Modeling, Building and Management User Profiles: UPMS (User Profiles Management Service)

Taous Iggui
LIMED (Laboratory of Medical Computing), Science Computer Department,
University of A. Mira of Bejaia, Algeria.
iggui09gates@gmail.com

Hassina Nacer
Computer Science Department, University of Science and Technology-Houari Boumediene
USTHB, Algiers, Algeria
sino_nacer@yahoo.fr

Abdelkamel Tari
LIMED (Laboratory of Medical Computing), Science Computer Department.
University of A. Mira of Bejaia, Algeria.
tarikamel59@gmail.com

Abstract

Artificial Intelligence, Data Mining, E-Learning methods, Semantic Web technologies, all have as a purpose the satisfaction of user's needs. Nowadays, the user profile is considered as one of the most important constraint to take into account to perform the accuracy of the results returned by applications and computer systems *(personalization and recommendation systems)*. A user profile is a knowledge related to a specific user, it may be extracted from several sources having different format and, most of time, geographically distributed, this is why, this knowledge need to be collected, predicted, managed, filtered and represented in such away to be exploited. In this paper, we present a survey about the most important works done in this area, by dividing the profiling process into three major steps: modeling, constructing and management user profiles. We'll present the different techniques and methods used in: web, ubiquitous and cloud environment and we'll discuss some of the challenges and the future trends in the user profiling field.

Keywords: User Profiling, Semantic Web, Cloud Computing, Big Data.

1. Introduction

« Scientia potentia est » is a Latin aphorism claimed to mean «Knowledge is power », commonly attributed to Francis Bacon. Knowledge is the common relevant keyword we find in all computer systems. But *"What is Knowledge?","What is the difference between Knowledge and Information?*" and *"How can we get this Knowledge?"*. We know that data consists of facts, numbers (temperatures), texts, images and sound. When data is combined with interpretation and meaning, information emerges. Knowledge can be as simple as knowing who is the president of the United States, or it can be as complex as mathematical formula relating process variables to finish product dimensions. To distinguish between information and knowledge is not always straightforward (Jusoh et al, 2009). (McCallum, 2005) defined knowledge as "a fluid mixed of framed experience, values, contextual information, but until people use it, it isn't knowledge". While (Alfawareh et al, 2008) use knowledge definition taken from (Redfearn, 2006), considering the primary elements of knowledge are concepts and relationships between concepts.

The improvement of the accuracy of applications and computer systems results and the satisfaction of the user's needs in term of *knowledge*. has always been a very complicated task, and one of the biggest challenges for researchers, cause it (knowledge) needs to be extracted, collected, filtered, validated, managed, predicted and need to be normalized in order to be represented in such away to be exploited, and, of course, need to be secured. Given the different formats of data, the lack of interoperability, the difficulty to ensure semantic interoperability and having data geographically distributed in addition to the emergence of new IT paradigms such as Big Data and Cloud computing, all these constraints make knowledge so complicated to be acquired.
A user profile is a knowledge related to a specific user (context, interests, preferences, etc), a decisive constraint taken into consideration and exploited by researchers to better guide the applications results, which allow the satisfaction of user expectations, and, in some cases, suggesting relevant recommendations.
User Profiling is a very interesting area, being at the intersection of Artificial Intelligence, Data mining and Learning methods, it has turned the heads of many researchers, and even many science-fiction producers and screenwriters, giving birth to several Hollywood movies such as: **Eagle Eye**[1], **Enemy of the State**[2], **RoboCop**[3], **Person of Interest**[4], etc. Nowadays, user profiling is a fact and so vital in many areas in which it is essential to obtain knowledge about users of software applications (Schiaffino et al, 2009), such: e-recruitment, e-commerce, criminal domain, adaptive an recommendation systems, information retrieval, education, security, e-learning, (Soniya et al, 2013; kaczmarek et al, 2010; Alguliev et al, 2014; Jeong et Choi, 2012; Satler et al, 2010), as noticed in (Kanoje et al, 2014): personalization systems (Bedekar et al, 2008), (Doan et al, 2002), adaptive website, adaptive web stores and customer, e-tourism base website (Ouanaim et al, 2010), e-recruitment (Bradley et al, 2000), decision systems (Aquino and Filgueiras, 20) and many others.

