Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

User Modeling and User-Adapted Interaction
https://doi.org/10.1007/s11257-020-09272-6

MYRROR: a platform for holistic user modeling

Merging data from social networks, smartphones and wearable devices

Cataldo Musto[1] · Marco Polignano[1] · Giovanni Semeraro[1] · Marco de Gemmis[1] · Pasquale Lops[1]

Received: 8 August 2018 / Accepted in revised form: 30 June 2020

Abstract
In this article, we present a platform that allows the creation of a comprehensive representation of the user that we call a *holistic user model* (HUM). Such a representation is based on the intuition that users' personal data take different forms and come from several heterogeneous sources. Accordingly, we designed a pipeline that: (1) extracts personal data from three examples of important classes of such sources, namely social networks, wearable devices and smartphones; (2) processes these data through natural language processing and machine learning techniques; (3) stores the output of such processing in a user model that encodes different aspects of people's life, such as *demographic data, interests, affect values, social relations, activities* and *physical states*. The resulting representation is made available to the user and to external developers. In the first case, a web interface allows the user to browse through her own personal data and to consult different facets of her HUM, in order to improve her self-awareness. In the latter, holistic user profiles are exposed through a REST interface and can be exploited by third-party applications to provide personalized services based on HUMs. In the experimental session, we evaluated usability and acceptability of the HUM in a user study which investigated how people were willing to use it. The results confirmed the effectiveness of our design choices and built the foundations for future usage of these profiles in personalized applications.

Keywords User modeling · Personal data · Social networks · Semantics · Web engineering · Quantified self · Self-awareness

✉ Cataldo Musto
cataldo.musto@uniba.it

[1] Department of Computer Science, University of Bari 'Aldo Moro', Bari, Italy

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

1 Introduction

A recent statistic by IBM[1] showed that 90% of the data available today have been created in the last years. This scenario is the consequence of: (1) *the development of the web 2.0* (O'Reilly 2007), which changed the role of web users from *passive consumers* to *active producers* of information, thus making possible the growth of *collaborative platforms* such as Wikipedia, the creation of *social networking* applications such as Twitter, Facebook and YouTube; (2) the growth of the Internet of Things (Atzori et al. 2010), which fueled the trend of Quantified Self (Swan 2013) and Personal Informatics. Accordingly, *very inexpensive* devices based on sophisticated sensors and technologies can be used today to collect and store data about people's daily lives (Rapp and Cena 2016).

Both these trends led to an exponential and uncontrolled growth of the available data and intensified the problem of *information overload* (Eppler and Mengis 2004). Indeed, users need more and more *support* to effectively sift through the large amount of information they have to deal with, and this issue fueled the research in the area of *personalized search engines* (Shen et al. 2005), *recommender systems* (Resnick and Varian 1997), and *intelligent personal assistants* (de Barcelos Silva et al. 2020).

All these technologies share the common idea of adapting their behavior based on some information about the user, like her*preferences* or her *needs*. Such information is typically encoded in a *user model* (Kobsa 1993), a digital representation of the person that stores information about the individual which is obtained by collecting and merging data explicitly provided by the user (demographic data, *ratings* on items, *reviews* of products) and data inferred by implicitly analyzing her behavior (e.g., web navigation, people followed on social networks, etc.) or the context (e.g., position gathered through GPS data).

Clearly, as the amount of available personal data grows, the need for tools and methods to effectively store and process these data and to build a *profile* of the user grows as well. However, we can point out that most of the approaches to build user models or to store personal data are currently affected by two main drawbacks:

1. Despite the heterogeneity of the available data, most of the platforms acquire and model a single source of information. As an example, platforms based on the gathering and the analysis of the footprints spread on the *web* do not (or just partially) exploit information gathered from *smartphones and wearable devices* (e.g., physiological data about the person, visited places, activities) and vice versa.
2. Most of the information about a person is stored and exploited by a *single* platform that does not communicate with other (similar) systems the information it holds. This problem, which is typically referred to as *data silo* problem, negatively affects the resulting *profiles*.

1FL01 [1] https://www.ibm.com/analytics/us/en/big-data/.

Both these issues cause a *sub-optimal* representation of the user, since the exploitation of a more comprehensive and richer set of data typically leads to better *profiles* (Rapp et al. 2018). Accordingly, in this work we present a platform that fills in this gap by building a user model that merges information collected from *social networks*, information coming from *smartphones* and physical and physiological data extracted from *wearable devices* in a *single* representation of the user.

Such a representation that we called a *holistic user model* (HUM) is built through a two-step process: first, personal data concerning the user are gathered from several heterogeneous sources. In this article, we took into account *six different data sources*: four social networks (Facebook, Twitter, LinkedIn and Instagram), Android smartphones and FitBit devices.

Next, we processed and enriched these data by exploiting a pipeline of natural language processing (Manning and Schütze 1999) and machine learning techniques, whose goal is to infer new and descriptive characteristics of the user that are used to populate different facets of the *holistic user model*. Such a holistic representation of the user, which is built and updated in real time as long as the user exploits her digital devices to produce or consume information, is finally made available to both the users themselves and to third-party services. In the first case, a web interface allows the user to access and browse among her own personal data, in order to improve her *self-awareness*. In the latter, holistic user profiles are exposed through a service-oriented architecture and can be used by external developers to integrate the HUMs in their own personalized applications.

To summarize, this article provides the following contributions:

- We introduce a conceptual model that we called a *holistic user profile* that supports the construction of a comprehensive user profile based on the aggregation of heterogeneous personal data;
- We present a platform called Myrror that allows the concrete creation of such user models through a *privacy-aware* and *transparent* profiling strategy that relies on six different sources: Facebook, Twitter, LinkedIn, Instagram, Android devices and Fitbit;
- We design a *mapping mechanism* to populate the facets of the holistic user profile based on the personal data held by the system.
- We carried out a user study that involved 40 persons, which evaluated the acceptance of the platform and the willingness of the users to provide their own personal data to build a *holistic user profile*.

The rest of the paper is organized as follows: Sect. 2 provides an overview of related work in the area of user modeling and emphasizes the distinctive features of the current work. Section 3 introduces our *holistic user models* and describes the facets we encoded in this conceptualization. Next, Sect. 4 depicts the data sources we exploited in the current work and present the overall architecture of the system. Section 5 shows the results of the user studies we designed to evaluate users' willingness to provide their own data to build HUMs as well as the perceived effectiveness of the system. Finally, conclusions and the ideas for future work are reported in Sect. 6.

2 Related work

In this section, we provide an overview of the literature related to the current work. Specifically, we aim to discuss and to identify: (1) the most suitable and reliable *data sources* to take into account to build user profiles; (2) the *dimensions* to be encoded in a comprehensive representation of the user; (3) the overall *architecture* of a system aiming at acquiring and merging heterogeneous data about the user.

2.1 User profiling strategies and data sources selection

From the early 2000s, the web has become a primary source of information in the area of user modeling. Indeed, the idea of replacing stereotype-based user profiles (Rich 1979) with *keyword-based profiles* and to use the web as a primary source of information has been definitely acknowledged in this period (Kobsa et al. 2001).

Another important shift in the area was observed in the early 2010s, when the concept of *social web mining* (Russell 2013) has been introduced. In this phase, several research investigated how to build richer profiles based on the information extracted from social networks. As an example, Abel et al. (2011) use Twitter as a source to infer user preferences. The usefulness of Facebook and LinkedIn data for user modeling and personalization has been investigated by Shapira et al. (2013), Musto et al. (2012) and Lops et al. (2011). Another interesting and recent trend concerns the exploitation of semantics-aware representations to model user profiles (Bontcheva and Rout 2014). As an example, Orlandi et al. (2012) combined social data with Linked Open Data (Bizer 2009) for preference modeling and prediction. The extraction of social data resulted as a very promising research line also to infer features different from users' interests. As an example, Golbeck et al. (2011) presented an approach to predict users' personality traits by processing content generated on social media.

According to the current literature, social networks and social media represent a fundamental source to collect data about the users and to build user profiles. Accordingly, in our system we connected four different social networks (e.g., Facebook, Twitter, LinkedIn and Instagram) in order to gather textual data and to use them to automatically infer both user interests and more fine-grained and particular features such as personality traits, emotions and inclination to empathy.

Moreover, several work recently tried to exploit signals and information different from those available on social media to build user profiles. As an example, a relevant trend is to gather and analyze users' personal data available on smartphones and on personal tracking devices. The early work in the area (Verkasalo 2010) showed that smartphone data can be a reliable source to analyze user behaviors. This intuition is also confirmed by Shye et al. (2010), who showed that smartphone data can be used to detect users' activities, and by Seneviratne et al. (2014), who use information extracted from personal devices to automatically detect users' traits.

To sum up, the findings emerging from the analysis of related literature support the idea of acquiring data coming from smartphone and wearable devices as well, in

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

order to significantly widen the nature and the type of information we acquire in our system. This design choice will lead to a more comprehensive representation of the user that relies on a larger set of data concerning different aspects of their life.

2.2 Categories of user attributes

As shown by foundational work in the area, such as the systems belonging to the Personis family (Kay et al. 2002), several *user attributes* can be acquired from the previously mentioned *data sources*. In this area, a substantial body of research investigated how these attributes can be organized in high-level categories (or *facets*). Early approaches, as that proposed by Kobsa et al. (2001), grouped user attributes in a set of five basis user dimensions: *demographic data, user skills, user knowledge, preferences and goals*.

Next, with the advent of context-aware (Abowd et al. 1999) and ubiquitous (Kuflik et al. 2012) computing, such categories have been extended in order to include also physiological (*heart beat, blood pressure*) and contextual data (*spatial position*, *emotions*, etc.) as well. As an example, the General User Modeling Ontology (GUMO) proposed by Heckmann et al. (2005) dates back to this phase.

Finally, the recent advances in social networks have required a further extension of this categorization in order to include new attributes, such as users' social connections. In this research line, we can mention the work by Plumbaum et al. (2011) and the recent conceptual model proposed by Cena et al. (2018).

The facets described in the real-world user models presented in Cena et al. (2018), which are based on eight different categories—i.e., demographic data, interests, needs, mental and physical state, knowledge, behaviors, contextual data and individual traits—represent the more comprehensive and complete conceptualization of users models currently proposed in the literature. Accordingly, we have adopted that schema as a starting point to encode our HUM. More details about this will be provided in the next section.

2.3 Architectures for user profiling

Architectures for building user profiles are split into three main categories: *centralized approaches*, *decentralized approaches*, *mixed approaches*.

Centralized approaches are typically referred to as *User Modeling Servers (UMS)* (Kobsa 2001) and rely on two main assumptions: (1) the evidence about a user can come from several different sources; (2) the profiling step should be decoupled from the adaptation and the recommendation ones, so a UMS should be devoted to the creation and the update of a user profile while arbitrary adaptive applications should just consume the profile a UMS has exposed. As an example, UM Toolkit (Kay 1994) and Doppelganger (Orwant 1991) fall into this category: the main idea behind these early attempts was to collect information about user's preferences, knowledge, needs and demographic data and to store them by exploiting an*internal representation* which is made available to external applications.

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

More recently, this research line has evolved into the idea of *lifelong user models (LUM)* (Kay and Kummerfeld 2009). The intuition behind the LUM is to build a unique representation of the user that stores all the information about an individual throughout her life, by merging data collected through many different devices. These principles are implemented in Portme (Kay and Kummerfeld 2010), a platform merging explicit feedback provided by the users with data gathered from external sources called *tellers*. For the sake of completeness, it should be pointed out that these models can be also implemented through decentralized approaches that will be discussed next.

