Product vision

Paul Bakker
pbakker3
4326091

Robin Borst
robinborst
4291972

Matthijs de Groot
mawdegroot
4171683

Julian Hols
jthols
4247930

Jan van Rosmalen
javanrosmalen
4318609

April 30, 2015

Contents

1 Introduction

For this project we will focus on the behavior of patients having a new kidney who are using the website mijnnierinzicht.nl for self monitoring. Patients have to measure their creatinine level, blood pressure and weight. The website gives feedback by looking at the creatinine level. This program will sort the data in such a way that we can easily check if the website is used as intended. This program will look at the behavior of one patient. To do that we will take data from three different sources to look at the behavior of the patient. The first file is the data from the mijnnierinzicht.nl website about what the patient entered to the website. It will also give information about what kind of feedback the patient got. The second source is the file from the devices used for measuring the creatinine level, so we can see what and how many times the patient has measured. The last file is the visit log of the hospital where the patient is registered. This program will take data from multiple sources and can combine these sources using 8 standard transformations. This makes it easy to create one new data set which is easy for further analysis using SPSS.

2 Customer

This product will focus on the behavior of the patients which are using the website mijnnierinzicht.nl, so this program is first and foremost intended for the researchers at the ADMIRE project. The ADMIRE project has created the website mijnnierinzicht.nl which makes it possible for patients to use a home-based medical device for measuring the creatinine level. Conventionally these patients had to go to the hospital for every measurement, which is really costly and time consuming. The website then gives feedback using the creatinine level and gives a course of action for the patients if needed.

They have run their first test round of the website. Our product will make it possible for the researchers of the ADMIRE project to process all the raw data so that they can use it in their statistical analysis. They will use this to look how the website is used by the patients and look where there could be improvements for the website. This program will output all the relevant information specified by the analyst. The ADMIRE project can then look at how patients interact with the website using their own statistical analysis program, like SSPS.

But we are also looking to the possibility to extend this program to be used with similar projects. As this program will be using a scripting language and the data is only sorted according to scripts it could be possible to change it for use in more projects. As an example it could be used for the similar project which has the same concept but is used by diabetes patients who measure their blood sugar level.

3 Customer needs

The customer has indicated that they want to be able to process the data for a single patient. This data is extracted from three different sources as mentioned in the introduction. The customer wants to import this data on which he wants to run a script containing different data processing commands. When the script is completed the user wants to be able to export this data to a text file in which the user can specify a delimiter so it can be used to import in a different statistical analysis program for further processing.

The user wants to process all the raw data is such a way that he can easily answer questions that he might have about the behavior of a patient. These questions are for example:

- How long does it take between the measurement and the entering of that data in the website?
- Does the patient always enter the correct value measured by the StatSensor?
- Does the patient always enter each measurement?
- Does the patient keep to the given measuring schedule?
- Does the patient measure again if the website gives that feedback?
- Does the patient contact the hospital only when the website gives that feedback?

Also the user wants to visualize the processed data with a time-line graph or a transition matrix. These visualizations can illustrate the answers to the questions above. Ideally the user would also like to export these visualizations to an image file so it can be used in reports and presentations.

4 Crucial product attributes

4.1 Parse

The product must be able to parse the data that is generated by the Statsensor, the hospital records and the website mijnnierinzicht.nl. The data records from the different data sources are not static, meaning they can change over time. The product is therefore required to have parsing functionality that is not hardwired to the current data set, but a function that dynamically adjusts the parser to handle the input format. The parsed input is available for the program to run operations on.

4.2 Transformations

4.2.1 Chunks

The product must be able to perform chunking. Chunking is the aggregation of adjacent data that the researcher views as coherent.

4.2.2 Comments

The product must be able to link comments to each step of the data analysis. Comments are syntactically unstructured notes that are linked to any part of an analysis.

4.2.3 Codes

The product must be able to link codes to data elements or previous analysis steps. Codes are syntactically structured labels that are linked to data elements or chunks. We will use these codes for each events to easily detect common patterns.

4.2.4 Connections

The product must be able to specify connections between different data fields. Connections represent either the relations between ESDA products of a similar type or links between qualitatively different ESDA products that are nonetheless based on the same data element.

4.2.5 Comparisons

The product must be able to perform comparisons on different chunks or subresults from previous analyses. Comparisons allows the analyst to judge the effect of certain factors on sequential data. Secondly, an analyst may wish to compare different parts of one subjects data or to compare different subjects data possibly collected under different conditions.

4.2.6 Constraints

The product must be able to place constraints on the data. Constraints represent a selection of part of the data for further transformation and a temporary exclusion of the remainder.

4.2.7 Conversions

The product must be able to convert data into a new representation of the data. Conversions are data transformations that allow new patterns to emerge.

4.2.8 Computations

The product must be able to perform computations on the data. Computations refer to the broad range of formal procedures for analyzing sequential data.

4.3 Scripting language

The researchers requested the ability to dynamically create queries to run on the data. The product must give the researches the possibility to create a query by writing a script in the product's scripting language. The scripting language describes the syntax in which the researchers can specify the query they want to run on the data. The script can also be saved for a later time. The product runs on the data of a single patient, to save time the researchers requested a save and load function so they do not have to recreate the script for every patient.

4.4 Output

The output of the product will be used with a statistical analysis software package, such as IBM SPSS. The different statistical tools the researchers use have different interpretations of CSV file format. The CSV file format can be used with a variety of "delimiters". The researchers requested support for the CSV file format with the possibility to choose which delimiter is used.

5 Timeframe and budget

The final product is set to be finished on June 19th 2015. The product is developed using the SCRUM method so there is a tested version every week. The budget of the product is hard to determine because the developers of the product are students. The budget can, however, be described by the amount of scheduled man-hours each developer. The Contextproject is set for 10EC per person, we have a group of 5 persons and each EC represents 28 hours so the budget of the product is 1400 man-hours.

6 Comparisons

One tool similar to the one we have to build, is uFREASI, "an integrated software system to process, analyze, and assess the quality of ICP-MS data". Here ICP-MS is an abbreviation used in chemistry, which is not important for us. What this tool basically does, is getting some raw input data, processing it using some algorithm, doing calculations to show results and generating an output file (Tharaud, Gardoll, Khelifi, Benedetti, and Sivry, 2015). Although this tool has a similar task in the sense of data processing, uFREASI was made specially for use in chemistry. For ADMIRE or similar self-measurement research programmes, this tool would not suffice.

Since self-measurement for patients is quite new, analysis on that data is new as well. That's why there are no existing products to compare our product to. We can only compare to similar tools like uFREASI mentioned before, but that is not very useful, since both programs are built for different topics.

Our tool has the unique selling point that it is specifically designed for evaluating data of self-measuring devices. This is because it executes the algorithm for generating feedback to the patient - displayed as a traffic light in ADMIRE - on the actual data from the measuring device, and compares this to the data the patients entered. Also it can combine other data such as hospital visits to study behavior of the patients, such as how they react to the feedback. Besides that, the program can easily be adapted to deal with different kinds of feedback algorithms, and so it can be used for several self-measurement analyses.

References

Tharaud, M., Gardoll, S., Khelifi, O., Benedetti, M., & Sivry, Y. (2015). UF-REASI: User-FRiendly Elemental dAta procesSIng. A free and easy-to-use tool for elemental data treatment. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84923107053&partnerID=40&md5=ac07704d0376caedc17eab5d057d5beb