Users interact with lots of new applications every day, running on different platforms in different environments, requiring different information, having different goals and providing different services. In recent years, much effort has been put into the discovery and representation of the user's basic features (Salter et al, 2010). Web, Ubiquitous and even on Cloud computing, many systems have been proposed to collect, organize and manage users information by analyzing their different interactions with several applications and web services, exploiting different modeling technologies and knowledge management methods.
Many works have done on the user profiling, but the lack of surveys that we noticed during our previous work in this area (Iggui et al, 2016) led us to present, in this paper, a complete and comprehensive one about the most important works, approaches and mechanisms developed for user profiling. To do this, we divided the profiling process into three major steps: modeling, constructing and management user profiles, we discuss the different techniques and methods used according to each step. In the second section, we present the new IT paradigms: Cloud and Big Data, and we propose an architecture for managing user profiles on the Cloud computing. A such survey is a very powerful tool since by knowing the characteristics of the existing approaches, it will be possible and easier to propose new and better mechanisms to enrich and perform the user profiling area.
The reminder of this paper is organized as follows: Section 2 present the related works divided into three subsections: modeling, constructing and management user profiles in the different environments (Web, Ubiquitous, Cloud), while a discussion of the different methods will be presented in Section 3. Section 4 will present the most important notions about Cloud computing and Big Data. Then, we illustrate the different modules of our architecture for user profiles management. Finely, we conclude the paper by Section 5.

2. Related Works

Table.1 shows the classification we adopt to present an overview of the main works done in the user profiling area.
In this section we divide the profiling process into three major steps: modeling user profile, building user profile and management user profile, as follows:

Table 1. Name of the table

Management User Profiles	Modeling User Profile
	Building User Profile

[1] **Eagle Eye** is 2008 American science fictional technology action thriller film directed by D. J. Caruso, produced by Steven Spielberg.
[2] **Enemy of the State** is a 1998 American spy-thriller directed by Tony Scott and produced by Jerry Bruckheimer.
[3] **Robocop** is an American science fiction directed by Paul Verhoeven.
[4] **Person of Interest** is 2011-2016 American science fiction crime drama, directed by Jonathan Nolan.

2.1 Modeling User Profiles

In recent years, much effort and several works have been done on the representation of the user's basic features, many models have been proposed for modeling user information in order to answer his Needs, Requirements and Desires (NRD) (Salem and Rauterberg, 2004) on a preferred system (Liu et al, 2009). The user profile is used to eliminate the ambiguity, guiding the system's actions and to improve the quality of user's interactions (Kay, 2001). We can divide user information into two categories : static and dynamic information (Iggui et al, 2016) which varies from one application domain to another, from simple *personal demographic information to* complex information *as background knowledge, interests, preferences, goals, behaviors, user's individual characteristics and user context, etc (Schiaffino et al, 2009) (Liu et al, 2009). This content depends on the application field, which is strongly linked to mechanisms and tools used for modeling.*
We distinguish from literature several techniques for user modeling enabling the resolution of many challenges in different environments (web, ubiquitous, cloud), such as heterogeneity, semantic interoperability, supporting scalability and expressiveness, Etc. We illustrate in what follows the most used by recent researchers: (Kuflik et al, 2012) (Schmohl and Baumgarten, 2008) (Strang et al, 2004) (Vanathi et al, 2011):

- *Keywords-based models: known by Bag-of-keywords and vector's representation*, the most simple and common model, considered as a basic and representation, the user is represented by a set of terms or concepts, possibly weighted (Kuflik et al, 2012) (Malone et al, 1987). Generally, the keywords are extracted from consulted document and text sources and weighted using TF/IDF (Manning et al, 2008). The concepts represent user's interests and the weights represent the degree of interest. Values can also be binary (0 or 1), to indicate behaviors such as purchase or not, clicked or not or they can be integers (Hafez and Xu, 2013). In (Alguliev et al, 2014), the authors worked on detecting a masquerader in the cloud environment. In order to discover an illegal attack, a behavior profile is created for each user as a vector taking into account the user's features and his interests, the normal behavior of a user is modeled as a cloud model, then the other models are compared and the deviations from this behavior are evaluated, if a deviation value is above the threshold, the user trying to get access to the system is evaluated as *an illegal user*. In (Tayari et al, 2009) authors proposed a multi-dimensional model based on an algebraic representation of user emotions in a Victor space, where emotions were represented by vectors.

This kind of classic representations is most of time used in Information Retrieval (IR), specially known to be chosen for the modeling of both static and dynamic user information from explicit and implicit sources. (Liu et al, 2009).

