Visualizing Quantified Self Data Using Avatars

Isabella Nake, Aris Alissandrakis, and Janosch Zbick

Department of Media Technology
Faculty of Technology
Linnaeus University
Växjö, Sweden
Email: isabella.nake@gmail.com, {aris.alissandrakis,janosch.zbick}@lnu.se

***Abstract*—In recent years, it is becoming more common for people to use applications or devices that keep track of their activities, such as fitness activities, places they visited, the music they listen to, and pictures they took. These data are used by the services for various purposes, but usually there are limitations for the users to explore or interact with them. Our project investigates a new approach of visualizing such Quantified Self data, in a meaningful and enjoyable way that gives the users insights into their data. The paper discusses the feasibility of creating a service that allows users to connect the activity tracking applications they already use, analyse the amount of activities, and then presents them the resulting information. The visualization of the information is proposed as an avatar that maps the different activities the user is engaged with, along with the activity levels, as graphical features. Within the scope of this work, several user studies were conducted and a system prototype was implemented to explore how to build, using web technologies, such a system that aggregates and analyses personal activity data, and also to determine what kind of data should and can be collected, to provide meaningful information to the users. Furthermore, it was investigated how a possible design for the avatar could look like, to be clearly understood by the users.**

Keywords–Quantified Self; Avatars; Data Visualization.

I. INTRODUCTION

Making active use of most entertainment, productivity, lifestyle, health and fitness, etc. applications and services can produce a large amount of personal data. There is a growing interest in collecting data generated from users' everyday lives, indicated by the increasing emergence of technologies and applications available to track users' activities. One of the commercial purposes for doing this is recommendation. An application used to listen to music does not only keep record of the songs that were played, but can also give recommendations to further songs and artists a user might like, based on the information of what other users are listening to. This is done, e.g., by Spotify [1]. A similar approach is used by Amazon, who provides recommendations for products based on what a customer bought in the past, what they have in their virtual shopping cart, their ratings and likes, as well as the behaviour of other customers who viewed or bought similar items [2]. There can also be other purposes, such as for maintaining personal records. A user might use an application in combination with her fitness activities to get detailed information on, e.g., how long and how fast she was running. But although many of the applications and services keep track of activity data, they do not necessarily provide detailed feedback to the users. Users are also not always aware of how often they are actually making use of some applications, e.g., how often they use an application that allows them to take pictures to share with their friends.

The concept of the Internet of Things (IoT) is rapidly growing, connecting more and more real-world physical objects, like household appliances, buildings and human bodies, as well as items like food or clothing, to the Internet [3]. Many devices are now available that, besides communicating with each other, can sense and collect a variety of data from users' behaviour. These include ubiquitous mobile phones and smart wearable devices. These substantial and heterogeneous data are used by the different services, but are not necessarily shared with the users, and usually there are limited ways to analyse and process the information.

Recently, interest has been growing in collecting data about health, sports, sleep and other activities within our everyday lives [4]. Although people worry about sharing too much information about themselves via social media platforms and being tracked by ad networks, the interest in generating more personal data is increasing [5], and this movement is called *Quantified Self*. As Swan [6] describes, "The quantified self (QS) is any individual engaged in the self-tracking of any kind of biological, physical, behavioural, or environmental information." To track the data, either applications or devices are used. The QS website (http://quantifiedself.com/guide/) lists more than 500 tools, which can be used for the data collection. Examples for tracked data are "weight, energy level, mood, time usage, sleep quality, health, cognitive performance, athletics, and learning strategies" [6]. One main aim of collecting QS data is to examine ones life [5]. Though some people might just be interested in storing information about themselves to remember it later on, others want to get a deeper meaning out of their data [7]. Those want to discover patterns to gain self-knowledge and self-awareness of their behaviour and thus have the possibility to change and improve their lives.

Large amounts of data are generated by and collected (usually by the different service providers) from users' everyday activities, as well as from making use of smart devices, wearables and of course mobile phones. The difficulty lies, as usually noted in the "big data" domain discussions, in interpreting the data into useful information (cf. [8]). Numerous applications exist which can visualize the data, but further research is necessary to offer easily-understandable information visualization, especially to cross-reference different types of data [9]. Although these "big data" are used by the applications and services for a variety of purposes, they remain mostly unavailable to the users, with limited ways to interact, explore and maybe get insights based of their own information. The work presented in this paper attempts to address this and explore ways to reclaim the ownership of these "big data" for more personal uses.

Our project investigates in particular how to build a system, using web technologies, to aggregate and analyse personal activity data (related to QS) in order to provide meaningful information to the user. The focus lies on identifying social media (and other) services that can be considered as sources of data, what kind of data can be collected from each of these services, how the data could be collected and aggregated into this system, as well as how often the collected data should be updated. Another equally important aspect of this project is to determine a possible (graphical) design for the visualization of the information so it is meaningful to, and understood by, the users. Based on previous work, it was assumed that a suitable visualization for such a system could consist of an "avatar" that would map the different activities the user is engaged with, along with the levels of those activities, as different graphical features.

The remaining of the paper is structured as following: Section 2 presents related work. The methodology is described in Section 3. An overview of the system prototype, which was developed in the context of this project, can be found in Section 4. Results and analysis of the study are shown in Section 5. Finally, Section 6 summarizes and concludes the study.

II. Related Work

Chang [10][11][12] wrote a three-part series about QS and *Gamification*, which can be defined as "the use of game design elements in non-game contexts" [13].He emphasizes the importance of making the data collection, and thus the sensors, ubiquitous and hidden, since people are not likely to change their usual behaviour just for getting some numbers, e.g., by having to enter all their activities manually in an application. Furthermore, he suggests to give scores and rewards for activities, based on predefined rules which are known by the users, rather than presenting numbers and graphs.

Within the scope of our project, a new approach of visualizing QS data is explored. The users should not have to start using a new tracking application for collecting their data, but having the data gathered from existing services they already use. After analysing the collected data, these would be visualized as an avatar. The avatar should show the users how active they are using certain services; this is something which they may not be aware of, or it could not directly correspond to their personal impression. It could also make them aware of the extent that the different services collect personal data from them. Such avatars could also be shared with friends, for example via social media platforms (posts, status updates, or even profile pictures). This sharing could allow for comparisons, or competing for which avatar shows the most or least activity for a certain service, which brings elements of Gamification into the QS domain.

