A Personalized Health Information System to foster Preventive Medicine

Sebastian Klenk, Julia Möhrmann, Andre Burkovski, Jürgen Dippon, Peter Fritz and Gunther Heidemann

***Abstract*—The first and foremost task of all health information systems is to inform the users about their current health level. Good systems give information on what action might change their status quo for the better. A excellent system would engage the user in these actions to improve their health in a sustainable way.**

In the course of this paper we will demonstrate how such a system might look like. Our main emphasis will be on its attainability with currently available data sources.

I. Introduction

> People are traveling in their cars along a highway, only to find that the road heads directly off a cliff. Not surprisingly, this creates a pileup at the cliffs bottom with all sorts of injuries and fatalities. So, where do you put the hospital?

This quote taken from Goetzs book "The decision tree" [5] illustrates the dilemma we have with our current heath care system: We treat people only after they fell of the cliff and do not prevent them from falling. In this paper we will present a system that is intended to make people think about performing a u-turn before they reach the cliff. We will start with a small example which will serve as a guidance through this paper.

Example 1: You, a health concerned user, decide to do something about your health. You consult your doctor and buy a lot of books and magazines. The information you get from your doctor is rather medical as well as detailed and you get the advice to exercise more and perform a healthy diet. The facts and information you get from books, magazines or health portals are also rather generic and in no way personalized. After a number of days with healthy food and irregular exercise your motivation drops. You neither see any progress nor does the abstract idea of better health allows you to further remain obedient.

What is clearly lacking in this example is the connection between you current situation, its outcome and the improved outcome after exercise and a better diet. The same example with feedback on the actions taken would result in a dramatically different outcome.

Example 2: You, a health concerned user, decide to do something about your health. You consult your doctor and buy a lot of books and magazines. The information you get is very specific to your current situation. You get detailed information on how more exercise and a healthier diet improve your health and reduce the risk of fatal diseases. After each of your exercise you can see how you life expectance changes and after each healthy meal you see the risk reduction for different diseases. This kind of continuous feedback keeps you on track for the next few years.

Goetz [5] argues in his book that personal monitoring and direct feedback allows for more conscious decisions. In the next few sections we will propose a system that can support such a process. It monitors personal lifestyle data, compares the data with epidemiological data, estimates probable outcomes and proposes alternatives.

Sebastian Klenk, Julia Möhrmann, Andre Burkovski, Jürgen Dippon and Gunther Heidemann are with the University of Stuttgart, Stuttgart, 70569, Germany (phone: +49 (0)711 685 88-241; email: klenksn@vis.uni-stuttgart.de).

Peter Fritz is with the Institut für Digitale Medizin, Stuttgart, 70192, Germany

II. The data

Changing peoples behavior requires them to be knowledgeable about their current actions and what consequences of these actions are. Further it is necessary to demonstrate how a change in action positively influences their health. At first sight, this seems to be rather trivial and the, therefore necessary, data easy to obtain. Basically we would need data on

1) the current health condition, the current lifestyle and
2) on the expected progress or decline of the persons health.

But obtaining objective, quantitative data on peoples lifestyle is still a question of active research. How do we measure the average stress level, or the overall amount of exercises people actually perform? How to measure peoples diet, alcohol or smoking habits? For obvious reasons, having them record each of these factors by hand does not work. We need a (semi-) automatic way of determining these aspects.

A. Personal lifestyle information

As lifestyle data we will consider any data that is in any way connected to the personal lifestyle. This could be dietary facts, the current stress level, the amount of exercise, relationship status or parental status. Even though this data is all around us, it is not easy to access. We belive that a semi automated approach is most promising. We will now discuss how the raw data can be obtained and we will present methods how to calculate actual information from the different sources.