- *Markup scheme models:* They consist of hierarchical data structures based on markup tags including attributes and comments. Usually implemented as derivatives of SGML[5] such as XML[6]. In some cases, markup scheme models are used to describe context as extensions of Composite Capabilities / Preferences Profile (CC/PP)[7] and User Agent Profile (UAProf)[8] to cover the high dynamics of contextual information (Schmohl & Baumgarten, 2008). Ensuring the interoperability and being portable, several works of modeling user profile are based on structured languages, as (Abel et al, 2011) based on the RDF[9] graph which connect Twitter posts with articles, the authors represented the user as a combination of two profiles: Topic-based profile which include user's interests (sports, politics, etc) and Entities-based profile representing user's interests as persons, organizations or events. (EARL, 2006) based on XML for Modeling user emotions. (Jeong and Choi, 2012) where the authors proposed an XML-based model, considering user information, services and device information, for user authentication on Cloud computing.
- *Ontology-based approaches*: An ontology is a specification of a conceptualization (Gruber, 2008) and thus provides a mean for the organization and management of knowledge in such a way that is shareable and reusable among the applications of different domains. Once the knowledge is organized in the form of ontology, inference capability further provides the facility to extract implicit relations from user interests. Ontology is a powerful tool to represent and share knowledge about user, group of users or even a domain (Han and Chang, 2002) (King et al, 2007). For this, researchers, in many works, use different basic reference

[5] Standard Generalized Markup Language (iso 8879:1986). http://www.iso.org, June 2017.
[6] eXtensible Markup Languages.
[7] Composite Capabilities/ Preferences Profile (CC/PP). http://www.w3.org/Mobile/CCPP/, June 2017.
[8] User Agent Profile. http://www.wapforum.org, June 2017.
[9] RDF : Ressource Description Framework.

ontologies as FOAF ontology, Wordnet Dictionary, ODP (Daoud and Tamine, 2008), Open Directory and even Wikipedia to categorize concepts, which facilitates a vast number of different applications to semantically interact with the same knowledge base (Wasim et al, 2011). The user's interests and background knowledge are represented by concepts or topics as in (Michelson et al, 2010) using Wikipedia to disambiguate the entities convoyed by user's tweets and to map them to the categories that define the topic profile, and (Umamaheswari and Patil, 2012) where the knowledge (interests) is extracted from *stored documents, browsed web pages and composed/received emails.* Indicated as the most expressive model (Strang et al, 2004) (Couto et al, 2005), ontologies are used in many fields such as Web Intelligence, Recommendation Systems and, of course, Information Retrieval.

- *Graph-based Models*: Graphical models consist to represent entities and their relationships graphically, many modeling languages have been developed to this issue such as Unified Modeling Language UML[10] and Object-Role Modeling ORM[11] which represent the user and his context by identifying facts and enriching those by types and roles (Schmohl & Baumgarten, 2008). Within the scoop of APMD[12] project (Personalization Access to Mass of Data) (Bouzeghoub and Kostadinov, 2005) (Kostadinov, 2005) discussed and proposed a user profile modeling approache for knowledge management and for adapting the execution of queries in order to have better personalization results. The authors express the user preferences as a combination of user information (interests, personal data, security data) and context information in a generic and extensible mutli-dimensional meta-model providing a support for a profiles management platform, enabling the import/export of user profiles, profiles specialization/instantiation, the matching and the evolution of user profiles. To deal with the ambiguity generated from different user context sources, (Sklab et al, 2013) inspired by (Baldauf et al, 2007) (Chaari, 2007) (Dey et al, 1999), proposed a generic user context model, instantiable in different domains in pervasive environments (e-business, medical field) using graphic formalism, logic ontologies for a semantic description and XML for storage. We find also (Kankainen and Parkinnen, 2001) who proposed a graphical representation of User Profile (GUP) for telecommunication services, using the same idea of (Jonassen et al, 1993), where the authors take account user's background, user's context, his life sector and images.
- *Machine Learning based models*: The user model is developed based on training data, invisible for users. This approach is not used much as it requires extensive data for modeling user interests. Also, the dynamic nature of user interests requires a continuous learning process which is commonly not the case in machine learning methods of user interest learning (Michelson et al, 2010).
- *Logic-based-models:* represent highly formal modeling approaches, based on logics which define conditions on which concluding expressions on facts may be derived from sets of other expressions or facts. Those conditions are described by rules in a formal system, so that the facts, expressions and the rules put together define the context (Schmohl and Baumgarten, 2008).
- *Object-oriented-models:* Consist of encapsulating contextual information into objects. (Schmohl and Baumgarten, 2008).

2.2. Building User Profiles

User profiling has become present among the majority of information systems, the motivation of building user profiles is that users differ in their preferences, interests, background and goals when using software applications. Discovering these differences is vital to providing users with personalized services (Schiaffino et al, 2009). The construction of a user profile depends on the information sources and the application field. (Kanoje et al, 2014) presented a survey about trends and user profiling techniques, the authors identified three types of user profiling approach according to the implication of the user in the profiling process. a) *Explicit approaches*: based on the information delivered by the user, usually through Web registration forms, questionnaires, or specially designed psychometric instruments. But revealing their information to every information system became boring. A vast majority of users are reluctant to provide explicit information about their preferences and interest areas (Salter el al, 2010). So researchers presented b) *Implicit approaches*: based on user's behaviors and taking into account the dynamicity constraint, many methodologies, techniques and learning methods have been proposed in literature to predict information about the user and his needs, as explained by (Poo et al, 2003). c) *Hybrid approaches*: this type combined the advantages of both previous approaches.