In particular, *decentralized approaches* aim to create a standard representation of the users (e.g., by using ontologies) that uses rule-based approaches or reasoning techniques to build a general meta-model of the users. Several approaches fall into this categories, such as the General User Modeling Ontology we previously mentioned (Heckmann et al. 2005), the User Behavior Ontology (Angeletou et al. 2011) and the recent Social Web User Modeling ontology by Plumbaum et al. (2011). In all these cases, the authors built a very general ontological representation of the user and mapped rough information to the aspects they modeled in the profile.

Regardless of the specific approach used to build a comprehensive representation of the user, the merge of (heterogeneous) data coming from different sources often leads to *conflicts* between the data. As an example, two different sources may populate the same features with different (and maybe conflicting) values. Popular strategies to tackle this issue range from the detection and the resolution of conflicts *before* the user model is built, as proposed by Zapata et al. in e-learning domain (Zapata-Rivera and Greer 2004), to the design of specific *resolvers*, as proposed by Kay (1994), that acquire *all* the available data and implement conflict resolution strategies based on different heuristics. As we will show in the next section, we relied on the latter strategy, since we defined some *priority rules*, which are partially inspired by those proposed in the UM toolkit (Kay 1994).

Finally, it should be pointed out that a significant research effort has been devoted to the development of techniques for *transparent user modeling*. In this area, the concept (also referred to as *scrutable user modeling*) was first introduced by Kay (2006); Kay and Kummerfeld (2013), who implemented these principles in the Personis System (Kay et al. 2002). A similar architecture aiming at building transparent user profiles was also proposed by Kyriacou et al. in Kyriacou (2008).

In our framework, we decided to further investigate this research line and we proposed an architecture for building transparent user models that meets the principles of the recent GDPR regulations. Indeed, as stated in the regulation (see Article 22[2]), "*the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling [...] the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision*".

[2] Automated individual decision-making, including profiling. https://gdpr-info.eu/art-22-gdpr/.

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

Accordingly, our idea is to implement also a privacy-aware profiling strategy, where the final user has to explicitly decide which facets of her profile she wants to unveil to external applications, thus giving her control and awareness of the information encoded in the holistic user model.

2.4 Summary

We want to conclude this overview of related work in the area by emphasizing the hallmarks of our research and by framing it in the current literature.

- We propose a *mixed architecture* to build user models that tries to take the best out of the current literature: first, it is inspired by both Kobsa's (2001) work about Generic User Modeling and recent approaches that build a mediated representation based on social data as in Abel et al. (2013). Indeed, our approach relies on a central profiling component, but the user profile is built by acquiring the single models stored in external data sources (e.g., Facebook, Twitter, etc.) and by defining some *translation rules*, similar to those proposed by Van Der Sluijs and Houben (2006), that map the data points encoded in the user models to the facets we defined in our own holistic user profile.
- We aim to build a *transparent user model*, by giving the user control of the information about her that is spread through social networks and via personal devices. According to our privacy-aware profiling strategy, the user has to explicitly indicate which information she wants to extract from each data source she connects to the platform and has to indicate which facets of the profile she wants to unveil to third-party applications.
- A distinctive feature of the work is the *integration of the data coming from smartphones and from devices for tracking personal data*, such as FitBit. As a consequence, we will propose a very general and wide conceptualization of the dimensions to be encoded in the user profile that goes beyond all the approaches and the architectures currently proposed in literature.

In the next section, we will thoroughly describe these aspects, by introducing the concept of *holistic user profiles* and by describing the platform we developed to construct such user profiles by gathering and merging heterogeneous personal data describing the user.

3 Holistic user models

According to our vision, a *holistic user model (HUM)* is a comprehensive representation of the user which is obtained by merging heterogeneous personal data collected from social networks and personal devices. As previously introduced, our conceptual model is inspired by the one proposed in Cena et al. (2018); thus, it consists of the following facets: *demographics, interests, affective aspects, psychological aspects, behaviors, social connections, physical states*. In the following, we present a description of each facet. Specifically, we use this section to provide a general overview of the features that are included in each facet, while the design choices and the implementation details concerning the single portions of the user model will be discussed in the next section.

Demographics This facet includes all the *personal demographic information* about an individual. This group of features is typically domain independent and has a very low variability or no variability at all (e.g., the city of birth has no variability, while the current city does not change frequently). The usefulness of these features for user modeling and personalization tasks has been largely demonstrated in the literature (Kobsa et al. 2001; Wang et al. 2012).

Interests This facet stores all the information about what a user likes and what she is interested in. This is a fundamental source of information for every application that is designed to tailor its behavior based on of user preferences and needs, such as recommender systems (Linden et al. 2003). Differently from demographic data, such features are typically *domain dependent*. In general, we can state that users' interests can be modeled as a set of couples $\langle keyword, relevance \rangle$, where the *keyword* is a unique representation of something the user is interested in, while the *relevance* is a weight representing to what extent the user is interested in the keyword. It should be pointed out that we used the term *keyword* just for the sake of simplicity. As we will thoroughly describe in the next section, more sophisticated methodologies to model user interests based on semantics-aware representations that rely on the *entities* available in the Linked Open Data cloud have been exploited in this work.

Affective aspects This facet stores all the information about *users' mood and emotions*. This class of features is domain independent and has a high or even very high variability. As shown in the literature, mood and emotions can lead to a more precise modeling of the user (Tkalčič et al. 2013).

Psychological aspects This facet models information about the personality of the user, her empathy and other psychological aspects. Differently from the users' affective aspects, psychological aspects are stable and domain-independent traits, whose importance for user modeling and personalization was confirmed by Kelly and Tangney (2006).

Behaviors This facet models and manages information about the behaviors of the user and her activities. This facets encodes two kinds of data: (1) information about user's working place and about the points of interests she visits; (2) information about users' physical activities, such as running or walking activities, which are gathered by exploiting the sensors available in smartphones and wearable devices.

Connections This facet encodes all the social connections and the relationships of the user. As previously stated, neither Heckmann nor Plumbaum explicitly modeled this aspect in their representations. However, this is a very important facet since social ties represent a very relevant source of information to model the users and to predict their behavior.

Physical states This facet stores all the physiological and physical data points about the person. These data include user's physical parameters like *heart rate*, *blood pressure* as well as mental states such as *stress* and *anxiety*. In our case, these are short-term, domain-independent information and many of them can be directly detected using sensors in wearable devices (Rapp and Cena 2014).

4 MYRROR: a platform for building holistic user models

In this section, we introduce MYRROR, a platform that allows the users to connect their own digital identities in order to acquire personal data and to process them to support the creation of *holistic user profiles*. In the following, we will describe the general architecture of our platform and we will provide all the implementation details.

4.1 Design of the system

As shown in Fig. 1, MYRROR is organized by following the typical layered architecture consisting of a *data acquisition layer*, a *data processing and enrichment layer*, a *holistic profile builder* and a final layer for *data visualization* and *data exposure*.

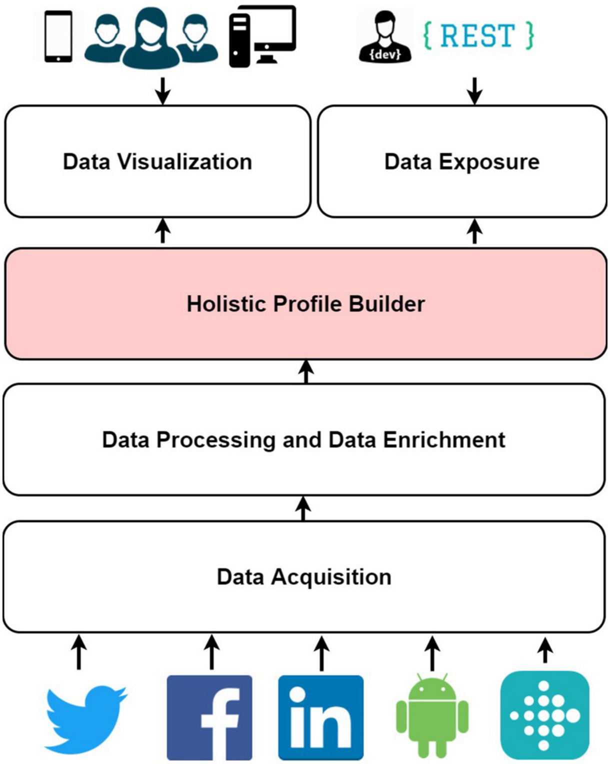

Fig. 1 Organization of the framework

4.1.1 Data acquisition

The goal of this layer is to create a bridge between Myrror and the data sources that feed our user model. In the following, we provide a description of the information we gathered from each source.

Twitter The official Twitter APIs allow to extract information about the *posts* written by the user as well as her own *connections*, along with her *demographic* features. As for the Tweets, we gathered the content of the post, its popularity (*retweets* and *favorites* count), the date of the Tweet, its language, and the information about the latitude and longitude (if any).

Facebook We extracted from Facebook basic *demographic* information (name, age, gender, language, picture, city), the content of her *posts*, the names of her *friends* and finally the name, the description and the categories of the *pages* she likes.

LinkedIn LinkedIn APIs allow to extract basic *demographic* information (as those available on Twitter and Facebook) and data about the current *working position* of the user.

Instagram Instagram APIs[3] allow to extract information about the *photographs* published by the user as well as basic *demographic* features. Specifically, Myrror gathers all the pictures published by the users, the *hashtags* used to annotate the images as well as their *description*. For each picture, the number of *likes* received by the picture and the geo-localization of the image were extracted, as well.

Android Three different groups of data are extracted from this source: *GPS data*, modeling the position of the user in terms of latitude, longitude and accuracy; *Contacts*, containing the names of the people in the contact list and the number of interactions (calls, messages) with the current user; *App Usage*, encoding the information about the apps more frequently used (along with their categories).

FitBit FitBit APIs[4] allow the extraction of information about *sleep habits* (amount of time passed in bed, minutes to fall asleep or to get awake, sleep trend, quality of sleep), *food habits* (daily calories, daily menu and type of food taken), *heart rate* (average heart rate, peak heart rate, etc.) and the *activities* of the user, such as the number of daily steps, her running exercises and daily cardio activities. Moreover, the platform also manages some *demographic information*: most of the features are already available in the other sources, such as the *name* or the *gender* of the person, but FitBit APIs also include very specific features which are not covered by the other data sources, such as the *weight* or *height* of the user.

4.1.2 Data processing and data enrichment

In the second layer of our architecture, all the data extracted through the Data Acquisition layer are processed to obtain a better representation of the data or to

3FL01 [3] https://www.instagram.com/developer/.
4FL01 [4] https://dev.fitbit.com/.

infer new characteristics of the users. Specifically, all the data stored in Myrror are processed by exploiting two different pipelines:

- *Natural language processing (NLP) pipeline*, which is designed to process textual data, such as the posts written by the user and the description of the Facebook pages she likes. *Instagram pictures* are considered as textual data as well, since a description of the images and some hashtags are provided. Our pipeline consists of six algorithms, which are all commonly used to process textual data (Manning and Schütze 1999): *language detection, tokenization, stopwords removal, lemmatization and entity linking and entity enrichment*.

  In particular, entity linking (EL) and entity enrichment algorithms have a straightforward application in our system. Indeed, EL algorithms disambiguate *polysemous* and *ambiguous* terms (such as *apple* that are stored in a user profile), and allow to understand that the target user is an *Apple fan* potentially interested in technology, rather than a vegan user interested in some ideas for her weekly menu based on *apples*. In this way, a more precise representation of users' interests is obtained.
- *Machine learning (ML) pipeline*, which is designed to process *textual* and *non-textual* data by means of ML models. These models are used to further improve the comprehension of the text, by adding extra information such as the general *topic* the content is about or the *opinion* it conveys. In particular, both a topic modeling algorithm based on latent Dirichlet allocation (Blei et al. 2003) and sentiment analysis based on the algorithm presented by Basile and Novielli (2014) were implemented in this release. Moreover, ML models were also used to automatically infer characteristics of the user, such as *emotions* and *personality*, through pre-trained models for *emotion and personality detection* (Polignano et al. 2017) and *inclination to empathy* (Polignano et al. 2018).