A. Data Collection and Visualization

Research has been conducted to design personal visualizations and personal data presentations. Choe et al. [14] investigated how a visualization system should be designed to support Quantified Selfers, who might not be visualization experts or data scientists, with getting insights to their own data. The following four areas were proposed for the development of personal visualization systems: self-reflection as a personal insight, gaining valid personal insights, communicating personal insights and visual annotation for highlighting insights. Wang et al. [15] presented users different personal visualization designs to identify important elements in personal visualizations. Results showed that through abstract visualizations the users are more encouraged to explore their data and thus, those visualizations give them more insights.

To track QS data, usually either an app like *Moves*, or devices like *Fitbit* or *Nike+* can be used. Those services usually offer a website or a mobile application where the users can see values and related graphs for their activities.

An application that gives immediate feedback to the user in a more *playful* way is *FitCat*. FitCat is an application for the *Pebble* smartwatch, which tracks the fitness activities of the user. With the FitCat application, the watch shows a cat which behaves in the same way as the user: if the user is walking, the cat is walking; if the user is relaxing, the cat is relaxing. The more active the users are, the more virtual money they get within this application, which they can then use to make the cat happy.

A tool which allows the users to track any kind of data and tries to visualize them in meaningful and easy understandable graphs is *Loggr*. Similar to this tool is *mem:o*, which main focus lies on the design of the visualization using colourful circles. Although both tools give a nice visualization on the tracked data, the disadvantage of those two applications is that the user has to enter all the data manually. Since already a lot of applications exist which help the user to track different kinds of data, automatically aggregating the data in order to get a visualization would be preferable to manual data entry.

B. Avatar Creation

Related to the chosen visualization, the avatar, is the Musical Avatar [16], a personal avatar generated based on the users music preferences. Favorite music tracks, provided by the users, are analyzed and tagged with semantic descriptors. Those descriptors are then used to create a user profile which is then represented as a cartoonish avatar. This visualization only shows music preferences, and is not further updated based on current user activity.

Research has been conducted on using an avatar as motivational or biobehavioral feedback. Murray et al. [17] proposes an avatar as a digital representation of a user for mobile health applications. Their work presents a theoretical description on how the appearance and behaviour of an avatar can trigger changes of the user's health-related behaviour. Feedback via avatars was found to be easier to understand for the users, since they are human-like and we are interacting with other humans in our everyday lives, whereas interpreting graphs is not done so often and needs to be learned first. Scott et al. [18] investigated how avatars should be designed to be expressive and communicate feedback.

A lot of research has been conducted regarding personalized avatars for virtual environments (cf. [19][20][21][22]). The user's body structure as well as kinematic properties should be reflected in the avatar. One possible purpose of building an avatar based on the user's body is to create a personalized shopping avatar; the user can then put clothes

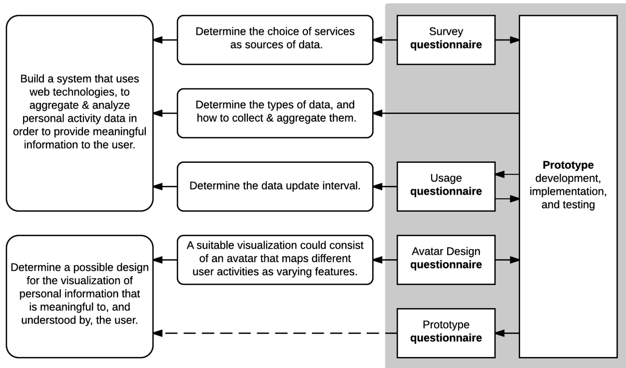


Figure 1. How the online questionnaires and the prototype development informed the research.

on the avatar "to preview the appearance of the items on the user" [23]. All those studies are building avatars based on the appearance of the user, but do not take the user's behaviour into account.

III. Methodology

Within the scope of this project, a series of online questionnaires were designed and a prototype was implemented. Figure 1 illustrates how the different questionnaires and the development of the prototype informed the research.

A. Online Questionnaires

Over the course of this project, three questionnaires were designed. An additional questionnaire, for the participants that allowed to have some of their activities monitored over a period of weeks, was also designed.

1) Survey Questionnaire: The aim of this initial questionnaire was to find out the popularity of different social network services and smart/wearable devices among potential users of the proposed system. By knowing which services and devices are familiar and actively used, an informed initial decision could be made regarding which services and devices the system prototype should support. Furthermore, the reasons why people started using those services/devices, the reasons why they keep using them, and how they share their activities, were investigated.

2) Avatar Design Questionnaire: This questionnaire was created to explore options and validate the graphical design of the avatar, in particular regarding the different graphical features that were intended to correspond to categories of activities, as identified by the previous Survey Questionnaire (and also considered feasible to support by the prototype). The aim of this questionnaire was to investigate if the participants could recognize, and interpret the meaning of, these features. Furthermore, it aimed to investigate if the way of indicating the different levels of those activities (by variations of the graphical features) was clear.

3) Usage Questionnaire: The final questionnaire was designed to investigate potential users' preferences regarding to how an avatar should be updated according to usage patterns, identified based on data that were collected using the prototype, as well as whether it should include date information.

B. Prototype Development

To investigate the feasibility of the proposed system, a prototype was developed. For testing purposes, ten participants were asked to register with the system, allowing to collect activity-related data from them over a period of a month. The purpose for collecting these data was to get a better understanding and identify patterns and levels of usage in the behaviour of users, regarding their activities. Those patterns were also used as the basis of the Usage Questionnaire.

1) Prototype Questionnaire: Additionally, as a thanking gesture for their assistance, once the collection period was over, the data from each participant were analysed and a personalized avatar was generated, corresponding to their activities during the testing period. But before their personal avatar was presented to them, to be shared on social media or used as a profile picture, they were shown the list of all the generated avatars and were asked to pick the one they thought represented them.

IV. System Prototype

A. Functionalities

The implemented system is a tool that allows users to connect their own accounts from other applications. The tool then aggregates the usage data and, based on analysis of how actively those applications were used, it can generate a graphical avatar which represents each user and (the level of) their activities.

