Journal of Biomedical Informatics 69 (2017) 188–202

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Data summarization method for chronic disease tracking

Dejan Aleksić [a], Petar Rajković [b,*], Dušan Vučković [b], Dragan Janković [b], Aleksandar Milenković [b]

[a] University of Nis, Faculty of Science and Mathematics, Department of Physics, P.O. Box 224, 33 Visegradska, 18000 Nis, Serbia
[b] University of Nis, Faculty of Electronic Engineering, Laboratory for Medical Informatics, 14 Aleksandra Medvedeva, 18000 Nis, Serbia

ARTICLE INFO

Article history:
Received 23 January 2017
Revised 10 March 2017
Accepted 17 April 2017
Available online 19 April 2017

Keywords:
Medical information system
History of disease summarization
Chronic disease management
System usage scenarios

ABSTRACT

Objectives: Bearing in mind the rising prevalence of chronic medical conditions, the chronic disease management is one of the key features required by medical information systems used in primary healthcare. Our research group paid a particular attention to this specific area by offering a set of custom data collection forms and reports in order to improve medical professionals' daily routine. The main idea was to provide an overview of history for chronic diseases, which, as it seems, had not been properly supported in previous administrative workflows. After five years of active use of medical information systems in more than 25 primary healthcare institutions, we were able to identify several scenarios that were often end-user-action dependent and could result in the data related to chronic diagnoses being loosely connected. An additional benefit would be a more effective identification of potentially new patients suffering from chronic diseases.

Methods: For this particular reason, we introduced an extension of the existing data structures and a summarizing method along with a specific tool that should help in connecting all the data related to a patient and a diagnosis. The summarization method was based on the principle of connecting all of the records pertaining to a specific diagnosis for the selected patient, and it was envisaged to work in both automatic and on-demand mode. The expected results were a more effective identification of new potential patients and a completion of the existing histories of diseases associated with chronic diagnoses.

Results: The current system usage analysis shows that a small number of doctors used functionalities specially designed for chronic diseases affecting less than 6% of the total population (around 11,500 out of more than 200,000 patients). In initial tests, the on-demand data summarization mode was applied in general practice and 89 out of 155 users identified more than 3000 new patients with a chronic disease over a three-month test period. During the tests, more than 100,000 medical documents were paired up with the existing histories of diseases. Furthermore, a significant number of physicians that accepted the standard history of disease helped with the identification of the additional 22% of the population. Applying the automatic summarization would help identify all patients with at least one record related to the diagnosis usually marked as chronic, but ultimately, this data had to be filtered and medical professionals should have the final say. Depending on the data filter definition, the total percentage of newly discovered patients with a chronic disease is between 35% and 53%, as expected.

Conclusions: Although the medical practitioner should have the final say about any medical record changes, new, innovative methods which can help in the data summarization are welcome. In addition to being focused on the summarization in relation to the patient, or to the diagnosis, this proposed method and tool can be effectively used when the patient-diagnosis relation is not one-to-one but many-to-many. The proposed summarization principles were tested on a single type of the medical information system, but can easily be applied to other medical software packages, too. Depending on the existing data structure of the target system, as well as identified use cases, it is possible to extend the data and customize the proposed summarization method.

1. Introduction and motivation

Describing chronic medical conditions and documenting them properly is one of the key Electronic Health Record (EHR) functionalities enabling an effective primary care [1,2]. When the data have

* Corresponding author at: University of Nis, Faculty of Electronic Engineering, Laboratory for Medical Informatics, 14 Aleksandra Medvedeva, Lab 534, 18000 Nis, Serbia.
E-mail addresses: alexa@pmf.ni.ac.rs (D. Aleksić), petar.rajkovic@elfak.ni.ac.rs (P. Rajković), dusan.vuckovic@elfak.ni.ac.rs (D. Vučković), dragan.jankovic@elfak.ni.ac.rs (D. Janković), aleksandar.milenkovic@elfak.ni.ac.rs (A. Milenković).

http://dx.doi.org/10.1016/j.jbi.2017.04.012
1532-0464/

been collected from different sources, or by following different procedures, the data aggregation must be introduced [3] and accompanied by the adequate tracking tools. The ways of tracking chronic diseases constitute an important part of sustainable healthcare and this is classified as a long-term healthcare tracking [4]. As it is recommended in [5], HER-based system should collect data about various events related to chronic (and generally to any other) diseases and allow users to continuously track the health status of the patient. In addition, different trending tools and summarizing methods can be developed to ease medical professionals' operations and allow a further data mining [6,7].

The data summarization is especially important when new patients suffering from chronic diseases need to be identified [8]. Its importance is further emphasized by the fact that, in many countries, the percentage of patients with chronic diseases is around one half of the total population [9]. This large number of patients makes chronic diseases an important social issue. From the medical point of view, such prevalence categorizes certain diseases as epidemic. A closer look reveals that one of the most frequent chronic diseases is hypertension. It affects 35–45% of the world population [9]. A similar situation is noted in the Republic of Serbia, where the most frequently diagnosed chronic disease is high blood pressure [10].