4.1.3 Holistic profile builder

The techniques implemented in the Data Acquisition and Data Processing layers allow the extraction and the processing of user's personal data. However, such preliminary processing is not enough since all the heterogeneous data points previously collected still need to be aggregated and merged in order to build a comprehensive *holistic user profile*.

To this end, the third step of the pipeline is carried out by the Holistic Profile Builder. In turn, this module is split in two smaller components: a Data Mapper and a Data Manager, whose goal is to populate the user profile and to manage privacy-related aspects and *conflicts* that may happen in the data mapping process, respectively.

Data Mapper The goal of this component is to aggregate the data previously collected and to map them to the facets of our *holistic user model*. As an example, the name and the surname of the user are copied in the *demographics* facet of the HUM, while the information about physical activity of the person is stored in the *behaviors* section of the user model. Such a mapping is carried out by means of some *mapping*

Table 1 Mapping between data sources and facets of our holistic user model

	Twitter	Facebook	LinkedIn	Instagram	Android	FitBit
Demographics	X	X	X	X		X
Interests	X	X	X	X	X	
Affective Asp.	X	X		X		
Psychological Asp.	X	X		X		
Behaviors	X			X	X	X
Connections	X	X			X	X
Physical St.					X	X

If a specific data source contributes to the information encoded in the specific facets, an "X’’ is reported in the table

rules that identify the most suitable facet for each information extracted from the data sources.

Table 1 provides an overview of the mapping mechanisms we implemented in Myrror. As shown in the table, each data source contributes to different facets, and each facet is populated through heterogeneous data that come from different sources. A list of different *mapping rules* we designed to populate different facets of our *holistic user model* follows. For the sake of simplicity, we can state that every time an "X’’ is put in the table, a mapping rule that translates the data collected from the source reported in the column to the facet reported in the row exists.

Demographics Our HUM includes eleven different demographic features: *name, surname, e-mail, gender, location, picture, birthday, height, weight, working position, industry*. These features are chosen by analyzing related literature, such as the general user model ontology (GUMO) (Heckmann et al. 2005) and related resource and vocabularies, such as FOAF.[5]

To encode demographics features in HUMs, we just carried out a *copy* of the available data in the corresponding facet of the profile. It should be pointed out that some of the features, as *height* or *weight*, are available on a single source (FitBit, in this case), while other features, such as the *name* or the *gender*, are available in multiple sources.

Interests Information about user interests are collected and stored in three different forms: (1) categories of the Facebook pages a user likes (e.g., politics, technology, etc.); (2) categories of the apps the user frequently uses (e.g., social networking, games, sport news, etc.); (3) *topics* that are typically discussed by the user as well as the *concepts* that are mentioned in her own posts.[6] In this case, we defined three different *mapping rules* to populate this facet of the profile. In particular, we stored: (1) the keywords describing Facebook pages; (2) the keywords describing the apps used; (3) the *entities* and the keywords extracted from users’ posts, along

[5] http://www.foaf-project.org/.

[6] From now on, the term "posts’’ is used to indistinctly refer to Facebook posts, Instagram posts and Tweets.

with the *topics* returned by the LDA algorithm. In all these cases, we obviously rely on the output previously obtained from our NLP pipeline.

However, in order to effectively model users' interests, it is necessary to handle *interests temporal decay*, whose management has been largely discussed in user modeling community (Barua et al. 2011). In this case, a background routine implemented in the Holistic Profile Builder is launched every day to slightly decrease the relevance of each element we stored in this facet of the user profile. When the evidence about a new interest is collected, the relevance score is set to 1. Next, we applied a *linear decay function* that decreases the relevance score of 0.01 every day. This value was set through a simple heuristic. This means that after almost 4 months an interests is removed from the HUM, as long as the user does not provide any more evidence about it. As future work, we will take into account and evaluate different strategies to implement interests' decay in our HUM, inspired by the findings presented in related work (Ayalon and Toch 2017; Hu et al. 2016; Rui and Zhang 2017).

Affective aspects Affective aspect, such as *mood* and *emotions*, is inferred from textual content. Accordingly, to populate this facet, we defined a simple mapping rule that relies on the output of the models for sentiment analysis and emotion detection we run in the machine learning pipeline.

In our case, we considered mood and emotions as *highly variable*, so the routines we implemented update this facet on a *daily basis*. In both the cases, the input for the models is represented by the posts written by the user during the last day, and the output is the *sentiment* (or the *emotion*, respectively) of the user predicted by the machine learning model, based on the available data. It should be pointed out that we stored in our HUM all the *emotions* and the *sentiment scores* detected by the algorithms throughout the usage of the platform.

Psychological aspects Psychological aspects like *empathy* and *personality traits* are inferred from textual content, as well. As well as for the affective aspects, we define a *mapping rule* that exploits the textual content produced by the user to populate this facet.

As for the personality traits, we used textual content as input and we stored in Myrror the scores for her Big Five Personality traits (*openness, conscientiousness, extraversion, agreeableness, neuroticism*) (Goldberg 1993) returned by the ML model for personality detection, while as for the inclination to empathy (Hogan 1969), a categorical score (*high, medium, low*) is obtained and stored.

Behaviors Information about users' behaviors can be obtained by exploiting two different data sources: (1) FitBit or Android data. (2) geo-localization information. Accordingly, two mapping rules were defined. In the first case, all the activities gathered from FitBit (running, walking, etc.) are collected and used to fill in this section of the profile. Alternatively, information coming from GPS sensors can be used to infer whether the user is making some activities. In this case, we acquire information about users' activities available in Android phones and we store them in the user profile. In both cases, this facet is updated *every day* by aggregating the raw data gathered from the data source.

Moreover, information about users' behaviors can be also obtained from geo-localization data gathered from the posts written by the user. In this case, we define a further mapping rule that browses among the geo-localized posts of the user and encodes in the *holistic user model* the name of the places or the cities visited by the user throughout her usage of the system.

Connections Social connections are filled in by gathering data coming from both Android phone and social networks. Specifically, this facet is populated through a mapping rule that executes the following two steps: (1) each contact extracted from all the data sources linked to the system is stored in the facet as a social connection; (2) the strength of the tie between the user and the contact is calculated based on the number of phone calls or on the number of interactions on social networks (*likes, favorites, retweet, etc.*) they have.

Physical states This facet is filled in through a simple rule that *maps* FitBit data to the attributes of our HUM. Specifically, all the information about *sleep* and *heart rate* is stored in this section of the profile. As for sleep, data are gathered on a *daily basis* and are used to obtain some insights about the average number of *hours of sleep*, *time to get awake* and so on. Similarly, data about heart rate are copied in this section. In this way, data about the average *heart rate* and *peak heart rate* are stored in a HUM. To summarize, these data represent physiological data points updated in real time based on the data we got from wearable devices.

Data Manager The aggregation of the data carried out by the Data Mapper allows the construction of our holistic user profiles. However, the design choices concerning privacy-related aspects and the resolution of the conflicts among the data encoded in the profile are another fundamental part of the system that is worth to be discussed.

As for the privacy, the Data Manager component takes charge of managing which data the user *wants to include* in her own HUM and which facets the user *wants to expose* to third-party applications. To this end, we designed a transparent profiling strategy where the user can control the process. Specifically, she has to (1) explicitly authorize the data sources she wants to connect to her own HUM and (2) to select which kind of data she is willing to provide, for each source. As an example, a user may authorize Facebook and may decide to allow the extraction of her posts and to forbid the extraction of her friends.

Similarly, each user can decide which portions of her HUM she wants to share with third-party developers and applications. As an example, she can decide to label her interests or her demographic data as *public* and to maintain as *private* her personal emotions, her psychological states or her connections.

Moreover, the Data Manager handles potential conflicts (such as *duplicate* or *inconsistent* information) among the data stored in different facets of the profile. As for demographics and behaviors, conflicts are tackled by introducing the concept of *priority rule*. As previously stated, these rules are inspired by those proposed in the UM toolkit (Kay 1994). The goal of a priority rule is to select *the most reliable data source*, among those connected to Myrror, for that specific attribute. Our priority rules were designed by exploiting background knowledge as well as by defining some simple heuristics based on the analysis of social network dynamics. Our choices rely on common sense knowledge and on a partial analysis of how people use social networks. As shown by previous research in the area (Kay and Kummerfeld 2013), this is a reasonable choice. As future

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

work, it is likely that a more sophisticated implementation of priority rules will be explored.

4.1.4 Data visualization and data exposure

The goal of the Data Visualization and Data Exposure layer is to make *holistic user profiles* available to both end users and external developers. In the first case, holistic user profiles are shown through a visual interface, and the users can browse among the data stored in all the facets to improve their own *self-awareness* and *consciousness*.

In particular, we designed four different methods to interact with the data: *tables, word clouds, plots and maps*. In the next section, we will provide more details about the data visualizations we made available for each facet of the HUMs, by showing the main components of Myrror user interface.

Finally, another distinguishing aspect of the system is the support to external developers who want to exploit the information encoded in the profiles in their own applications. In this case, we made available the data stored in each facet of our holistic user models (those the user selected as *public*, of course) through a set of high-level REST APIs. It should be pointed out that, due to a precise design choice, we did not allow the access to low-level and raw data extracted from the single data sources. External applications can only access through a REST interface to the information encoded in each facet of the holistic user models and can use the data to personalize and adapt their own applications.

4.2 Implementation of the system

In this section, we provide all the details concerning the implementation of the first release of Myrror. It should be pointed out that all the following screenshots are taken from the fully working online prototype of the web application,[7] that can be used in all the functionalities. Moreover, a screencast showing the organization of the system is also available on YouTube.[8]

Preliminaries Before logging in to Myrror, it is clearly necessary to sign up to the platform by providing the classical details, like *name, e-mail, password, etc*. After logging in, the user has to explicitly link the data sources that will be connected to her HUM.

As an example, Fig. 2 shows a user profile that has linked three different data sources: Twitter, Instagram and LinkedIn. Each identity is linked by clicking on the name of the data source on the left part of the user interface and by providing the credentials (e.g., Twitter login data) to connect the user profile to Myrror. As previously explained, once the data sources are connected, the user has to explicitly define which data she wants to extract from each source.

7FL01 [7] http://90.147.102.243:9090.

8FL01 [8] https://www.youtube.com/watch?v=3YRlcUhNZnQ.

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

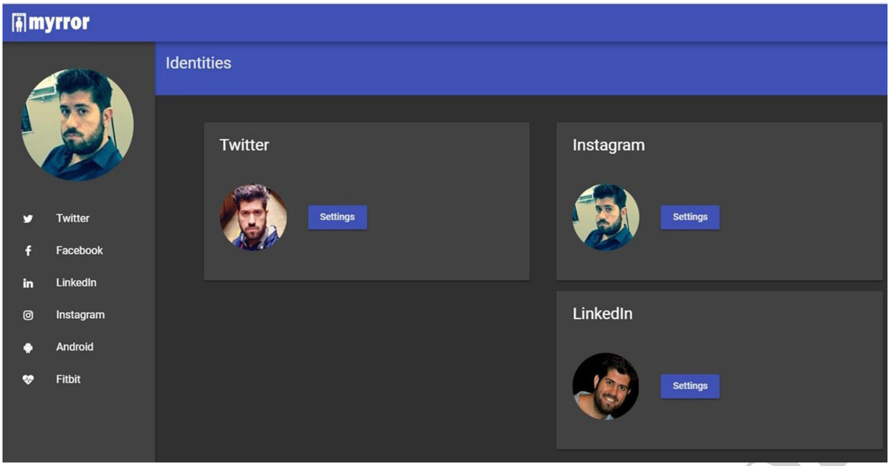

Fig. 2 Linking data sources in the data acquisition layer

Once the user has enabled the extraction of her own personal data, the profiling process can start. Clearly, the process is carried out in *background* and *periodically repeated*, in order to always provide the HUM with new and fresh data about the user.