1) Movement data: Lifestyle data, which is relatively easy to obtain, is movement data. How much does the person move, or at least how much does the smart phone of a person

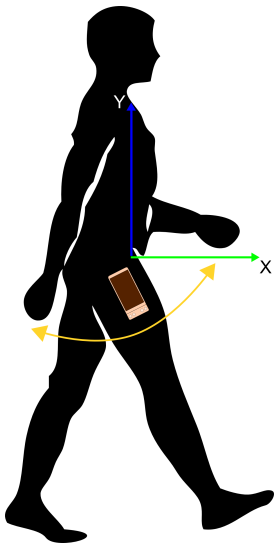

Fig. 1. The structure of the personalized information system. On the left side of the brick wall is the medical information system. On the right side is the patient information portal.

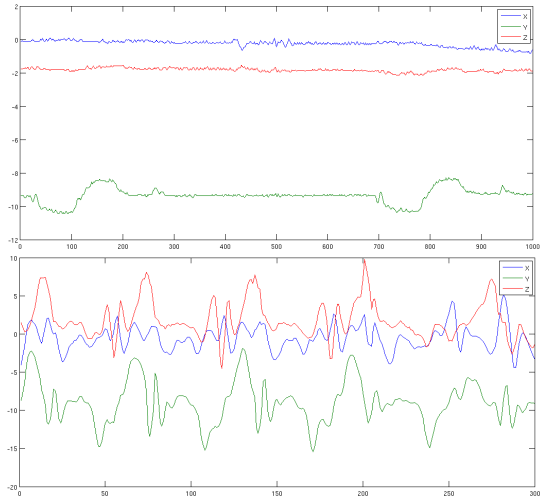

Fig. 2. Acceleration curves for a elevator ride (top) and a walk upstairs (bottom).

move? The difficulty with this kind of data is that, as long as it is not processed, it is of little use.

Figure 1 shows how the data is collected. Most of todays smart phone are capable of collecting acceleration data. There are three sensors, one for each axis in a three dimensional space. Each of these sensors collects information on the acceleration along its axis. Once the data is inside the phone it has to be classified. This is necessary because not all movement is actually due to self induced body movement. Some movement might be cause by the movement of a car the user is in or because he or she is taking the elevator instead of walking up the stairs. Figure 2 shows the acceleration curves for these two cases. There is already some promising research being done in this area [3], [1]. These approaches use time-frequency domain features and let users label their data themselfs. The result is a time series of different discrete blocks of activities. These include movement related aspects such as walking, running or climbing stairs, but also things like driving a car or taking an elevator.

With Information about different activities performed by a person and the duration of each of these activities it is easy to obtain a measure for the level of exercise performed by that person. This is already a good health indicator, but focusing sole on the amount of movement might also lead into wrong directions. A high stress level could result in a high degree of personal movement which would lead to the conclusion of a healthy life. We therefore have to include further information.

2) Stress level: Obtaining information on psychological aspects of a persons health is a difficult task. Stress for example is perceived and handled different by different people. There are however physical manifestations of stress such as in speech [6]. The voice and the articulation changes significantly when people are under stress. This is used, for example, in the area of driver safety to detect the stress and distraction level of a driver [2]. A similar approach could be used to detect the stress level of a person answering the phone. Given the speech data from phone calls, the same algorithms can be used to determine whether the speaker is currently under pressure or not. With this information we could determine the stress level at the moment of the phone call. With the data on all phone calls a user performs it would be possible to calculate a stress measure that could be used as a health indicator.

3) Eating habits: Exercise and Stress are just two important aspects of personal health. Others are eating habits. Again, modern smart phones can help us gather information on this lifestyle fact. Because, especially in urban areas, people eat out most of the time, eating habits can be obtained from their geographical location.

Most restaurants are visited for the food they are famous for. Therefore concluding from the restaurants people visit to the food they eat is not too far of a reach. People visit fast food restaurants for fast food, not for salad. They go to a sportsbar for beer and wings and not for juice and vegetarian food. If we accumulate data from several restaurant visits we might get a quite good idea of eating habits.

The difficulty with this kind of information is the classification of restaurant. Not all restaurants are already labeled with a suitable class label. We would therefore require the user to provide some initial information on the type of restaurant he is visiting and on the kind of they serve.