1) Website: Part of the prototype is a website which includes a User Management System. Users can register and afterwards, they can connect their accounts from other applications/services. Since several services exist which support the same activities (e.g., *Instagram* and *Flickr* both let users upload and share their photos), the services are categorized. At the current stage, *Human API* ("fitness"), *Last.fm* ("music"), *Instagram* ("photos"), and *Twitter* ("social networks") are supported (see Survey Questionnaire results and Table II). After a period of collecting some initial data, the users are able to see their personalized avatar. This avatar can be exported (saved as an image) and then, e.g., be upload as the user's profile picture on Twitter or Facebook.

2) Data Collection and Aggregation: Based on the results of the Usage Questionnaire, once a day the data from all the services of each user are requested, analysed, and the level of each activity is saved into the database.

The same type of activities are collected from each service within a certain category. For example, for "fitness" the distance and duration of the sport or physical activities can be considered, for "music" the number of songs the user listened to, for "photos" the number of pictures the user uploaded or liked, and for "social networks" the number of posts the user wrote, commented, shared or liked. For each of these categories, the different dimensions (e.g., walking, running and biking for fitness, or posting, commenting, sharing, and liking for social networks) need to be combined into one value to be represented on the avatar. This is done in a very simplistic way and based on subjective impressions of the researchers based on the results of the data collection.

The activity value A for each category is calculated as follows. For *fitness*, the (sum of) the distance for each type

 ISBN: 978-1-61208-468-8

TABLE I. ACTIVITY LEVELS AS DEFINED IN THE PROTOTYPE, BASED ON PARTICIPANTS' DATA COLLECTION OBSERVATIONS.

		activity level	
category	**low**	**medium**	**high**
fitness	x$< 10,000$	$10,000 \leq$x$< 100,000$	$100,000 \leq$x
music	x< 100	$100 \leq$x$< 1,000$	$1,000 \leq$x
photos	x< 5	$5 \leq$x< 10	$10 \leq$x
social networks	x< 5	$5 \leq$x< 10	$10 \leq$x

provided by Human API was considered:

$$A_{fitness} = d_{walk} + 2 \times d_{run} + 3 \times d_{cycle} + d_{unknown} \quad (1)$$

weighted based on average speed assumptions. A way to correspond distance and duration for different types of sports was investigated, but nothing practical was found. One practical way would be to consider the calories that were burned, but only some wearable fitness devices can provide such information. For the "unknown" type, we simply assume it to be equivalent to "walking", especially if no different physical activities are combined.

For *music*, the number of songs provided by Last.fm was considered:

$$A_{music} = n_{songs} \quad (2)$$

For *photography* the number of pictures uploaded on Instagram was considered:

$$A_{photography} = n_{uploads} \quad (3)$$

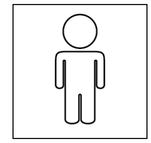
For *social networks*, tweets, retweets and favourites were considered:

$$A_{socialnetworks} = n_{tweets} + \frac{n_{retweets}}{2} + \frac{n_{favourites}}{4} \quad (4)$$

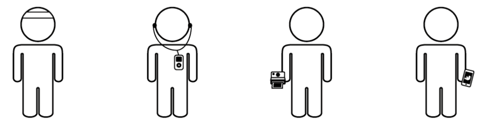
Posting new content should be considered more active than responding or liking existing content.

3) Avatar creation: The level of activities can correspond into one of the following: *not connected*, *connected* (but no activity), *low activity*, *medium activity*, *high activity*. For each category, two limits were defined: any value lower than the first limit would indicate *low activity*, a value between those two limits would result in a *medium activity* and everything above the second limit would be considered as *high activity*. Table I shows the levels per category. Based on the category of those activities, features and feature variations can be added onto the base avatar figure and generate the most recent version for each user.

The overall design was intended to be monochrome (black and white) and minimalistic in an abstract, "cartoon", style (cf. the "Big Triangle" in [24]). The base avatar is a simple outlined figure, as shown in Figure 2a. To indicate that the user has connected a service, an additional feature is added to the avatar, as shown in Figure 2b. These are: a headband for fitness, earphones for music, a Polaroid camera for photos, and a smartphone for social networks. To indicate the levels of activity in each of them, three levels of variation are added: drops of sweat from the headband, music notes coming out of the earphones, photos dropping out of the camera, and birds (implying activity on Twitter) flying out of the smartphone, as seen in Figure 3. Each of the features and its variations was meant to occupy a separate and distinct part of the avatar, with no overlapping. Figure 4 shows different avatar examples.



(a) Base.



(b) Left to right: fitness, music, photo, social networks.

Figure 2. The avatar along with the features that indicate the different services.

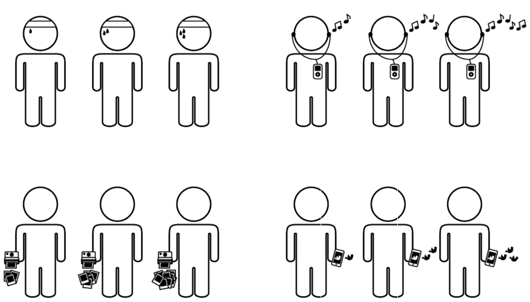


Figure 3. For each of the avatar features, variations that indicate increasing (left to right) levels of activity.

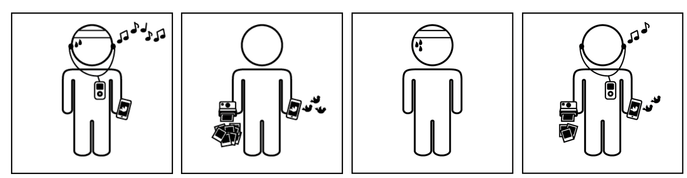


Figure 4. Different avatar examples.

B. System Architecture

The implemented system is distributed over three components (see Figure 5). The three components share one common database, but work independently from each other. The *User Interaction Server* is the only one that the user interacts with. This server hosts the website where the users can register and login, respectively, and where they can view their avatars. Using this server, they also connect external services (at the current stage *Human API*, *Last.fm*, *Instagram*, and *Twitter*) to be used for data gathering. During this connection process, the server obtains the users authentication credentials for each service and stores them in the database. Once the user is registered and has services connected, the *Data Collection Server* will request the user's data from each service, once a day. After the data is collected, it is analysed and the amount of activities are written into the database. The actual data, which were collected from the service, are not stored. The *Avatar Creation Server* queries the amount of activities for each user from the database on a regular basis and generates the avatar accordingly. The created avatar is then stored in the database.