Over the past few years, after the introduction of the Medical Information System (MIS) to primary healthcare in the Republic of Serbia, opportunities for the chronic disease tracking have emerged [11]. Primary healthcare centers are organized on municipal level, and almost all of them have the MIS in place. Installed MISs are successfully used for obtaining data and reporting them to interested parties, such as the Ministry of Healthcare (MoH) and medical insurance funds (HMIFs). Once uploaded, data are processed and the national level statistics are generated. Also, medical institutions create their own reports and statistics.

From a medical practitioner's point of view, the use of the MIS in daily operations offers several functionalities that can help in the chronic disease tracking. The MIS users are informed about certain guidelines during their training that, if properly followed, produce very clear patient-level reports with abundant information. Unfortunately, not all users consider the mentioned guidelines important enough and consequently, several different data collection routines have been observed. These aberrations reduce the transparency of patient-level reports, and data are being scattered across different categories. Management-level reports that require extraction of full data summaries for a specific patient can then produce false positives. This is in line with the findings presented in [12].

As it has been stated in [12], the introduction of the EHR-based MIS into primary healthcare has its positive and negative effects. In this case, the positive side is the fact that the volume of the data related to chronic conditions are properly collected and registered. This contributes to the creation of a valid general statistics report from a medical institution. Now, an additional effort is needed to improve the data structure and upgrade personal-level reports.

The development group managed to pinpoint five distinctive routines that the MIS users had in the data entry process. Some of them were in line with general guidelines, but others were the product of the non-envisaged software use, generating the data that could not be easily linked to the main history of disease (HoD) for a specific diagnosis.

This is why an algorithm was defined to help with the data summarization and enable all existing aggregation tools to display a much cleaner data spectrum. The proposed method will not introduce any changes into the existing data, but rather build an appropriate data structure around it. In this way, no data will be changed or structural integrity corrupted, which should ensure the end-user's satisfaction with the quality of reports.

The data collected in the Nis Primary Healthcare Center were used for this research. Its MIS had more than 600 active users in the monitoring period, covering the population of more than 200,000 patients from the Nis metro area. The Medical Information Systems research group from the Faculty of Electronic Engineering, Nis, Serbia, had been actively involved in the MIS development and deployment since 2002. In 2008, the development of the EHR-based MIS [11] started. This was predominantly aimed at primary and ambulatory care centers [13]. From the beginning of 2016, the system has been in use by 31 healthcare institutions, with user bases ranging from 20 to 600 medical professionals per institution.

2. Related work

It is common knowledge that patients suffering from chronic diseases usually visit their doctor more often [14]. Since their data are collected from several departments, and since the MIS users have different data entering routines, there are instances where some of the data related to some chronic disease are not properly structured and not easily accessible. As it has been presented in [15], the data related to chronic diseases must contain an additional set of internal relations in order to increase the chances of being accepted by medical professionals.

It is important to mention that doctors spend a lot of time reviewing medical charts. According to [18], the time needed to do this varies from a few minutes to half an hour. When data are not properly structured and displayed, this time increases, thus raising the chances of overlooking some of the important facts. As it has been stated in [18] "properly structured data leads to a reduction in treatment time vs. increased unproductive documentation-gathering time". Although this conclusion is valid for nephrologists, it is a rather generally applicable statement. Therefore, we decided to extend the existing data structures by introducing a minimal set of entities and relations.

Consequently, the end-users will get one additional set of overviews and reports that will eventually help them when important decisions regarding the medication process should be made. Applying the summarization not only to a single patient, but rather to specific social groups, can trigger preventive actions and health campaigns. The importance of such campaigns is discussed in [16,17]. Notwithstanding the fact that the collected data can give meaningful results, the weak spot is the data collection based on questionnaires and surveys which can be "polluted" by subjective answers. This fact induced us to move on in a different direction. Our data summarization approach is based only on the analysis of the data stored in the MIS, and thus, the obtained results are expected to be more objective.

In the Primary Care Center Nis, we identified the doctors that effectively used the available features and those that checked patients' printed records and used the MIS to enter data and print one-page-long summaries after the visit. The data summarization proposed in the next section is intended to improve their daily routine. Considering the available reference materials, there are many articles describing different data summarization methods and tools often connected to chronic diseases.

Some state-of-the-art summarizing tools are presented in [19,20]. The tool presented in [19] named HARVEST is a Web tool basically designed as a table view. The tool [19] supports also the timeline style overview. The concepts of table-based and timeline overviews are also applied in our summarization tool. Unlike HARVEST, the summarization engine developed by our research group is part of the Windows-based rich client. The main summarization overview is also a table view, but supported by a multi-leveled, strong filtering support. Our timeline is, by default, focused on the main events, such as a therapy change. Thanks to that approach, our summarization tool does not have the data cloud