4) Semi automatic classification: So far we assumed that given the data and a smart enough algorithm we can deduce almost any information. Movement can be classified by looking at different frequencies, listening closely gives us information on the stress level and the restaurant visits reveal the kind of dish a person likes. Unfortunately it probably will not work that smoothly. The algorithms will most likely need further information. Such input could be a verification of a classification result: was the classification the algorithm

performed right or should it make adjustments. "Was it right to deduce that you just had steak after you visited this steak house? No this place is famous for its fresh salad bar, I had salad". Semi automatic approaches to machine learning come in different flavors and are long known in the pattern recognition community [7].

B. Medical data

Besides personal lifestyle data we will also need data on the outcome of the current lifestyle of people. This data can be obtained from different clinical or epidemiological sources.

From the early days of medicine data on the success or failure of treatments has been gathered. Evidence based medicine has increased the importance of such data collections. They have become the foundation of treatment decisions. Especially for chronic diseases, such as heart diseases, stroke, cancer or diabetes there are large collections of data that cover numerous aspects of a person as well as follow-up information. These data bases have been used as rich source for epidemiologist and should now be opened up to all people.

1) Epidemiological data: From the point of view of those people involved, these large collections of data are a good thing because all the relevant data is already available. The drawback is that, in its current complexity, most likely, people will be overwhelmed by the amount of information available. Fortunately, given the information about peoples current condition, a small number of key diagnosis dates is sufficient to calculate all necessary probabilities. Therefore it is easy to calculate the personalized expected development of a chronic disease of a person, given only his current age and as little as five to ten other variables. The obtained data can then be compared to the expected development under a different condition. All the data which is necessary for such a calculation is already contained in different research data sets. The computations required to analyze them are mostly known for a quarter of a century. Almost all statistics packages are therefore perfectly capable of analyzing them. The resulting plots, Kaplan-Meier diagrams or hazard curves are easily interpretable, even by non experts.

2) Further information: So far, we only discussed data that can be statistically analyzed. Such data is an important source of information when it comes to determining what the expected outcome of a persons current lifestyle is. Besides this quantitative data there is also qualitative data that is of interest for people. Such information might be data on different diets or exercises, the side effects of a special treatment or whether any of this is covered by insurance. Most of this information is already publicly available in different sources and different qualities.

One major source of qualitative information are medical publishers such as Thieme or Springer. Most of them provide some form of online service. These services provide information on diseases, treatments, drugs and other health care related topics. Depending on the targeted audience the treatment of the subject ranges from coarse to very fine and detailed. Besides services that stem from print products there are a number of native online services such as WebMD, MedicineNet or Healthline.

Other important sources of information can be found in user generated content such as the open encyclopedia Wikipedia or public health portals and forums such as iMedix or eHealthForum. Of course the degree of quality in these sources varies extremely depending on the person who contributed the content.

All these sources of health data provide information that can be searched for, and found, with simple key word based queries. Such a search benefits mostly from the vast amount of data available online.

III. The information system

When you decide to change your lifestyle towards a more sustainable way of life and you are confronted with a system that bombards you with all the data described above, you will soon stop using such a system. Even worse, instead of having reassuring guidance you will feel confused and insecure.

Data is probably the most important aspect of an information system, but the extraction of the valuable information from the vast amount of data available is also the most difficult aspect of such a system.

When presenting personalized information, focusing on the right information is a major challenge. In this section we will show how system architecture and information flow can be constructed such that only relevant and suitable information will be presented to the user.

The personalized information system we are proposing consists of three parts. Each of these parts performs its own data processing. The structure of the system can be seen in Figure 3. The first part is the, already existing, medical information system (left to the brick wall). Its main objective is the statistical analysis of large patient data corpora. The second part (right to the brick wall in Figure 3) is the personal health information portal which gathers and aggregates data from different sources. The third part, and one of the data sources, is the mobile application that collects the personal lifestyle information.