Browsing user personal data Once the data have been correctly acquired and processed, the user has two choices: (1) to access the single data points MYRROR has extracted and (2) to browse among the facets that compose her own holistic user model. In the first case, the user has to click on the *Data* tab on her profile, while in the latter the *Profile* tab provides access to the aggregated representation of the user (Fig. 6).

Next, by clicking on the left menu of the *"Profile''* tab, the user can access to the information encoded in the single facets. As an example, users' interests are shown through a *tag cloud* like that presented in Fig. 3. Given such an interface, the user can interact with the available filters to choose (1) the selected time frame and (2) a different visualization of her interests, chosen among five alternatives: *Likes*, *App Categories*, *Hashtags*, *Concepts* and *Topics*.

As previously explained, for each alternative a different representation of the interests is adopted. As an example, Fig. 3 shows a semantics-aware representation of users' interests based on DBpedia entities.

Next, users can analyze the trend of their mood and emotions through the Affect facet. In this case, they have to choose the selected feature and the time frame, and a plot like that reported in Fig. 4 depicting the trend of the emotions over time is shown.

Similarly, the trends of the activities are shown under the *Behaviors* facet. In this case, the user can see a *line chart* showing the steps as well as the amount of physical activity she did in a certain period of time (Fig. 5), which is a very useful data visualization for Quantified Self-related goals.

Journal : SmallExtended 11257	Article No : 9272	Pages : 35	MS Code : 9272	Dispatch : 7-7-2020

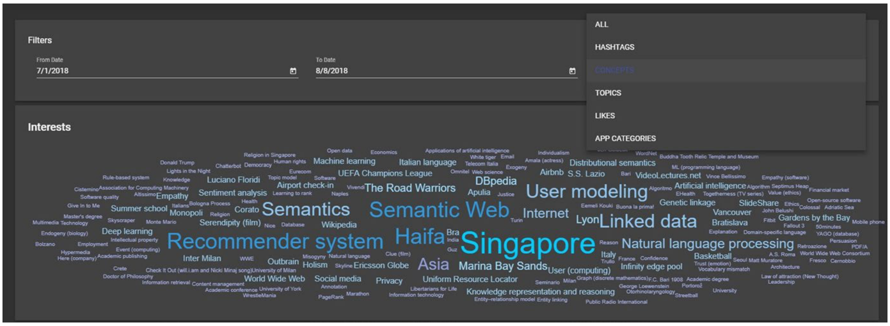

Fig. 3 Word cloud modeling users' interests in a certain time frame

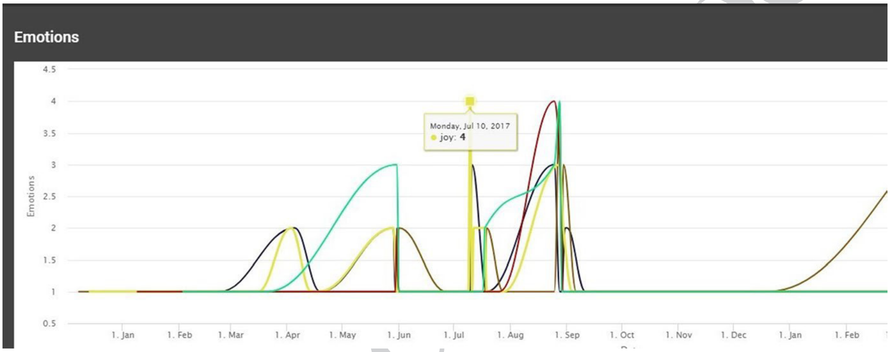

Fig. 4 Trend of users' emotions over time

Finally, as previously stated, a very important aspect of the platform concerns the privacy mechanisms we implemented in the system. In general, the platform gives to the user control over the facets of her profile that are opened to third-party applications. As shown in Fig. 6, the user is provided with the complete set of the facets of her HUM and she can click on each facet to *enable* or *disable* its visibility to external applications.

5 Discussion and limitations

The current implementation of MYRROR provides a solid foundation for further developments of the platform and its integration in third-party recommender systems. In the following, we provide an overview of the current *limitations* of MYRROR. For the sake of readability, we split the discussion by analyzing each component separately.

Data acquisition This release of the system mainly focused on data gathered from social networks and tracking devices, such as FitBit. However, the system does not

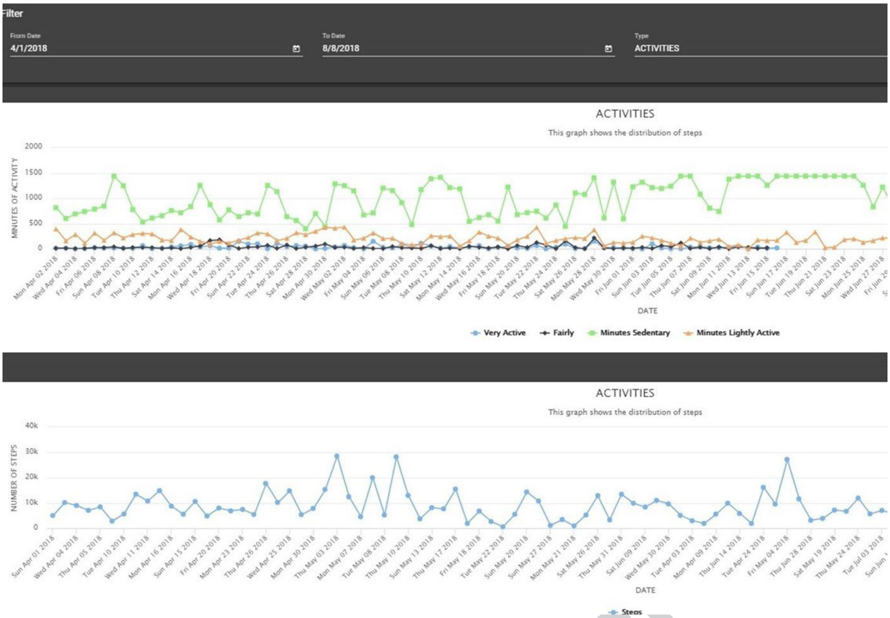

Fig. 5 Trend of users' activities over time. The first plot shows the amount of physical activity for each day, while the second shows the distribution of the steps. The time frame can be set through the date picker

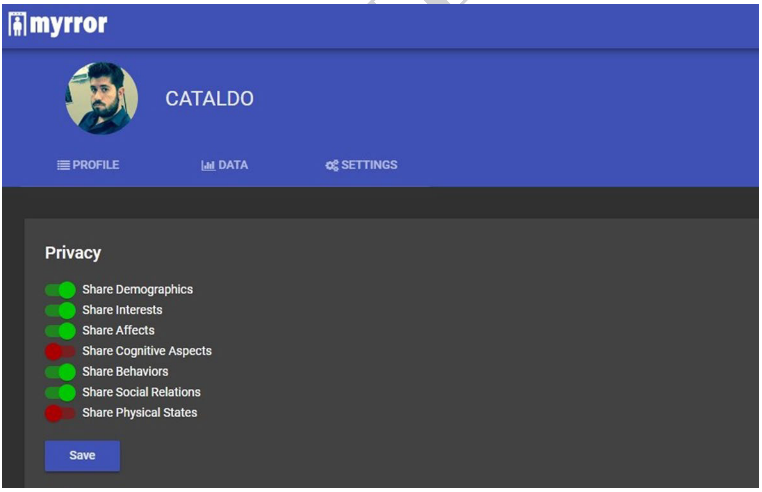

Fig. 6 Privacy and control mechanisms in Myrror

include data points concerning other aspects of people life, such as eye gaze, blood pressure and conversations. This will be investigated in future work.

Data processing Even if a good number of machine learning models are already included in the current implementation, the introduction of new algorithms would be useful to extend the information about the users which is currently held by Myrror. Examples of these algorithms are: detection of sedentary users based on the data extracted from FitBit (Lepp et al. 2013); detection of frequently visited places from GPS data (Ashbrook and Starner 2002); detection of users' interests from review data (Musto et al. 2017); and detection of emotions by using input different from textual content (e.g., (physical and physiological) data points).

Data mapping The current release of the system does not consider the concept of *context*, which is very important for several facets (preferences, affective aspects and behaviors, in particular). As an example, when the evidence about some preference of the user is gathered, we did not store the contextual situation in which that evidence has been acquired. This aspect, which is fundamental for a precise modeling of the person, is left as future work.

Data managing The current implementation of the *privacy management* mechanism is too coarse grained for a real use. As an example, a user should allow a *music recommender system* to access to her mood and may deny this privilege to a *news recommender system*. Currently, this is not possible in Myrror. Moreover, our system does not take into account the concept of *user identity*. As an example, the *same person* being in contact with the target user on both Facebook and Twitter is considered as *two different contacts*, since no alignment between different sources is done. This can be tackled through the introduction of techniques for *identity detection*.

Data visualization As for data visualization, a partial limit is represented by the fact that the widgets in the user interface are statically defined. As an example, a user can't cross and merge the data coming from two different facets (e.g., mood data and activities data) in a single visualization. Also this improvement is left as a future work.

6 Experimental evaluation

In the experimental evaluation, we carried out a user study to evaluate usability and acceptability of the HUM and to analyze whether the profiles built through the platforms matched users' beliefs and preferences. Finally, we also evaluated users' engagement and the overall usability of the system.

Specifically, we aimed to answer to the following research questions:

- *RQ1: effectiveness of the conceptual model.* Which facets of the HUM do the users consider as more relevant for getting personalized suggestions?
- *RQ2: user engagement and adherence of the resulting profiles.* How frequently do the users interact with the system? What do they think of the holistic user models the system can build?
- *RQ3: self-awareness, discover and remember capabilities.* Do the data visualizations we made available in Myrror increase users' self-awareness of their personal data? Does the system allow the user to discover or remember information about herself?

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

6.1 Experimental protocol

In order to answer to the research questions, we arranged a user study lasting 28 days (4 weeks) that involved *40 users* (80% male and 20% female). Participants were recruited by following the classic *availability sampling* strategy, a widespread technique to organize user studies in recommender systems and user modeling areas (Carmagnola et al. 2009; Semeraro et al. 2012; Lops et al. 2009). We included people having different age ranges (from university students to adults) and a different knowledge of social networks and technology in general.

Age range of the participants was 20–65. Most of the participants were under 26 years old (14 participants, 35.0%), while 13 participants (32.5%) were between 26 and 35. Finally, 9 participants (22.5%) were between 36 and 45 and just 4 participants (10%) had more than 45 years. As for the employment, most of the participants were students or worked in a private company (15 participants each, 75% in total). Next, 8 participants worked in a public company and 2 were self-employed. As for the education, 5 participants held a Ph.D., 20 participants held a degree (8 master degree and 12 bachelor degree). The remaining participants just completed high school.

Clearly, all the users owned a social network account and/or a wearable device. As for the frequency of usage of social networks, 28 participants told that they daily use social networks and personal devices, while 9 people weekly use these technologies. Just 3 participants stated that they monthly use social networks. Generally speaking, we tried to recruit a sample as much heterogeneous as possible, in order to maximize both the internal validity and the external validity of the results.

