility. In some cases they even offer just predefined solutions for creating reports. The work is structured by the design science research process. According to Pfaff et al. the design science research process consists of six stages, which are problem identification & motivation, objectives of solution, design & development, demonstration, evaluation and communication [PTG06]. A problem identification and motivation was given in section 1, objectives of a solution were determined through interviews with experts and potential key users, although not yet concluded, gives reasonable grounds for the desired objectives. At this stage a conceptual and a prototypical approach were developed, which will both be presented in the upcoming sections.

3.1 Conceptual Approach

Based on the findings of the interviews, various technologies were examined which could support the method and the questions obtained. With MUSTANG (Multidimensional Statistical Data Analysis Engine) a software exists which is in use in various institutions in the health care sector. It is used for conducting exploratory analysis. Whilst MUSTANG offers flexible analysis it belongs to the class of AIS and only works with an underlying DWH. Therefore a direct analysis of unstructured data is not possible [KAS13]. Hence, another tool which is capable of doing so is needed. With Apache Solr exists an extendable and adaptive solution for realizing salable search applications. Apache Solr is based on Apache Lucence for the search component and Apache Tika for handling unstructured data. The downside is, that each document has to be added to Solr's index. Following that preliminary consideration an approach consisting of four elements was developed.

Harmonization of given data sources: For being able to work with heterogeneous data in a first it is necessary to harmonize the given sources as far as possible. Thus, create a unified scheme for the structured data. Together with potential stakeholders it has to be decided which data is absolutely necessary for answering the various questions raised.

Implementation of the analysis of unstructured data: It is not possible to harmonize unstructured data in the way as structured data for integrating them into an AIS.

Therefore the application of data mining technologies is necessary to extract the required information. Moreover it is important that the unstructured data can be searched.

Development of a suitable process: Depending on the systems and sources from which data should be brought together, a process must be developed. In this process must be specified which an how the data is brought together and how it should be prepared for making it available to the analysis software.

Implementation of the necessary software applications: For conducting the analysis the development and implementation of a portfolio of software applications is necessary. It is important to ensure that new data is automatically added to the data basis. Furthermore potential end users must be able to conduct analysis on their own.

3.2 Prototypical Approach

A first and very simple prototype was developed to evaluate the technical feasibility of the discussed ideas. It was applied to the sample data and the results were discussed with potential stakeholders. The prototype consists of a MUSTANG and a Apache Solr installation. The analysis of the structured data was handled by MUSTANG. Keywords from the results were then manually added to Solr's search query. It was now possible to search and find suitable documents by their meta data. Moreover it was possible to do a full text search on all MS Word documents. A correlation between the result sets has to be made manually. At this stage, the ETL process for the data warehouse is done by the hospital and the copying process for the unstructured data has to be carried out manually. In the further development process it is planned to adjust the ETL process and implement an automated procedure for deploying the unstructured data.

As a result of this first test with the provided sample data it can be concluded that the combination of an AIS with a platform for search and retrieval of information in unstructured data is not only possible but very promising.

4 Planned Evaluation and Outlook

The involvement of potential key users and stakeholders of a solution for a combined analysis of structured and unstructured data ensures that the conceptual work and the development of the prototype are valuable and appropriate. Moreover important question formulations can only be carved out in co-operation with key users. Although the work is in a very early stage, feedback on the conceptual and a prototypical approach was very positive. For a next step more and more detailed interviews with key users are planned. It is necessary to clarify further the requirements and objectives for the method and the accompanying software solution. As for the prototype further tests are needed. In addition to this date it is uncertain how to proper integrate the AIS with the search and retrieval platform. Approaches which will be explored include integration on service level and

integration on GUI level.

For demonstrating method and prototype it is planned to use them in cooperation with an hospital to measure and monitor rehabilitation of patients from a monitoring station. For this purpose, method and prototype must be kept in development. It is planned to involve the medical staff in the whole process. Two approaches are intended for evaluation of the presented approach. On the one hand, extensive expert interviews will he held. On the other hand, the prototype will be used for retrospective analysis. The result will be compared with existing ones.

References

[BFG⁺14]   Henning Baars, Carsten Felden, Peter Gluchowski, Andreas Hilbert, Hans-Georg Kemper, and Sebastian Olbrich. Shaping the Next Incarnation of Business Intelligence. *Business & Information Systems Engineering*, 6(1):11–16, January 2014.

[CCS12]   Hsinchun Chen, Roger H. L. Chiang, and Veda C Storey. BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly*, 36:1–24, 2012.

[CG10]   P. Chamoni and P. Gluchowski. Analytische Informationssysteme - Einordnung und Überblick. In *Analytische Informationssysteme : Business Intelligence-Technologien und -Anwendungen*, pages 3–16. 2010.

[HBB⁺06]   Peter U Heuschmann, Marcel K Biegler, Otto Busse, Susanne Elsner, Armin Grau, Uwe Hasenbein, Peter Hermanek, Rudolf W C Janzen, Peter L Kolominsky-Rabas, Klaus Kraywinkel, and Others. Development and Implementation of Evidence-Based Indicators for Measuring Quality of Acute Stroke Care The Quality Indicator Board of the German Stroke Registers Study Group (ADSR). *Stroke*, 37(10):2551–2573, 2006.

[HW08]   Rada Hussein and Alfred Winter. Towards More Integrated Implementation of Healthcare Information Systems: Using the 3LGM2 for Modeling the IHE-Scheduled Workflow Integration Profile. *2008 21st IEEE International Symposium on Computer-Based Medical Systems*, pages 650–652, June 2008.

[KAS13]   David Korfkamp, Hans-Jürgen Appelrath, and Eunice Sirri. CARESS als integrierte Auswertungssoftware zur Berechnung von {Ü}berlebenszeitanalysen. In *GMDS 2013: 58. Jahrestagung der Deutschen Gesellschaft f{ü}r Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)*. Deutsche Gesellschaft f{ü}r Medizinische Informatik, Biometrie und Epidemiologie, German Medical Science GMS Publishing House; D{ü}sseldorf, 2013.

[PTG06]   K Peffers, T Tuunanen, and CE Gengler. The design science research process: a model for producing and presenting information systems research. In *DESRIST*, pages 83–106, Claremont, CA, 2006.

[STM⁺11]   E. Swart, D. Thomas, S. March, T. Salomon, and O. Knesebeck. Die Eignung von Sekundärdaten zur Evaluation eines Interventionsprojekts. *Prävention und Gesundheitsförderung*, 6(4):305–311, September 2011.

[USG10]   Lothar Ullrich, Gabriele Stolecki, and Matthias Grünewald. *Intensivpflege und Anästhesie*. Thieme, 2., neu berarbeitete und erweiterte auflage edition, 2010.