A. Medical information system

The statistical analysis of patient data has a long and fruitful history in medicine. Most medical research work is based on thorough numerical evaluation. The results of these calculations are mostly left to the physician to interprete. Patients usually don't come in contact with this kind of data. There is a good reason for this: The statistically valid interpretation of subtle statistical differences is neither easy nor obvious.

In a health information system the software decides which are the different options that can be compared. This way only statistically valid queries can be formulated. Interpreting the resulting plots from these queries, is also made much more easy as only a very small number of intuitive plots are presented.

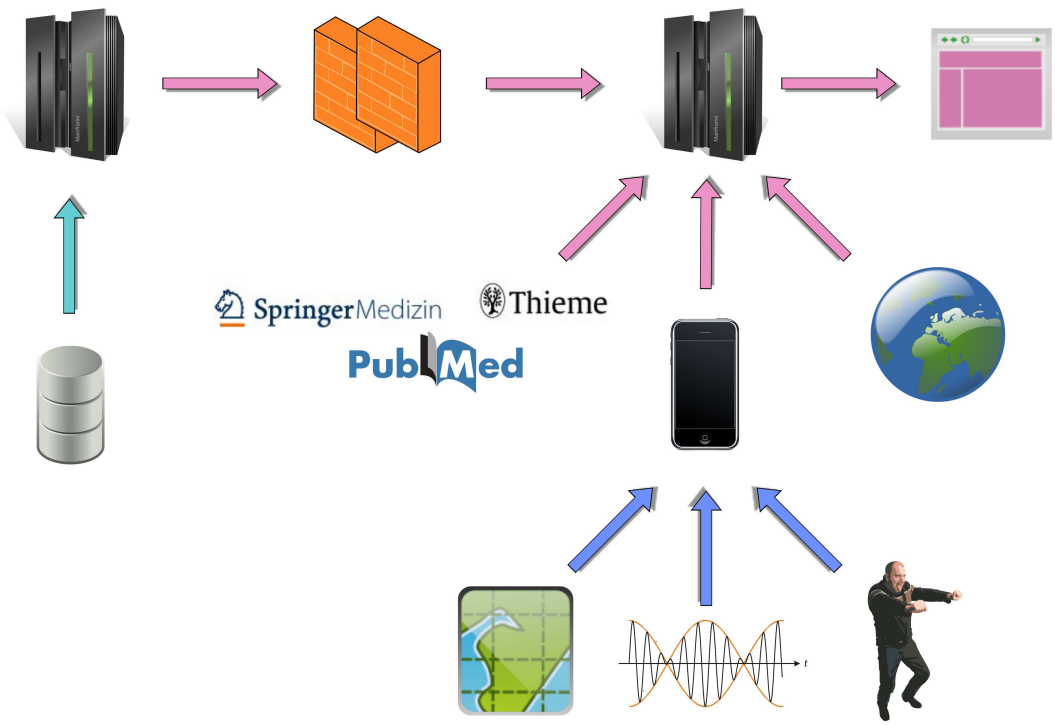

Fig. 3. The structure of the personalized information system. On the left side of the brick wall is the medical information system. On the right side is the health information portal with the mobile data collector.

B. *Personal health information portal*

Portals are places where data, people and services are integrated into one seamless presentation of information. The proposed personal health information portal integrates clinical and statistical data from the medical information system with personal health information, lifestyle information other publicly available data sources.

The clinical data will be obtained directly from medical information systems. Here the physical separation (represented by the brick wall in Figure 3) serves two purposes: First, it guarantees that only data will be presented to the user which is suitable for his needs. Second, it also protects the privacy rights of other patients in the database underlying the medical information system.

The publicly available data will be obtained from different sources. First of all there are medical publishers that provide information services. Some of them also provide standardized access methods such as SOAP, REST or the, currently under development, InfoButton Specification [4]. The main difficulty when connecting to these services is the standardized nomenclature. Different services require different name spaces. The InfoButton for example requires ICD-10 codes whereas other sources allow for ordinary key word search.