Our user study was organized in the following four sessions:

1. *Introduction and training*. First, all the participants recruited for the experiments were involved in a training session aiming at introducing the concept of *holistic user modeling* and the goal of the experiment. Specifically, we thoroughly explained how to connect each single identity to the platform, how to enable data extraction and how the privacy is guaranteed. Next, we explained the meaning of all the facets we encoded in our holistic user profiles, we introduced the basics of the mapping mechanisms that we implemented to populate the profiles, and we provided instructions on how to answer the questionnaires. Finally, we also had a discussion with the users about their doubts concerning the implementation, the privacy and the trustworthiness of the system.
2. *Information gathering*. Next, we asked participants to fill in a questionnaire (from here on, PRE-Q questionnaire) designed to assess their willingness to make their personal data available for user modeling and personalization tasks, and we guided the users to register to Myrror. Finally, each user connected her own digital identities to the system;
3. *Usage of the platform*. We asked the users to use the platform for 4 weeks (28 days). Throughout this period, we asked them to freely connect to the system (not mandatory), to interact with the platform and to browse among different data visualizations available;

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

Table 2 Answers to Question 3 of PRE-Q, investigating the willingness to allow the extraction of users' personal data

	Not willing	Little willing	Quite willing	Very willing	Not owned	%Not or little willing	%Quite or very willing
Twitter	7	11	6	11	5	**51.4%**	48.6%
Facebook	2	21	9	8	0	**57.5%**	42.5%
Linkedin	6	9	8	16	1	38.4%	**62.6%**
Smartphone	5	17	10	6	2	**57.8%**	42.2%
Wearable	7	13	6	7	7	**60.6%**	39.4%
Instagram	8	16	7	6	3	**64.8%**	35.2%

4. *Evaluation.* After the second step, we recalled all the participants of the study. First, we asked them to interact with the system and to consult the resulting profile (regardless they already had interact with the system in the previous weeks). Next, we asked them to fill in a post-usage questionnaire (from now on, POST-Q questionnaire). Through the questionnaire, we collected information about: (1) what do the users think about the resulting user profiles; (2) whether the data visualizations they interacted with were satisfying or not; (3) the ease to manage personal data through the platform; (4) the ability of the platform of acting like a self-reporting tool and whether the system allowed the user to discover new information about themselves. Moreover, we also acquired users' ideas about future development and we asked them the likelihood of a future usage of the system. Finally, users had to compile the well-known System Usability Scale (SUS) questionnaire[9] to evaluate the overall usability of the system.

The outcomes emerging from the pre-usage questionnaire (PRE-Q) and post-usage questionnaire (POST-Q) were used to answer the aforementioned research questions. Specifically, PRE-Q was used to answer *RQ1* while POST-Q and the SUS usability questionnaires were used to answer to *RQ2* and *RQ3*. For the sake of brevity, we do not report the complete questionnaires, which are available online.[10]

6.2 Discussion of the results

Before going into the details of the discussion concerning our research questions, we exploited some of the answers we collected from *PRE-Q* to investigate users' willingness to unveil their personal data to get personalized services.

[9] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

[10] PRE-Q Questionnaire: https://forms.gle/zVaSaLRCyyu26urv9—POST-Q Questionnaire: https://forms.gle/4wvxhJU6JqEPLwjh6.

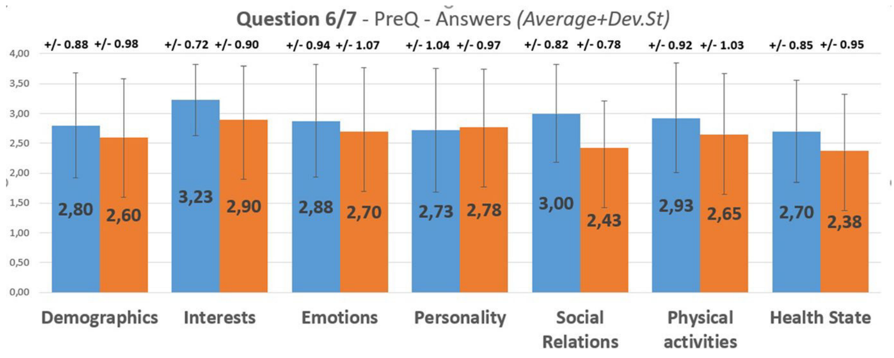

Fig. 7 Average Score and Standard Deviation of the answers collected for Question 6 (in *blue*) and Question 7 (in *orange*) of the PRE-Q questionnaire

As shown in the results reported in Table 2, users are particularly willing to provide their LinkedIn (62.6%) and Twitter data (48.6% of the participants). Conversely, the participants did not show the same willingness for more popular social networks like Facebook and Instagram and for personal devices like FitBit and smartphones.

The results emerging from this part of the experiment were quite expected, since the recent issues concerning the use of personal data by companies like Cambridge Analytica[11] raised the problem of the privacy and sensitized people toward a more careful sharing of personal data. Accordingly, these results provides two main outcomes: (1) users' need precise information about how their data will be exploited and what kind of personalized services they will obtain. This further emphasized the need to integrate our holistic user profiles in third party services that will be the focus of our next work; (2) regardless of the personalization strategy, it is necessary to design a *transparent* user profiling strategy, as that we implemented in Myrror, since the users need (and want) control of the information they share. Otherwise, it is likely that they will not be willing to provide their own personal information.

Research Question 1. Next, we analyzed the goodness of the conceptual model for *holistic user profiling*. Specifically, we asked the users about their willingness to share personal data to populate all different facets of the HUMs (Question 6 of PRE-Q), and we evaluated their opinion about the usefulness of the facets for personalization and recommendation tasks (Question 7 OF PRE-Q). Results of the comparison are reported in Fig. 7.

As shown in the figure, we can note a small decrease in the scores we obtained for all the facets. This is an expected outcome that confirms again the users' partial willingness to reveal their personal data, even if they considered as relevant all the facets of the profile. The decrease is particularly relevant for users' social relations that decreased from 3.00 to 2.43. This is probably due to the very personal nature of this facet.

[11] https://www.bbc.com/news/topics/c81zyn0888lt/facebook-cambridge-analytica-data-scandal.

Research Question 2. In order to evaluate the impact and the effectiveness of the system, we organized a second evaluation session *4 weeks* after the first one. In this time interval, the users had the opportunity (not mandatory) to connect to the system and to see how their holistic user profiles were built. Of course, during the second session the users had to mandatorily connect to the platform and to interact at least once with the resulting profiles. It should be pointed out that we recalled our sample just to answer the POST-Q questionnaire and to share some thoughts about the platform. All the data were collected *outside* the laboratory, by analyzing users' daily usage of the system.

The first aspect we investigated through the POST-Q questionnaire concerned the online identities connected to MYRROR (*Question 1*). By aggregating the answers we obtained, it emerged that 16 participants (40%) connected their Twitter online account, 36 (90%) their Facebook account, 17 (42.5%) their LinkedIn account, 15 (37.5%) their smartphone, 35 (87.5%) their Instagram account and 14 (35%) their FitBit device. In total, 5923 posts generated by the users, 4716 connections among users, 2040 likes to pages and 47,409 records from wearable and mobile devices were gathered and stored.

Next, we analyzed the answer to Questions 2–5 of the POST-Q questionnaire to answer to RQ2. First, Question 2 allowed us to investigate the frequency of usage of the system. As shown in Fig. 8, we obtained encouraging findings since most of the samples (35 out of 40, 87.5%) asserted that they used the system on a *weekly* basis, at least. Conversely, only 5 users out of 40 rarely used the system (less that weekly) throughout the weeks of the experimental.

This is a good outcome, since it means that a large majority of the users were interested in checking their profiles and following the building process throughout the time window of the experiment. Even if only a small amount of users (6 out of 40, 15%) stated that they used the system *every day*, this is not worrying. Indeed, the system was designed so that it can work in background without a continuous and explicit input of the user; thus, it is not necessary an everyday interaction. In our opinion, a weekly usage—especially for a prototype version as that we evaluated in this experiment—is an encouraging and satisfying outcome.

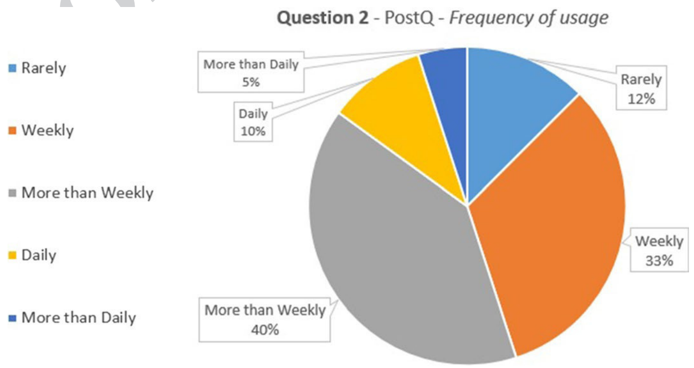

Fig. 8 Frequency of usage—Question 2 of the POST-Q questionnaire

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

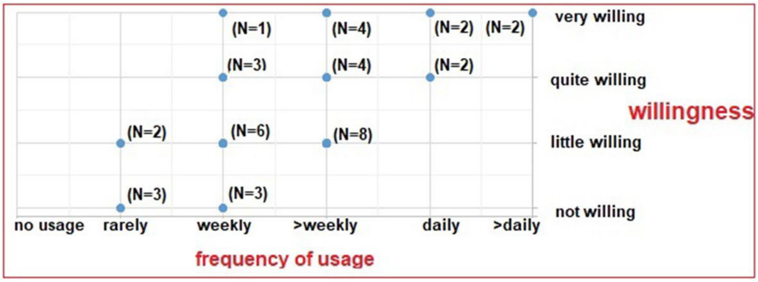

Fig. 9 Relationship between Myrror frequency of usage of and users' willingness to provide their data

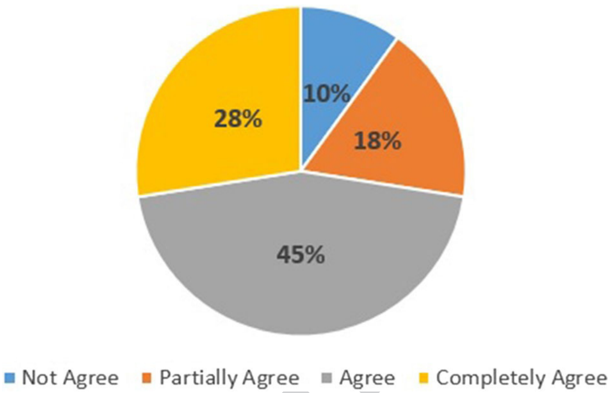

Fig. 10 Answers to Question 3 of POST-Q, evaluating the opinion of the users toward data extraction and privacy mechanisms in Myrror

Moreover, in order to deepen the analysis concerning the characteristics of the users who used Myrror more frequently, we analyzed the relationship between *frequency of usage* of the platform (as reported in Fig. 8) and *users' willingness* to provide their data, which is discussed in Table 2. It should be pointed out that labels were assigned to the users based on an strategy inspired by *majority vote* (i.e., a user who selected ``little willing'' for 3 data sources and ``quite willing'' for 2 data sources was provided with the label ``little willing''). In total, 6 users were labeled ad *not willing*, 16 users as *little willing*, 9 users as *quite willing* and 9 users as *very willing*. Results are presented in Fig. 9.

As shown in the figure, a *linear* relationship between frequency of usage and willingness emerged. Indeed, users who were willing to provide their data used Myrror more frequently. Similarly, users who are little or not willing at all rarely used the platform.