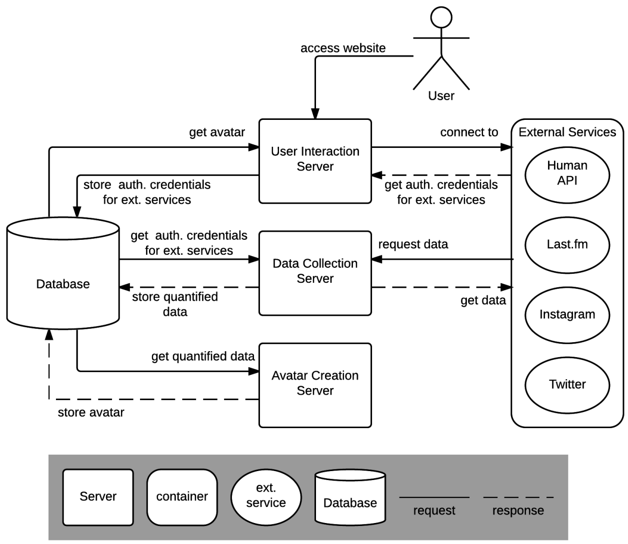


Figure 5. System architecture overview.

The system prototype architecture aims to provide maintainability, scalability and extensibility. Application Programming Interfaces (APIs) can change over time, services that do not have an API yet could offer one and also new services can come up in the future. To be able to cope with those changes and integrate new APIs without much effort, the system needs to be maintainable as well as extensible. Scalability helps not only to deal with more and more data, which is requested daily for each user, but also to handle a possibly increasing number of users.

Splitting the three functionalities of the system – user interaction, data collection and avatar creation – to three servers instead of having all functions on one server supports maintainability, scalability and extensibility. The system is "maintainable" due to being able to have two parts function properly even if an error occurs in the third part. Furthermore, the error cause can be searched within a limited environment, i.e., on the particular server that showed unusual behaviour, instead of the whole system. The "scalability" of the system is given, since not one single server has to handle all the requests from the users and the ones to the external services. While one server is responsible for managing the user requests and retrieving the authentication credentials, another server gathers the actual data from those services. Further scalability could be achieved by having two separate databases, one for storing the amount of activities and the other one for storing the avatar images. The separation of the three different functionalities makes it easier to add new functions to the system and makes is easily "extensible". Possible extensions are adding a new category or a new service in an already existing category, having more levels of activity for each category, and so on. In addition, the possibility of adding another server with new functionalities, which makes use of the data in the database(s), is given.

C. Technological Background

The development of the prototype application was intended to use currently emerging technologies. An important aspect was to build a maintainable, scalable and extensible system. To achieve this, not only a proper architecture is needed, but also the use of technologies that should support these characteristics. In this section, the tools and technologies that were used for the implementation are described.

1) HTML5, CSS3, and JavaScript: The client-side of the prototype application is built with *HTML5* in combination with *CSS3*. Additionally, *JavaScript* and its library *jQuery* is used.

2) Node.js, Express, and Jade: The server-side is based on *Node.js*. This platform allows to build fast and scalable network applications [25]. It makes use of a single-threaded and event-driven programming style together with a non-blocking, asynchronous I/O model [25][26][27].

Traditionally, when a process wanted to do some I/O operations, the process had to wait until the I/O operations were done and could only then continue running. To avoid those waiting times and having the possibility of several users at the same time, multi-threading was introduced. Each process can use different threads; within one process, all threads share the same memory. Using this model, another thread can use the CPU, as long as one thread is waiting for I/O operations to finish. If several CPUs are available, various threads can run at the same time. The disadvantage of this model is that programmers do not know which thread will be executed at which time, and need to take care of concurrent access to the shared resources. It is not unusual that bugs appear randomly, which can be hard to debug. Event-driven programming is an alternative to multi-threading, and used by Node.js. It is based on an event loop, i.e., one process contains only one thread. It basically runs two functions, event detection and event handler triggering, continuously. Every time the event loop detects that an event happened, it triggers the corresponding event handler, which is usually registered as a callback function. When I/O operations are performed, the process is not blocked and thus, several I/O operations can be executed in parallel.

In 2003, Kegel described first the C10K problem: how can one server support ten thousand concurrent clients [28][29]. Multi-threading allows serving multiple clients at the same time, but the number of threads is limited since only a certain amount of memory and processing power is available. Thus, multi-threading does not support a high capacity. In contrast, event-driven programming is more (resource) efficient. Using events as light-weight constructs and its non-blocking I/O model, Node.js is able to handle high capacity and thus, not facing the C10K problem.

Furthermore, the programming language used for Node.js is JavaScript. Its closures and first-class functions make it useful for event-driven programming. In addition, JavaScript is widely used in (client-side) HTML-based web applications. With Node.js, developers use JavaScript also on the server-side and therefore, only one programming language is needed. Another benefit of using JavaScript on both sides, is the *JSON* format, which is native to JavaScript and useful for any data exchange between the server and the client.

For the prototype, together with the Node.js platform, the *Express* framework is used to handle the middleware and *Jade* as template engine for HTML templates.

3) MongoDB: As data storage, the document-oriented, non-relational database *MongoDB* is used [30]. Non-relational – also called Not Only SQL (NOSQL) – databases like

 ISBN: 978-1-61208-468-8

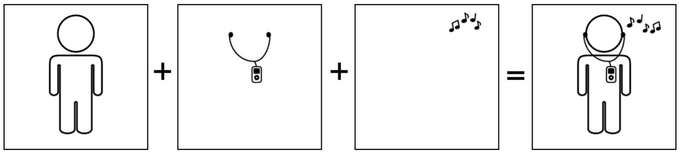


Figure 6. Composition of different images (base, feature, variation) to create each avatar.