Accessing open sources such as Wikipedia is usually rather easy. Here XML-APIs are provided. Other social web sites have to be included into the portal on a individual basis.

C. *Personal lifestyle app*

Applications running on smart phones, usually called apps, have become extremely popular especially because of there simple interfaces and their easy handling. This should also be the driving force when developing such an application for health information collection: simplicity and ease of use [8].

The app in this system serves as a data entry method only. This means that its only purpose is to collect data, query information from the user and send this data to the information system. Therefore not much user interaction is required. The user has to be able to respond to queries and alter the information the app has generated. This might be the case when the app falsely assumes the user is performing some exercise or is at a certain type of restaurant. In such a situation the user has to be able to change the current data that is generated by the app.

D. *Related work*

There is, of course, already a number of methods and systems that try to tackle the kinds of problems discussed in this paper. There are, for example numerous videos or questionnaires that serve as health guidances or decision aids. Most of them are either non inter active, such as videos, or in form of questionnaires that present answers to predefined questions. Especially the interactive and explorative character of the system proposed in this paper is aimed at fostering a change in lifestyle that people can identify themselves with and have trust in.

Besides several approaches to foster a healthier lifestyle there are several web sites that server a more educational purpose (such as WebMD – http://www.webmd.com or heart.org – http://www.heart.org). These lack the interactivity and the personalization that is achieved through integrating the patients clinical data. The same holds true for web sites for people with chronic diseases, where they can organize their drug regiments and symptomatology [9].

IV. Conclusion

A sustainable change towards a healthy lifestyle requires a thorough commitment. Such a commitment is only possible if the person truly believes that such a change is necessary. We have demonstrated, in the course of this paper, how different data sources and information on the personal lifestyle can be combined in a personalized information system that provides a user with enough information to come to such a conclusion.

We argue that the current smart phone generation is well equipped to collect all data necessary to form a sufficiently clear picture of a users lifestyle. As we discussed whether state of the art technology is capable of extraction information on the users stress level, eating habits and amount of exercises performed.

Medical data and online sources on health questions are an other cornerstone of the system proposed. They augment the personal lifestyle data with life expectancy data or further information such as dietary facts or suitable exercises.

Personal health information is of concern to everyone. Making well informed decisions on our personal health should be the rule and not the exception. Public and personalized access to health information, as we propose it will benefit this cause.

References

[1] Bao, L., and Intille, S. S. Activity recognition from user-annotated acceleration data. *Pervasive 2004* (April 2004), 1–17.

[2] Bořil, H., Boyraz, P., and Hansen, J. H. L. Towards multi-modal driver's stress detection. In *Proc. of 4th Biennial Workshop on DSP for In-Vehicle Systems and Safety* (Dallas, TX, 2009).

[3] Brezmes, T., Gorricho, J.-L., and Cotrina, J. Activity recognition from accelerometer data on a mobile phone. In *IWANN '09: Proceedings of the 10th International Work-Conference on Artificial Neural Networks* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 796–799.

[4] Del Fiol, G. Context-aware knowledge retrieval (infobutton) decision support service implementation guide. Tech. rep., HL7, 2010.

[5] Goetz, T. *The decision tree*. Rodale, 2010.

[6] Hansen, J., and Patil, S. Speech under stress: Analysis, modeling and recognition. In *Speaker Classification I*, C. Mller, Ed., vol. 4343 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2007, pp. 108–137.

[7] Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. *The elements of statistical learning*, corrected print. ed. Springer, 2002.

[8] Linghao, Z., and Ying, L. On methods of designing smartphone interface. In *Software Engineering and Service Sciences (ICSESS), 2010 IEEE International Conference on* (2010), pp. 584 –587.

[9] Molpus, J., Ed. *The Impact of Personalized Medicine Today*. Breakthroughs Reports. HealthLeaders Media, April 2010.