Next, in Question 3 we analyzed the ease of usage and the perceived transparency of the extraction process we implemented in Myrror. As shown in Fig. 10 we obtained encouraging results as well, since 73% of the sample agreed that the

Journal : SmallExtended 11257	Article No : 9272	Pages : 35	MS Code : 9272	Dispatch : 7-7-2020

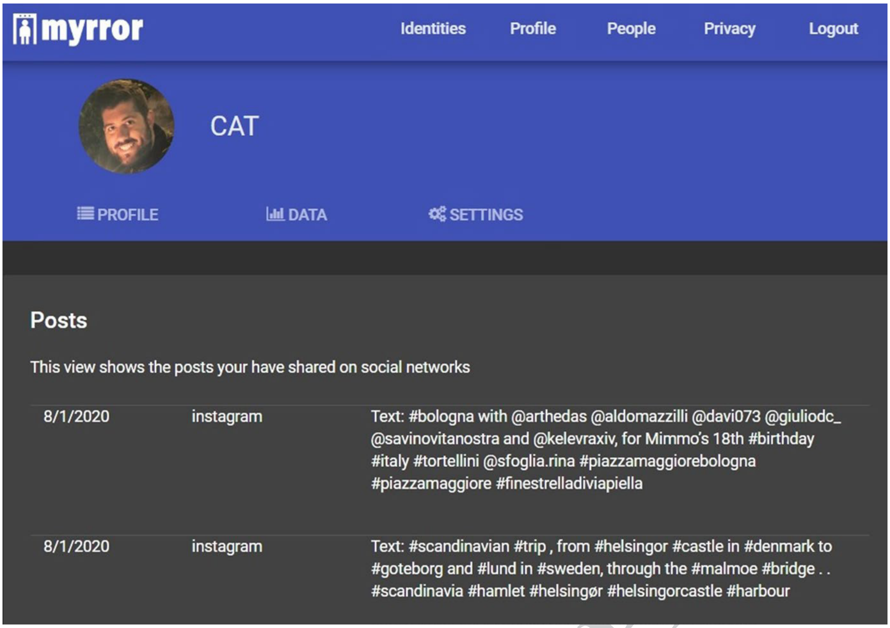

Fig. 11 Visualization of raw data in the ``Data'' section of the profile. In this case, posts of the user are shown

strategies we chose to give control to the users were understood and appreciated. This is an interesting finding that emerged from the usage of the system, which confirms that the insight of designing a *privacy-aware and transparent profiling process* is a good choice.

Another important outcome of the experiment is the impact of the aggregation strategies we encoded in Myrror through the Data Mapper module implemented in the Holistic Profile Builder. This is a fundamental part of our experiment, since it aims to evaluate whether our intuition of gathering and merging heterogeneous personal data in a smaller set of facets is appreciated by the users or not.

To this end, through *Question 4* of our POST-Q questionnaire we evaluated whether the aggregated data shown in Myrror were more effective than the raw data gathered from the single sources. Concretely, this was done by comparing the data visualizations available in the *Data* section, storing all the raw data (Fig. 11), with those available under the *Profile* section (Figs. 3, 4, 5).

As shown in Fig. 12, the results we obtained for this question are really promising since for almost all the facets of our HUM a percentage of users close to 90% partially or completely agreed that the insight of aggregating heterogeneous footprints spread over the web can lead to a better snapshot of the profile of the user.

Finally, Fig. 13 shows that 25 out of 40 user (62.5%) agreed or completely agreed that the resulting profiles are adherent to their personal beliefs. This is an encouraging outcome that further confirmed the goodness of the HUMs.

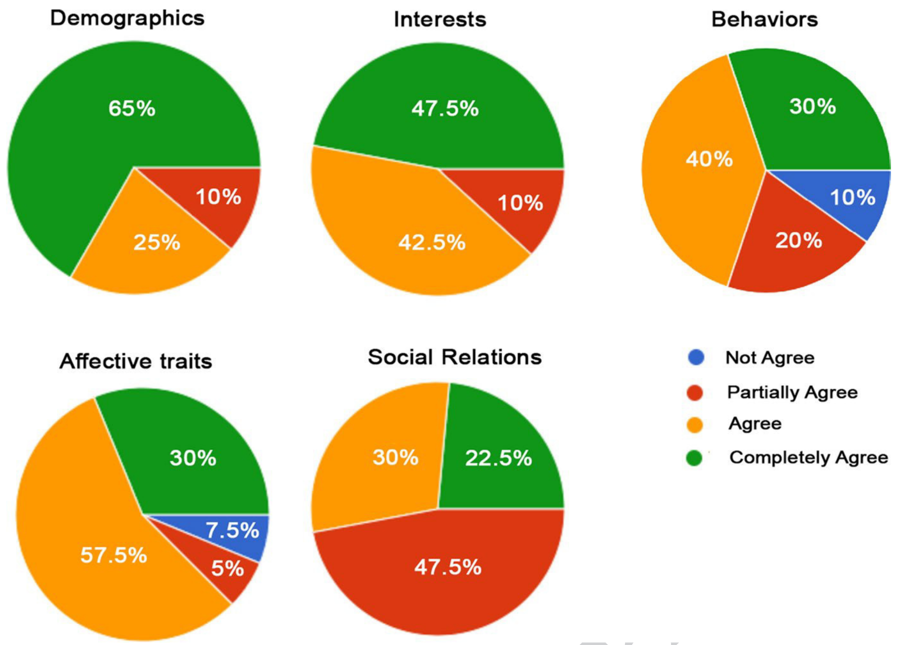

Fig. 12 Answers to Question 4 of POST-Q, evaluating the opinion toward the aggregation strategies implemented in Myrror

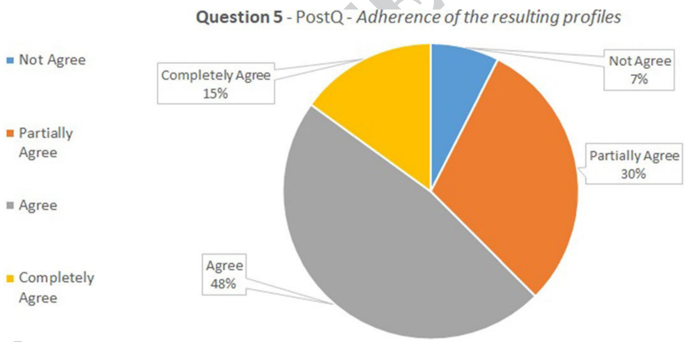

Fig. 13 Answers to Question 5 of POST-Q, evaluating the adherence of the resulting profiles available in Myrror

This outcome allowed us to positively answer to RQ2, since both user engagement and the quality of the profiles were encouraging and satisfying.

Research Question 3. Finally, through RQ3 we aimed to investigate to what extent Myrror could act as a Quantified Self tool. Specifically, in Question 6 the users evaluated the data visualizations available in the framework and stated whether they improved their self-awareness. Results are reported in Fig. 14 that reports the average answers of the users on a 4-point Likert scale.

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

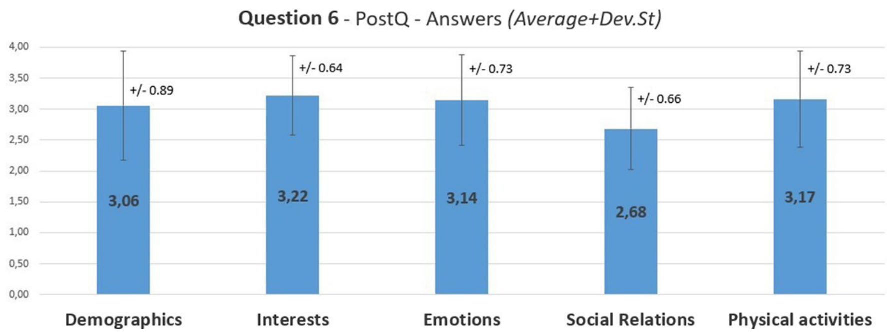

Fig. 14 Average Score and standard deviation of the answers collected for Question 6 of the POST-Q questionnaire

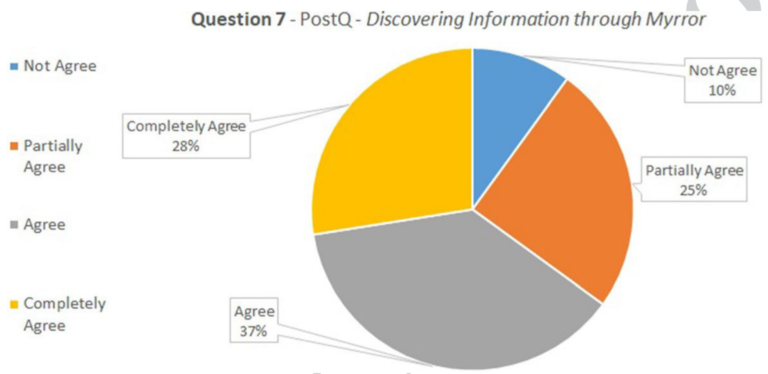

Fig. 15 Answers to Question 7 of POST-Q, assessing the ability of Myrror as a tool for ``*discovering*'' information

As reported, users generally had a positive opinion about the data visualizations available in Myrror. The tag cloud we used to model *users' interests* emerged as the most effective data visualization (3.22 out of 4 as average scores), followed by the charts we used to show *users' behaviors* and *physiological data* and those we use to report users' emotions, whose results were higher than 3 out of 4. In this case, the worst results are obtained by the visualization we used for social relations (2.68 as average score). This behavior is probably due to the fact that we did not implement any mechanisms for identity alignment, so it is likely that the results showed by Myrror for this visualization are not satisfying for the users.

Next, through Question 7 and Question 8 we were interested in assessing the ability of Myrror of acting as a *discovering* or *remembering* tool. In the first case, we investigated whether Myrror can allow the users to discover new information about themselves by connecting different and heterogeneous pieces of information, while in the second we asked the users whether the life-long

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

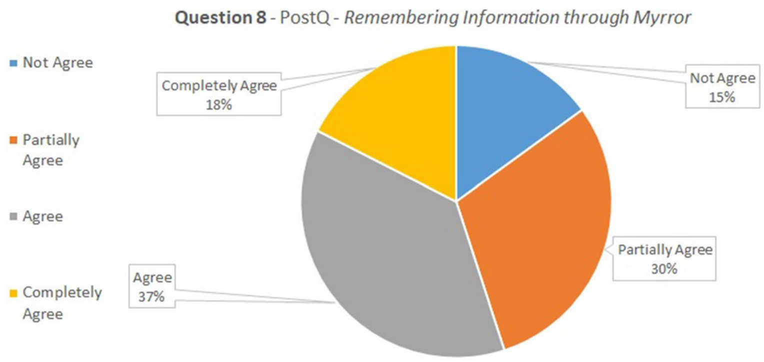

Fig. 16 Answers to Question 8 of POST-Q, assessing the ability of Myrror as a tool for ``*remembering*'' information

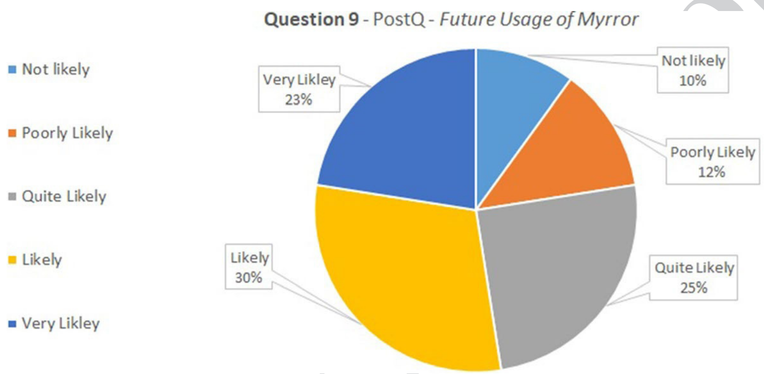

Fig. 17 Answers to Question 9 of POST-Q, about the likelihood of future usage of Myrror

storing capabilities of Myrror allowed the user to remember facts about their life. Answers to the questions are reported in Figs. 15 and 16.

As shown in the figures, in both the cases the majority of the users agreed that the system has such a capability. As for Question 7, 26 out of 40 users (65%) agreed or completely agreed that the system allows the discovery of new information through the aggregation of the data as well as through the available data visualizations. As for Question 8, the percentage of users who understood the potential use of Myrror as a lifelong logging tool decreased to 55% (22 out of 40 users). However, even this percentage can be considered as satisfying for our goals. Indeed, it is likely that such a capability of the system would emerge in longer and continuous usage of the platform, rather than in shorter experiment of 4 weeks.