TABLE II. PARTIAL RESULTS REGARDING POPULARITY OF VARIOUS SOCIAL NETWORK SERVICES AND WEARABLES.

service or wearable	used often or sometimes	never used or not familiar
Facebook	74	4
Google Drive	59	19
Spotify	50	28
IMDB	50	28
Google+	43	35
Twitter	42	36
Instagram	39	39
LinkedIn	36	42
Wordpress	29	49
Flickr	28	50
Nike+	13	65
Fitbit	3	75

MongoDB support scalability through *scaling out* (balancing data and load across multiple servers) which allows them to be used for a high amount of data [30][31][32]. In contrast, relational databases support only *scaling up* (getting a new and bigger machine), which makes them inefficient to be used for such a use case. In addition to effective scalability, MongoDB offers a good query performance even for a massive amount of data, which makes it very suitable for "big data" applications.

Since MongoDB uses "JSON-like" documents, it is easy to use together with the JavaScript code on the server-side through the Node.js platform [30]. Those documents can contain a complex hierarchical structure, which might correspond more to how object-oriented programmers think about their data than the, in traditional rational databases used, "rows". Furthermore, no definitions for those structures exist. Keys and values do not have a fixed type or size. It is easy, to add or remove fields within a document. This means, that the structure of the data can be easily changed if necessary. On the other hand, developers need to be careful to not end up in a complete mass of data.

4) ImageMagick: To generate the avatar based on the users activities, several images – up to two for each category (one for the feature and one for the variation) plus the base avatar – need to be composed dynamically on the web server (cf. Figure 6). For this purpose, the tool *ImageMagick* is used, which is a command line tool to create, edit, compose or convert bitmap images. The library *GraphicsMagick* allows to use the tool on a Node.js server.

5) Integrating APIs: The purpose of the system is to analyse each user's activities, via their use of various applications, and to then visualise the result as an avatar. Apart from determining popularity and usage from different services, to see which of those are reasonable to include in such a system, it was investigated which services offer a suitable API that would allow to gather the appropriate data.

One limitation can be the lack of access to an API, e.g., Nike+ has an API, but the company only gives access to specific partners and therefore, it could not be used for the current prototype.

On the other hand, there exist some APIs that already connect multiple services and allow developers to just integrate one API, while still be able to get data from all those sources. The users can then choose which of those service(s) they want to connect. The advantage of integrating only one combining API instead of separate ones for each service is not only less programming effort, but also guarantees that the data structure is the same, no matter from which application they are retrieved.

Developers integrating other services via their APIs need to be aware that those might change over time. Some changes only add new functionalities and do not affect currently running applications that already make use of this API. However, other changes require developers to integrate them in their applications, which can involve a lot of effort and time. For example, after the last change of the Facebook API, several applications needed to shut down because they were using functionalities which were not available any more (see http://techcrunch.com/2015/04/28/facebook-api-shut-down/).

Furthermore, some services do not allow to make use of any features that might not correspond to their core features and functionalities. For example, Facebook shuts down applications that would notify the users if someone unfriended them, because this would violate the Terms of Service (see Facebook Policy section 4.4, and Facebook Legal Terms section 3.2). Two examples were "Unfriend Finder" in 2013, and "Who Deleted Me" in 2015.

V. Results and Analysis

A. Survey Questionnaire

1) Demographics: Two identical versions of this initial questionnaire were created, one that was deployed in Sweden, and the other one in Austria. Note that many of the participants in Sweden were connected to Linnaeus University, and had an international background. There were no big differences detected in the results between the two groups, so in this section combined numbers are shown.

The questionnaire was filled out by 78 participants aged between 18 and 32 (average 24.8, standard deviation 3), whereof 40 were male and 38 female. 51 of the participants were located in Sweden, the rest in Austria.

2) Results: The initial questions presented two lists of social networks and wearables, respectively, and asked the participants to indicate which of those they are familiar with, and if so, how often they use them. Partial results are shown in Table II. Unsurprisingly, most participants used *Facebook*. In contrast, the majority of the participants were unfamiliar with smart devices and wearables (very few used *Nike+* or *Fitbit*).

The next questions investigated the reasons why the participants started to use social network services and/or wearables and why they are still using them. Most (around half) of the participants started to use different social network services to keep track of their activities, and that remains the reason that they still use them. But overall, it appears that most participants

 ISBN: 978-1-61208-468-8

did not consider that they keep using the services for the same initial intentions.

A distinct difference between sharing behaviours was also noted: in Sweden, more participants shared on social media platforms, and in Austria more participants shared with their friends offline.

According to the results of the questionnaire, participants are interested in keeping track of their activities and how much time they spend on them. Furthermore, there is a higher percentage of people who start using an application with the intention to share their activities with others, than users who engage in actual sharing. Although the investigation of the reasons for this behaviour are beyond the scope of this paper, proposing a tool which creates a visualization (e.g., an avatar) based analysis of the activities, might help to increase the number of people who are actually sharing their activities (via this visualization). In addition, it would not only allow others to see "how active" users are, but also the users themselves to see their own activity levels, something that might differ from their personal impression.

However, the main outcome of this questionnaire was the initial selection of services which were integrated and supported by the developed prototype. Informed by the results, different categories of potential activities could be defined: *social networks* (Facebook, Twitter, Google+), *work related* (Google Drive, LinkedIn, Wordpress), *music* (Spotify), *films* (IMDB), *photography* (Instagram, Flickr), and *fitness* (Nike+, Fitbit). The following subsections consider these possible identified categories, and discuss which APIs were finally supported. Based on the previously presented architectural decisions, the prototype can be extended with more APIs per category or more categories without much effort.

Social Networks: The most actively used social media platform is Facebook. However, Facebook changed its API recently and requires an elaborate and long-winded review process which was considered as out of scope for this project and thus it was not included in the prototype. Google+ and Twitter were considered similarly popular. Due to development time constrains, the choice between them was based on the fact that it is more complicated to change (via API) the user profile picture on Google+ than it is on Twitter. Therefore, Twitter is initially supported as it fits easier with the envisioned overall system workflow.

Work Related: None of the services identified in this category offers a suitable API. *LinkedIn* has a restricted API that gives access only to personal data (no possibility to get another user's activities) for open use, and only official partners are allowed to have more access (see https://developer.linkedin.com/ and http://mashable.com/2015/02/12/linkedin-closed-api/). *Google Drive* and *Wordpress* did not provide any useful data about the users' activities and were therefore also not included in the prototype.