Finally, we evaluated the likelihood of a future usage of Myrror by collecting the answers to Question 9. Also in this case we got interesting and satisfying outcomes, since the majority of the users (21 out of 40, 9 = very likely, 12 = likely) stated that they would have continued using Myrror in the future. By also including the users who answered that they would have used Myrror ``quite likely'' in the future, the

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

overall percentage of the users increases to 31 out of 40 user (77.5% in percentage), which is a very good outcome that confirms the good impact of the system on final users (Fig. 17). AQ2

The last aspect we investigated through our user study concerned the overall *usability* of the system. In this case, we asked the users to answer to the well-known SUS questionnaire (Brooke 1996), in order to assess about the overall usability of the platform.

As shown in Table 3 and Fig. 18, we obtained the highest results for Question 1 and Question 3, concerning easiness and frequency of use. This is not surprising, since the impact of these aspects on the overall user engagement was already discussed in the article. Similarly, we obtained satisfying results in terms of integration of different functions (Question 5) and learning curve (Question 7). The outcomes emerging from Question 7 are particularly good, given the high complexity of the system. As for the even questions, we obtained the lowest result for Question 10, concerning the training time needed to use the platform. This is an expected outcome that depends again on the complexity of the system. However, as emerging

Table 3 Results of the SUS questionnaire

#	Question	Average	SD
1	*I think that I would like to use this system frequently*	3.98	0.89
2	*I found this website unnecessarily complex*	2.23	0.95
3	*I thought this website was easy to use*	4.00	0.88
4	*I think that I would need the support of a technical person to be able to use this system*	2.18	0.78
5	*I found various functions in this system were well integrated*	3.95	0.95
6	*I thought there was too much inconsistency in this system*	2.33	0.92
7	*I would imagine that most people would learn to use this system very quickly*	3.88	0.85
8	*I found the system very cumbersome to use*	2.43	0.75
9	*I felt very confident using the system*	3.78	0.86
10	*I needed to learn a lot of things before I could get going with this website*	2.65	0.95

The higher the better for *odd* questions, the lower the better for *even* questions

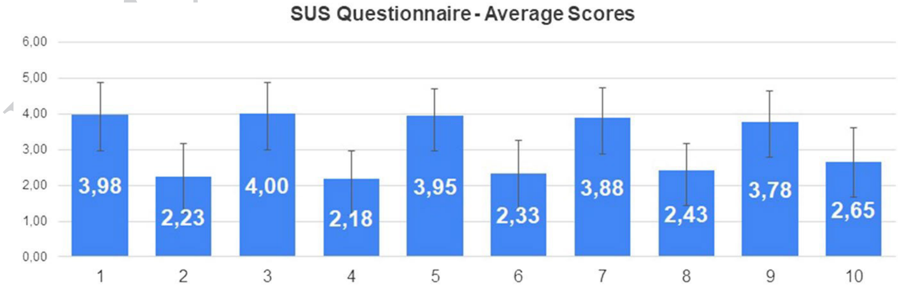

Fig. 18 Bar chart summarizing the results of the SUS questionnaire. Values on the *X axis* are mapped to the questions presented in Table 3

from Question 2, most of the samples understood that such a complexity is a necessary feature of a system whose goal is to acquire and manage such a huge number of personal data. Overall, we obtained an average SUS score of 69.44 (min = 42.5, max = 100, SD = 15.02). According to (Brooke 2013), we can conclude that the overall usability of the system is *good* (SUS score between 53 and 73).

6.3 Recap of the experiment

In the following, we want to synthesize the main findings and the main lessons we learned from this evaluation of Myrror, by answering to the research questions we introduced in this section.

- *RQ1: effectiveness of the conceptual model Which facets of the HUM do the users consider as more relevant for getting personalized suggestions?*

  Answer: we did not note a particular facet that significantly emerged as *more relevant*. In general, the users considered as important and relevant all the facets we encoded in our conceptual model, and this further confirmed the goodness of our design choices.
- *RQ2: user engagement and adherence of the resulting profiles How frequently do the users interact with the system? What do they think of the holistic user models the system can build?*

  Answer: user engagement was satisfying, in terms of both the amount of collected data and the average number of connections and login we got from the users recruited for the experiments. Overall, the users stated that the resulting user profiles were adherent to their personal beliefs.
- *RQ3: self-awareness, discover and remember capabilities Do the data visualizations we made available in* Myrror *increase users' self-awareness of their personal data? Does the system allow the user to discover or remember information about herself?*

  Answer: Yes, it does. This is a fundamental outcome of this experiment that finally confirmed that both our conceptual model and the implementation of Myrror can support the users in the creation of their own *holistic user profiles*. Indeed, the users appreciated the idea of gathering and merging their data to populate the facets of the profile. Moreover, the data we collected showed that the majority of the sample correctly perceived the system as a tool that allow to discover new information about themselves and to remember previous facts about their life by connecting different and heterogeneous pieces of information.

7 Conclusions and future work

In this article, we have presented a platform that allows to extract and process users' personal data to populate a *holistic user profile*, that is to say, a representation of the user that relies on the digital footprints left on social networks, smartphones and wearable devices. Our holistic user model is based on seven different facets,

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

as *demographic data, interests, affective aspects, psychological states, behaviors, social relations and physiological data*, and is populated through a profiling procedure that maps the raw data to the corresponding dimension of the holistic user model.

In the experimental evaluation, we carried out a user study aiming to evaluate the effectiveness of our design choices and the willingness of the users to share their personal data. The experimental results showed that the users appreciated the platform as well as the data visualizations we made available. Moreover, the study revealed that the users are more willing to provide access to their personal information only if they can see a *real value* in the personalized services they can potentially exploit. Overall, the system was appreciated by the users both in terms of functionalities and its general usability.

As future work, we will continue the development of the platform by introducing more algorithms and techniques in the data processing and enrichment layer, in order to infer new and better features describing the users. Finally, we will integrate holistic user profiles in a real use cases, like tourism personalization and food recommendation. Specifically, we aim to develop recommendation algorithms that can take advantage of the heterogeneous data points encoded in the user profile and lead to better suggestions.

References

Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing user modeling on twitter for personalized news recommendations. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 1–12. Springer (2011)

Abel, F., Herder, E., Houben, G.J., Henze, N., Krause, D.: Cross-system user modeling and personalization on the social web. User Model. User-Adap. Inter. **23**(2–3), 169–209 (2013)

Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: International Symposium on Handheld and Ubiquitous Computing, pp. 304–307. Springer (1999)

Angeletou, S., Rowe, M., Alani, H.: Modelling and analysis of user behaviour in online communities. In: International Semantic Web Conference, pp. 35–50. Springer (2011)

Ashbrook, D., Starner, T.: Learning significant locations and predicting user movement with gps. In: Proceedings on Sixth International Symposium on Wearable Computers, 2002 (ISWC 2002), pp. 101–108. IEEE (2002)

Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)

Ayalon, O., Toch, E.: Not even past: information aging and temporal privacy in online social networks. Hum. Comput. Interact. **32**(2), 73–102 (2017)

Barua, D., Kay, J., Kummerfeld, B., Paris, C.: Theoretical foundations for user-controlled forgetting in scrutable long term user models. In: Proceedings of the 23rd Australian Computer–Human Interaction Conference, pp. 40–49 (2011)

Basile, P., Novielli, N.: Uniba at evalita 2014-sentipolc task: Predicting tweet sentiment polarity combining micro-blogging, lexicon and semantic features. In: Proceedings of EVALITA, pp. 58–63 (2014)

Bizer, C.: The emerging web of linked data. IEEE Intell. Syst. **24**(5), 87–92 (2009)

Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)

Bontcheva, K., Rout, D.: Making sense of social media streams through semantics: a survey. Seman. Web **5**(5), 373–403 (2014)

Brooke, J.: SUS: a retrospective. J. Usability Stud. **8**(2), 29–40 (2013)

Brooke, J., et al.: Sus—a quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

Carmagnola, F., Vernero, F., Grillo, P.: Sonars: A social networks-based algorithm for social recommender systems. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 223–234. Springer (2009)

Cena, F., Likavec, S., Rapp, A.: Real world user model: Evolution of user modeling triggered by advances in wearable and ubiquitous computing. Inform. Syst. Front. **21**, 1085–1110 (2018)

de Barcelos Silva, A., Gomes, M.M., da Costa, C.A., da Rosa Righi, R., Barbosa, J.L.V., Pessin, G., De Doncker, G., Federizzi, G.: Intelligent personal assistants: a systematic literature review. Expert Systems with Applications, pp. 113–193 (2020)

Eppler, M.J., Mengis, J.: The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. Inf. Soc. **20**(5), 325–344 (2004)

Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. In: CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 253–262. ACM (2011)

Goldberg, L.R.: The structure of phenotypic personality traits. Am. Psychol. **48**(1), 26 (1993)

Heckmann, D., Schwartz, T., Brandherm, B., Kröner, A.: Decentralized user modeling with UserML and GUMO. In: Decentralized, Agent Based and Social Approaches to User Modeling, Workshop DASUM-05 at 9th International Conference on User Modelling, UM2005, pp. 61–66 (2005)

Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: GUMO–the general user model ontology. In: International Conference on User Modeling, pp. 428–432. Springer (2005)

Hogan, R.: Development of an empathy scale. J. Consult. Clin. Psychol. **33**(3), 307 (1969)

Hu, R., Liu, J., Wen, Y., Mao, Y.: User: A usage-based service recommendation approach. In: 2016 IEEE International Conference on Web Services (ICWS), pp. 716–719. IEEE (2016)

Kay, J.: The UM toolkit for cooperative user modelling. User Model. User-Adap. Inter. **4**(3), 149–196 (1994)

Kay, J.: Scrutable adaptation: because we can and must. In: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 11–19. Springer (2006)

Kay, J., Kummerfeld, B.: Lifelong user modelling goals, issues and challenges. In: Proceedings of the Lifelong User Modelling Workshop at UMAP, vol. 9, pp. 27–34. Citeseer (2009)

Kay, J., Kummerfeld, B.: Portme: Personal lifelong user modelling portal. Tech. Rep. TR647, School of Information Technologies, University of Sydney (2010)

Kay, J., Kummerfeld, B.: Creating personalized systems that people can scrutinize and control: drivers, principles and experience. ACM Trans. Interact. Intell. Syst. **2**(4), 1–42 (2013)

Kay, J., Kummerfeld, B., Lauder, P.: Personis: a server for user models. In: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 203–212. Springer (2002)

Kelly, D., Tangney, B.: Using multiple intelligence informed resources in an adaptive system. In: International Conference on Intelligent Tutoring Systems, pp. 412–421. Springer (2006)

Kobsa, A.: User modeling: Recent work, prospects and hazards. Hum. Factors Inform. Technol. **10**, 111–111 (1993)

Kobsa, A.: Generic user modeling systems. User Model. User-Adap. Inter. **11**(1–2), 49–63 (2001)

Kobsa, A., Koenemann, J., Pohl, W.: Personalised hypermedia presentation techniques for improving online customer relationships. Knowl. Eng. Rev. **16**(2), 111–155 (2001)

Kuflik, T., Kay, J., Kummerfeld, B.: Challenges and solutions of ubiquitous user modeling. In: Ubiquitous Display Environments, pp. 7–30. Springer (2012)

Kyriacou, D.: A scrutable user modelling infrastructure for enabling life-long user modelling. In: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 421–425. Springer (2008)

Lepp, A., Barkley, J.E., Sanders, G.J., Rebold, M., Gates, P.: The relationship between cell phone use, physical and sedentary activity, and cardiorespiratory fitness in a sample of us college students. Int. J. Behav. Nutri. Phys. Act. **10**(1), 79 (2013)

Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **1**, 76–80 (2003)

Lops, P., De Gemmis, M., Semeraro, G., Narducci, F., Musto, C.: Leveraging the linkedin social network data for extracting content-based user profiles. In: Proceedings of the Fifth ACM conference on Recommender Systems, pp. 293–296. ACM (2011)

Lops, P., de Gemmis, M., Semeraro, G., Musto, C., Narducci, F., Bux, M.: A semantic content-based recommender system integrating folksonomies for personalized access. In: Web Personalization in Intelligent Environments, pp. 27–47. Springer (2009)

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

Manning, C.D., Schütze, H., et al.: Foundations of Statistical Natural Language Processing, vol. 999. MIT Press, London (1999)

Musto, C., de Gemmis, M., Semeraro, G., Lops, P.: A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, pp. 321–325 (2017)

Musto, C., Semeraro, G., Lops, P., De Gemmis, M., Narducci, F.: Leveraging social media sources to generate personalized music playlists. In: International Conference on Electronic Commerce and Web Technologies, pp. 112–123. Springer (2012)

Orlandi, F., Breslin, J., Passant, A.: Aggregated, interoperable and multi-domain user profiles for the social web. In: Proceedings of the 8th International Conference on Semantic Systems, pp. 41–48. ACM (2012)

Orwant, J.: Doppelgänger—a user modeling system. Ph.D. thesis, Massachusetts Institute of Technology (1991)

O'Reilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software (2007)

Plumbaum, T., Wu, S., De Luca, E.W., Albayrak, S.: User modeling for the social semantic web. In: de Gemmis, M., De Luca, E.W., Di Noia, T., Gangemi, A., Hausenblas, P., Lops, M., Lukasiewicz, T., Plumbaum, T., Semeraro , G.(eds.) Semantic Personalized Information Management: Retrieval and Recommendation Workshop (SPIM 2011), CEUR, vol. 781, pp. 78–89 (2011)

Polignano, M., Basile, P., De Gemmis, M., Semeraro, G.: An emotion-driven approach for aspect-based opinion mining. In: Tonellotto, N., Becchetti, L., Tkalčič, M. (eds.) Proceedings of the 9th Italian Information Retrieval Workshop, vol. 2140 (2018). http://ceur-ws.org/Vol-2140/

Polignano, M., Basile, P., Rossiello, G., de Gemmis, M., Semeraro, G.: Learning inclination to empathy from social media footprints. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 383–384. ACM (2017)

Rapp, A., Cena, F.: Self-monitoring and technology: challenges and open issues in personal informatics. In: International Conference on Universal Access in Human–Computer Interaction, pp. 613–622. Springer (2014)

Rapp, A., Cena, F.: Personal informatics for everyday life: how users without prior self-tracking experience engage with personal data. Int. J. Hum. Comput. Stud. **94**, 1–17 (2016). https://doi.org/10.1016/j.ijhcs.2016.05.006

Rapp, A., Marcengo, A., Buriano, L., Ruffo, G., Lai, M., Cena, F.: Designing a personal informatics system for users without experience in self-tracking: a case study. Behav. Inform. Technol. **37**(4), 335–366 (2018). https://doi.org/10.1080/0144929X.2018.1436592

Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM **40**(3), 56–58 (1997)

Rich, E.: User modeling via stereotypes. Cogn. Sci. **3**(4), 329–354 (1979)

Rui, L., Zhang, X.: A tag-based recommendation algorithm integrating short-term and long-term interests of users. DEStech Transactions on Computer Science and Engineering (SMCE) (2017)

Russell, M.A.: Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. "O'Reilly Media, Inc." (2013)

Semeraro, G., Lops, P., De Gemmis, M., Musto, C., Narducci, F.: A folksonomy-based recommender system for personalized access to digital artworks. J. Comput. Cult. Herit. **5**(3), 1–22 (2012)

Seneviratne, S., Seneviratne, A., Mohapatra, P., Mahanti, A.: Predicting user traits from a snapshot of apps installed on a smartphone. ACM SIGMOBILE Mobile Comput. Commun. Rev. **18**(2), 1–8 (2014)

Shapira, B., Rokach, L., Freilikhman, S.: Facebook single and cross domain data for recommendation systems. User Model. User-Adap. Inter. **23**(2–3), 211–247 (2013)

Shen, X., Tan, B., Zhai, C.: Implicit user modeling for personalized search. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pp. 824–831. ACM (2005)

Shye, A., Scholbrock, B., Memik, G., Dinda, P.A.: Characterizing and modeling user activity on smartphones: summary. In: ACM SIGMETRICS Performance Evaluation Review, vol. 38, pp. 375–376. ACM (2010)

Swan, M.: The Quantified Self: Fundamental disruption in big data science and biological discovery. Big Data **1**(2), 85–99 (2013)

Tkalčič, M., Burnik, U., Odić, A., Košir, A., Tasič, J.: Emotion-aware recommender systems—a framework and a case study. In: ICT Innovations 2012, pp. 141–150. Springer (2013)

Van Der Sluijs, K., Houben, G.J.: A generic component for exchanging user models between web-based systems. Int. J. Contin. Eng. Educ. Life Long Learn. **16**(1–2), 64–76 (2006)

Verkasalo, H.: Analysis of smartphone user behavior. In: 2010 Ninth International Conference on Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR), pp. 258–263. IEEE (2010)

Wang, Y., Chan, S.C.F., Ngai, G.: Applicability of demographic recommender system to tourist attractions: a case study on trip advisor. In: Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03, pp. 97–101. IEEE Computer Society (2012)

Zapata-Rivera, J.D., Greer, J.E.: Interacting with inspectable Bayesian student models. Int. J. Artif. Intell. Educ. **14**(2), 127–163 (2004)

Cataldo Musto is assistant professor at the Department of Informatics, University of Bari. He completed his Ph.D. in 2012 with a thesis entitled "Enhanced Vector Space Models for Content-based Recommender Systems." His research focuses on the adoption of natural language processing techniques and models for fine-grained semantic content representation in recommender systems and user modeling platforms. He was a visiting researcher of Philips Research Center in Eindhoven (The Netherlands) in 2011, where he worked on the personalization of EPG (Electronic Program Guides). He was involved in various national and international research projects that dealt with natural language processing and recommender systems. From 2016 to 2019, he acted as project leader for a funded project regarding Semantic Holistic User Modeling for Personalized Access to Digital Content and Services. Since 2009, he published around 70 scientific articles in top venues and journals. He obtained the most inspiring contribution award at UMAP 2013, and he got a Best Paper Nominee at RecSys 2016. Finally, he regularly acts as a PC member on several top-tier conferences and co-organizes or co-chairs a number of workshops. Recently, he co-organized RecSys workshops about new trends in content-based recsys (2016), UMAP workshops about Holistic User Modeling (2017, 2018 and 2019) and UMAP Workshop on Explainable User Modeling (2020). He is one of the authors of the textbook "Semantics in Adaptive and Personalized Systems: Methods, Tools and Applications," edited by Springer.

Marco Polignano is a postdoc research fellow at the Department of Computer Science, University of Bari Aldo Moro, Italy, in the SWAP (Semantic Web Access and Personalization) research group. He earned a Ph.D. in computer science and mathematics in 2018, at the same university, with the thesis titled "An affect-aware computational model for supporting decision-making through recommender systems." He was a program committee member and reviewer for many journal and international conferences, the local organizing committee for the Ai*iA 2017 and CLiC-it 2019 conferences, organizer of the Evalita 2018 challenge—ABSITA about the aspect-based sentiment analysis and exUm 2020 Workshop at UMAP 2020 about user modeling and explanation. In 2016 and 2018, he was a Marie Sklodowska-Curie Research and Innovation Staff Exchange (MSCA-RISE) fellow, involved in the project N. 691071, titled "Seo-Dwarf: Semantic EO Data Web Alert and Retrieval Framework." His research interests include recommender systems, natural language processing, machine learning and user profiling.

Giovanni Semeraro is full professor of computer science at University of Bari Aldo Moro, Italy, where he teaches "Intelligent Information Access and Natural Language Processing," and "Programming languages." He leads the Semantic Web Access and Personalization (SWAP) "Antonio Bello" research group. In 2015, he was selected for an IBM Faculty award on Cognitive Computing for the project "Deep Learning to boost Cognitive Question Answering." He was one of the founders of AILC (Italian Association for Computational Linguistics) and on the Board of Directors till 2018. From 2006 to 2011, he was on the Board of Directors of AI*IA (Italian Association for Artificial Intelligence). He has been a visiting scientist with the Department of Information and Computer Science, University of California at Irvine, in 1993. From 1989 to 1991, he was a researcher at Tecnopolis CSATA Novus Ortus, Bari, Italy. His research interests include machine learning; AI and language games; recommender systems; user modeling; intelligent information mining, retrieval, and filtering; semantics and social computing; natural language processing; the semantic web; personalization. He has been the principal investigator of University

Journal : SmallExtended 11257 | Article No : 9272 | Pages : 35 | MS Code : 9272 | Dispatch : 7-7-2020

of Bari in several European, national and regional projects. He is author of more than 400 publications in international journals, conference and workshop proceedings, as well as of 3 books, including the textbook "Semantics in Adaptive and Personalized Systems: Methods, Tools and Applications" published by Springer. He regularly serves in the PC of the top conferences in his areas and is Program Co-Chair of CLiC-it 2019. Among others, he served as Program Co-chair of CLiC-it 2016, ACM RecSys 2015 and as General Co-chair of UMAP 2013. From 2013, he is the coordinator of the 2nd Cycle Degree Program in Computer Science at University of Bari. He is the coordinator of the 1st edition of the Master in Data Science at University of Bari. He is a member of the Steering Committee of the National Laboratory of Artificial Intelligence and Intelligent Systems (AIIS) of the National Interuniversity Consortium for Informatics (CINI) and of the Steering Committee of the ACM Conference Series on Recommender Systems.

Marco de Gemmis is associate professor at the Department of Computer Science, University of Bari Aldo Moro, Italy, where he received his Ph.D. in computer science in 2005. His primary research interests include content-based recommender systems, natural language processing, information retrieval, text mining and in general personalized information filtering. He authored over 100 scientific articles published in international journals and collections, proceedings of international conferences and workshops, and book chapters. He was program committee member for international conferences, including: ACM Recommender Systems; User Modeling, Adaptation and Personalization (UMAP), and served as a reviewer for international journals, including: User Modeling and User Adapted Interaction; ACM Transactions on Internet Technologies. He was invited speaker at several universities, including: University of Roma 3, University of Basque Country San Sebastian, University of Cagliari, University of Milano-Bicocca, University of Naples Federico II, and at Workshop on Semantics-Enabled Recommender Systems at ICDM 2016. He was Marie Curie Fellow in the SEO-DWARF project, funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 691071.

Pasquale Lops is associate professor at the University of Bari, Italy. He received the Ph.D. in computer science from the University of Bari in 2005 with a dissertation on "Hybrid Recommendation Techniques based on User Profiles." His research interests include recommender systems and user modeling, with a specific focus on the adoption of techniques for semantic content representation. He authored over 200 scientific articles, and he is one of the authors of the textbook "Semantics in Adaptive and Personalized Systems: Methods, Tools and Applications," edited by Springer. He regularly serves in the PC of the top conferences in his areas. He was Area Chair of User Modelling for Recommender Systems at UMAP 2016 and co-organized more than 20 workshops related to user modeling and recommender systems. He gave a tutorial on "Semantics-Aware Techniques for Social Media Analysis, User Modeling, and Recommender Systems" at UMAP 2016 and 2017; he was a speaker at two editions of the ACM Summer School on Recommender Systems. He was a keynote speaker at the 1st Workshop on New Trends in Content-based Recommender Systems (CBRecSys) at RecSys 2014. Finally, he gave the interview "Beyond TF-IDF" in the Coursera MOOC on Recommender Systems.