Music: Considering music, the API of *Last.fm* can be used. Users can listen to music directly on Last.fm, or they can connect this service with various other music services and applications, among others Spotify. Those connected applications then send the information about what songs the user is listening to Last.fm (Last.fm calls this "scrobbling"), which can then suggest songs to the user which are similar to the ones she listens to.

Films: *IMDB* has no official API and therefore was not supported by the prototype.

Photography: To keep track on how active the user is regarding photography, the *Instagram* API is used. Flickr also provides a suitable API but since more people stated to be using Instagram, for development time constrains, only the latter one was supported.

Fitness: As a variety of fitness applications can be used, similar with the music category, the decision was made to take advantage of the integration of the *Human API*, which combines many of them in one single API.

B. Avatar Design Questionnaire

1) Demographics: In contrast to the previous questionnaire, only one version was distributed, with an additional question in the end, stating if the participants were living in Sweden or abroad.

This questionnaire was filled out by 56 participants aged between 19 and 54 (average 27.5, standard deviation 7), whereof 29 were male, 25 female, one trans* and one who preferred not to disclose. 33 of the participants were located in Sweden.

2) Results: Based on the Survey Questionnaire and the technical feasibility, four categories of activities were identified and therefore considered as potential features of the avatar: *fitness* (via HumanAPI), *music* (Spotify, etc. via Last.fm), *photography* (Instagram) and *social networks* (Twitter). For each category for which the user has a service connected to gather data, a graphic feature could be added to the basic avatar. The levels of activity for each of these could be signified by variations of these features.

For the initial group of questions, the participants were presented different avatars and asked what they thought the avatar was doing. Most of the participants' interpretations corresponded to the design intentions: e.g., they associated headphones or earphones to listening to music, taking photographs with the Polaroid camera, and the smart phone with being active in social networks. But some designs were not so clear. In the avatar with the headband, 32% participants interpreted the "drops" as *blood* (and not as *sweat*), and therefore that the avatar was injured; however, 49% participants interpreted the same avatar design as being physically active.

The next group of questions showed different avatars and asked if the participants agreed that the avatar is doing a specific activity. Similar to the previous questions, most participants confirmed the intended design. But (in line with previous impressions), a few participants insisted that the "physically active" avatar was injured. Some additional comments were made (mainly by participants that did not agree) that the avatars were perceived as being too "passive", and not really "active".

The next group of questions examined the perception of the activity level variations. Two avatars were presented at the same time, having the same feature (headband, music player and earphones, Polaroid camera, or smart phone) but with additional variations (sweat drops, music notes, Polaroid photos, or birds). Encouragingly, very few participants misidentified the compared activity levels. Some offered alternative interpretations: e.g., some participants did not agree that being "more active" can be represented by an increased number of, e.g.,

 ISBN: 978-1-61208-468-8

photos or music notes; the user is simply taking more photos (which is not the same as having interest in photography) or listens to “louder” music.

The next group of questions presented avatars with a combination of features, and asked the participants to identify the activities, which the majority did correctly.

The final group of questions presented pairs of avatars with different activities and levels of activities, and asked the participants to compare them. Again, the majority of the participants did so correctly.

Overall, the results showed that the design was understandable by the participants, and that they seemed to understand what the different features where designed to represent.

The suggested feature for physical fitness, as well as the use of feature variations (amount) to symbolize different levels of activity, were not entirely clear to all participants. Many participants interpreted the drops as blood and thus did not see the avatar as physically active, but rather as injured. To overcome this issue, the use of colours might be taken into account. For example, *blue* drops might rather be associated with sweat than with blood. Another possible way to indicate activity levels would be to adjust the transparency and not the amount of, e.g., the music notes. This method could also potentially allow for more activity levels than the current “one plus three”.

C. Usage Patterns

Ten participants (friends and colleagues from Linnaeus University, Sweden) were invited to register with the prototype system and let their data be collected over the period of a month. The aim of this data collection was to identify patterns in the behaviour of the users regarding their activities.

The first identified pattern is having low activity during most days, with a *peak* of a lot of activity for a, usually single, day. A typical example for this pattern was a user who took long bike tours once a week, which were tracked by a fitness application, but who did not track any other fitness activities. The second identified pattern is almost the inverse: having typically a lot of activity, and a sudden *drop* for one or two days. This behaviour was exemplified, e.g., by users who listened to a lot of music while they are working during the weekdays, but who hardly listened to their music during the weekends. The third identified pattern is of having more or less *consistent* activity. A user who was using a wearable device to keep track of all their steps during each day typically showed this type of pattern.

D. Prototype Questionnaire

After the collection period, an avatar was created for each participant according to their (overall) behaviour. Before these were revealed to them, the participants were given a list of all of the created avatars and they had to choose the one they think that represents them. Out of ten people, four selected the correct avatar, and another two selected an avatar that matches all the features of their avatar but that had different variations for at least one item. From the remaining participants, two selected an avatar that was similar, but that had one feature more or less than their actual avatar and a diverse number of features for at least one item. Only two users selected an avatar that was very dissimilar to their actual one.

E. Usage Questionnaire

1) Demographics: This final online questionnaire was filled out by 29 participants aged between 19 and 57 (average 27, standard deviation 7), whereof 14 were male, 13 female and two trans*. No location information was asked.

2) Results: This questionnaire was designed to investigate potential users’ preferences regarding to how the generated avatar should be updated according to usage patterns, as well as whether it should include date information.

Interestingly, only slightly more than half considered that having some date information displayed with the avatar would be useful. Those that did also showed a strong preference for this information to be displayed below the avatar, mostly in a “name of month, date of month, year” format (e.g., May 15, 2015).

Based on observations regarding the identified activity patterns, two issues were investigated: how often should the avatar be updated (based on the activity information), and whether any peaks or drops of activity should persist over the subsequent days.

The participants were presented with three different stories (each corresponding to an activity pattern) and for each of those stories a series of figures with ways that the avatar could be updated. To keep things simple, each of the stories mentioned only a single category of activities – photos (Instagram), social networks (Twitter), and music – and the options were a (not explicitly stated) mix of update options (every day, every second day, every week) and persistence (peak or drop indicated on only according to the events of the story or smoothed over more days).

More frequent (every day or every other day) updates were shown to be the, not very surprising but consistent throughout, preference. What was not so expected was the preference for no persistence “smoothing” over time. These findings, together with the inclusion of date information, further informed the development of the system and the avatar design.

VI. Summary and Conclusion

It is hoped that the work presented in this paper can be further extended in terms of supported services and functionalities, any graphical design issues improved, and potential system users provided with a way to, at the same time, be entertained, as well as gain some insight by the visualization of their personal activities, reclaiming in a meaningful way the data that are collected from their personal activities.

A series of user studies (using online questionnaires) were conducted, and a system prototype was implemented within the scope of this research. This paper presents the development and the architecture of a proposed system, that can aggregate and analyse personal activity data in order to provide meaningful information to a user.

Based on the results of the Survey Questionnaire, a selection of most used social networks and wearable devices was identified and categorized. Not all of these could be currently supported; the implemented four categories (and the APIs) are *social networks* (Twitter), *music* (Last.fm), *photography* (Instagram), and *fitness* (Human API). Due to the extensibility of the proposed architecture, further services can be relatively easily added and supported in the future.

 ISBN: 978-1-61208-468-8

Not every service can be integrated in such a system, depending on the availability and suitability (restrictions, not relevant activity data) of an API. In some cases, there exist APIs that already combine different services and thus multiple services can be integrated with less implementation effort.

A selection of the available data were chosen to be analysed and visualized (in our context). For fitness, the distance of the sport/physical activities; for music, the number of songs the user listened to; for photography, the number of pictures the user uploaded; and for social media, the number of tweets, retweets and favourites. Every time the data are requested and analysed, only the category activity levels are afterwards stored in the system database, and not the collected data.

Based on the results of the Usage Questionnaire, the potential users would prefer to get updates frequently (every day or every other day), and the visualization not to consider any persistence ("smoothing" over time). Therefore, the data are requested and aggregated as well as analysed once a day, generating an updated (depending on the activities) avatar.

As a possible visualization of users' activities, the graphical representation as an avatar was chosen and further investigated and validated by the Avatar Design Questionnaire. For each category for which the user has a service connected, the base figure of the avatar gets an added feature (a headband to indicate fitness activities, earphones for music, a Polaroid camera for photography, or a smartphone for social networks). The representation of the levels of activity is done by adding some feature variations to the avatar: drops of sweat from the headband, music notes coming out of the earphones, photos coming out of the Polaroid, and birds (related to the Twitter logo) flying out of the smartphone. Although the overall graphic design was well received and understood by the participants, there is always further room for improvements (e.g., the feature and variations for the fitness category should be redesigned).

A. Limitations

During the development phase, as well as during the test run of the prototype, some aspects were identified which developers need to take special care of, and some which require further considerations and improvement.

When a "mashup" (an application that combines data requested from various other applications) is created, it is important to be aware of that this application will depend on those integrated services. One point of this dependency is the availability of those services. For example, on the day when the trial of the prototype was started, the Instagram API was (temporarily) not available and thus, the invited users who registered could not at that time connect their Instagram accounts. Luckily, a few days later, Instagram fixed that particular issue, and these users could then connect their accounts. Furthermore, only certain data can be requested via an API that the services provide. For instance, the Instagram API does not provide an easy way to get comments that a user wrote. Thus, only the uploaded pictures as well as the "likes" could be used.

Additionally, some services allow to specify a date for which the data are requested. This is very useful for the prototype, since the data is gathered once a day, for events that took place the previous day only. But some services do not provide this option. A few services (e.g., the Twitter API) allow to request all data after a specific id. This gives the possibility to get all new data since the previous request was made, but it is not certain that all those events happened on the specific day. However, it is hard to obtain some data on the first time the prototype requests them, especially if the user was already using this service for a long time. For example, again with Instagram, this caused a bug regarding the number of "likes", and therefore this data were not considered for the photography activity. Instead, only the number of photos is currently used. Further development could address and fix this issue.

Another aspect regarding the use of external services which needs to be taken into account, is that some services have a limited amount of requests which are accepted within a certain time frame. This is especially important if the mashup is used by a large number of users. For the trial of the current prototype, the API use stayed within those limits.

Another limitation of the proposed system is the rather simple analysis of the gathered data. Further investigations and user studies are necessary on how to combine the different dimensions per category into a composite value (e.g., walking, running and biking for fitness; posting, commenting, sharing and liking for social media or photography). The generation of avatars for the participants who registered for the system was determined on subjective impressions of the researcher based on the results of the data collection (to determine the activity levels). For further work on this project, a more general solution needs to be found.

Using an avatar as visualization of the activities, also has some limitations, since for example only certain levels of activities can be differentiated before the design gets too complex and incomprehensible for the users. The current system uses four different levels for each category (just connected, low activity, medium activity, high activity). Gathering data from more users could determine if this segmentation is enough or if more differentiation would be required.

B. Future Work

As mentioned in previous sections, the avatar design was kept simple, and this project has several possibilities to be enhanced by future work, further investigating the preferences of the users regarding a perhaps more complex design.

One interesting aspect would be to integrate more services in the existing categories, or even more different categories, to address more users.

An open issue is the analysis of the gathered data, especially when it comes to the combination of the different data dimensions of one service into one composite value. For example, fitness and health applications can report a variety of heterogeneous data (distance, time, heart rate, etc.); social networks as well. These could be represented as multiple subfeatures on the avatar, or convert them to a common unit of measure (e.g., for fitness this could be the burning of calories).

The presented avatar generation used global values to define the activity levels for each category, meaning these values were the same for all users, independent on their individual behaviour. Another possibility would be to have personalized limits, depending on a user's behaviour – e.g., if she is usually very active, she needs to be "even more" active to

 ISBN: 978-1-61208-468-8

have her avatar showing a “high” amount of activity, whereas other users who are usually inactive just need “some” activity to have their activity shown as high on the avatar. Another approach would be to compare the activity to the activity of the previous day(s) or even week(s). If the user is more active than she was during the previous timespan, the avatar shows high activity, if the user is approximately as active as during the previous timespan, the avatar shows medium activity and if the user is less active than during the previous timespan, the avatar shows low activity.

Since it was not clear for all participants that the higher amount of properties is connected to more activity, other ways of showing the different levels for each category could be investigated. As already mentioned before, one possibility would be to work with transparency. Another one might be to work with different colours for different levels. The use of colour could also disambiguate some features (e.g., clarifying that a physical active avatar is not injured and bleeding, by making the sweat drops, coming from the headband, blue).

REFERENCES

[1] M. Van Rijmenam, “How Big Data Enabled Spotify To Change The Music Industry,” 2013, https://datafloq.com/read/big-data-enabled-spotify-change-music-industry/391 [retrieved: March 2016].

[2] J. Mangalindan, “Amazon’s recommendation secret,” 2012, http://fortune.com/2012/07/30/amazons-recommendation-secret/ [retrieved: March 2016].

[3] M. Swan, “Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0,” Journal of Sensor and Actuator Networks, vol. 1, no. 3, 2012, pp. 217–253.

[4] C. Gurrin, A. F. Smeaton, and A. R. Doherty, “Lifelogging: Personal big data,” Found. Trends Inf. Retr., vol. 8, no. 1, 2014, pp. 1–125.

[5] P. McFedries, “Tracking the quantified self [technically speaking],” Spectrum, IEEE, vol. 50, no. 8, 2013, pp. 24–24.

[6] M. Swan, “The quantified self: Fundamental disruption in big data science and biological discovery,” Big Data, vol. 1, no. 2, 2013, pp. 85–99.

[7] D. Lupton, “Self-Tracking Modes: Reflexive Self-Monitoring and Data Practices,” SSRN Electronic Journal, 2014.

[8] d. boyd and K. Crawford, “Critical questions for big data,” Information, Communication & Society, vol. 15, no. 5, 2012, pp. 662–679.

[9] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz, “Understanding quantified-selfers’ practices in collecting and exploring personal data,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ’14. New York, NY, USA: ACM, 2014, pp. 1143–1152.

[10] T. Chang, “All the world’s a game,” 2012, http://techcrunch.com/2012/03/24/worlds-a-game/ [retrieved: March 2016].

[11] ——, “The ’so what’ of the quantified self,” 2012, http://techcrunch.com/2012/03/31/quantified-self-so-what/ [retrieved: March 2016].

[12] ——, “We are our scores: The aspirational self,” 2012, http://techcrunch.com/2012/04/14/we-are-our-scores-the-aspirational-self/ [retrieved: March 2016].

[13] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, “From game design elements to gamefulness: Defining ”gamification”,” in Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ser. MindTrek ’11. New York, NY, USA: ACM, 2011, pp. 9–15.

[14] E. K. Choe, B. Lee, and m. c. schraefel, “Characterizing Visualization Insights from Quantified Selfers’ Personal Data Presentations,” IEEE computer graphics and applications, vol. 35, no. 4, may 2015, pp. 28–37.

[15] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma, “Design and Effects of Personal Visualizations,” IEEE Computer Graphics and Applications, vol. 35, no. 4, jul 2015, pp. 82–93.

[16] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambó, E. Gómez, and P. Herrera, “Semantic audio content-based music recommendation and visualization based on user preference examples,” Information Processing & Management, vol. 49, no. 1, 2013.

[17] T. Murray, D. Hardy, D. Spruijt-Metz, E. Hekler, and A. Raij, “Avatar interfaces for biobehavioral feedback,” in Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience, ser. Lecture Notes in Computer Science, A. Marcus, Ed. Springer Berlin Heidelberg, 2013, vol. 8013, pp. 424–434.

[18] M. Scott, L. Pereira, and I. Oakley, “Show me or tell me: Designing avatars for feedback,” Interacting with Computers, 2014.

[19] N. Ahmed, E. de Aguiar, C. Theobalt, M. Magnor, and H.-P. Seidel, “Automatic generation of personalized human avatars from multi-view video,” in Proceedings of the ACM Symposium on Virtual Reality Software and Technology, ser. VRST ’05. New York, NY, USA: ACM, 2005, pp. 257–260.

[20] S. Piotr and K. Bozena, “Personalized avatar animation for virtual reality,” in Information Technology, 2008. IT 2008. 1st International Conference on, 2008, pp. 1–4.

[21] M. Zollhöfer, M. Martinek, G. Greiner, M. Stamminger, and J. Süß muth, “Automatic reconstruction of personalized avatars from 3d face scans,” Computer Animation and Virtual Worlds, vol. 22, no. 2-3, 2011, pp. 195–202.

[22] Y. Chen and Z.-Q. Cheng, “Personalized avatar capture using two kinects in a moment,” in SIGGRAPH Asia 2012 Posters, ser. SA ’12. New York, NY, USA: ACM, 2012, pp. 3:1–3:1.

[23] C. Lawrence, V. Tiscareno, and S. Ng, “Personalized shopping avatar,” US Patent 8,818,883, August 2014.

[24] S. McCloud, Understanding Comics: The Invisible Art, 1993.

[25] M. Cantelon, M. Harter, T. Holowaychuk, and N. Rajlich, Node.js in Action. Manning, 2014.

[26] S. Tilkov and S. Vinoski, “Node.js: Using javascript to build high-performance network programs,” IEEE Internet Computing, no. 6, 2010, pp. 80–83.

[27] P. Teixeira, Professional Node.js: Building Javascript Based Scalable Software. John Wiley & Sons, 2012.

[28] D. Kegel, “The c10k problem,” 2006, http://www.kegel.com/c10k.html [retrieved: March 2016].

[29] L. Griffin, P. Elger, and E. de Leastar, “Project zeppelin: A modern web application development framework,” in Formal Methods for Components and Objects, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7542, pp. 184–198.

[30] K. Chodorow, MongoDB: The Definitive Guide. O’Reilly Media, Inc., 2013.

[31] J. Pokorny, “Nosql databases: A step to database scalability in web environment,” International Journal of Web Information Systems, vol. 9, no. 1, 2013, pp. 69–82.

[32] C. J. Tauro, S. Aravindh, and A. Shreeharsha, “Comparative study of the new generation, agile, scalable, high performance nosql databases,” International Journal of Computer Applications, vol. 48, no. 20, 2012, pp. 1–4.