[10] Unified Modeling Language, http://www.uml.org, June 2017.
[11] Object Role Modeling, http://www.orn.net, June 2017.
[12] ACI Masses de Données : Projet MD-33 Accès Personnalise à des Masses de Données. http://apmd.prism.uvsq.fr/

This part of our survey will be more focused on *implicit approaches* where the user profile is constructed by an unsupervised profiling system. Inspired by (Iggui et al, 2016), (Kanoje et al, 2014) and (Jie et al, 2010) we can divide the user profiling process into four steps:

1) *Collecting data*: the extraction of user information from different sources as extracting data from web Jie Tang (Jie et al, 2010), from social media (Li et al, 2014), the analysis of user behaviors, we take the example of (Hijikata, 2004) whom traced the user's mouse operation to extract text parts which the user might be interested on from text of web pages. Some of the main operations are: text tracing, link pointing, text selection, scrolling, saving, printing, window movement, and window resizing. Many works used TF/IDF measures, ontologies or dictionary to identify relevant information such key-words, named entities, etc.
2) *Processing data:* after the extraction of relevant information, some collected data may be duplicated or incomplete, so some techniques are proposed to integrate user profiles and solve the disambiguation problem as Hidden Markov Random Field (Jie et al, 2010), machine learning, ontologies and NLP techniques (Natural Language Processing) (Iggui et al, 2016).
3) Some works (Poo et al, 2003), (Stuart et al, 2003) classified users into groups according to their interests to better guide the personalization process. This step is more considered as a management task, it will be discussed in management user profiles section.
4) *Updating profiles:* this step come to deal with the evolution characteristic of the user profile. In order to maintain a fresh data and qualitative user profiles (Shmueli-Scheuer et al, 2010). This step aims to update the user information (Iggui et al, 2016).

The construction of user profiles is at the intersection of many different fields using different methods, from the simplest to the most complex one, according to the type of data and information sources: *Text information, videos, images and behaviors, etc.*

The main part of user information is extracted from textual format where the user is expressing himself or showing his interest on. We found the work of (Shmueli-Scheuer et al, 2010), the authors proposed a scalable framework where user profiles are represented by the textual content using vectors of Bag of Words. The authors used on Kullback- Leibler (KL) divergence (Kullback et Leibler, 1951) on users blog content and implemented their approach on top of Apache Hadoop MapReduce, interesting results were obtained but lot of information is lost using keywords and the approach didn't take account the semantic constraint.

To take into account this later, many works exploited the power of ontologies such as (Salter et al, 2010) within AGORA project, authors used a set of statistical measures combined with the algorithm presented by (Widyantoro and Yen, 2001), enabling the representation of user preferences with a set of ontology concepts. (Lau et al, 2009) build a fuzzy domain ontology from text's concepts (blogs, emails and so on). (Han and Chen, 2009) proposed a fuzzy clustering method called FCOU allowing to the user to be represented by one or more ontology, enabling to have enriched profiles which will be more helpful to the personalization process as in (Umamaheswari and Patil, 2012), where the ontology is built by extracting global knowledge from the LCSH system (The Library of Congress Subject Headings) and user background knowledge from user local instance repositories (LIR) such as *his stored documents, browsed web pages and composed/received emails.* (Sieg et al, 2007) improve the quality of the user profiles by assigning scores to the ontology of concepts extracted from the consulted documents according to the user behaviors.

Effectively, user's behaviors are very indicative while building his profiles. Several works have done on user's behaviors, studying and analyzing them in order to get more information about this one. According to (Morita and Shinoda, 1994) software users spend more time reading those articles they are interested in, while authors in (Oard and Kim, 1998) investigates another kind of behaviors more relevant such as printing, saving, scrolling and bookmarking, they consider that if a user has these behaviors when browsing a document, they must have some interests in it. This idea was adopted by (Xing et al, 2011) while proposing (BUIE) an algorithm for User Interest Extraction based on his Behavior. The user behaviors can be observed, also, during his interaction with multimedia services (Cloud), to get the content of user's interest instead of browsing for it. The authors in (Stupar et al, 2014) use the DTV information stored in the content provider database to construct a viewing history (event, how long watched) to generate a user profile able to personalize the content provider. The approach could be more efficient if authors used annotations to take into account the degree of interest of the user about each event.

The exploitation of user behaviors still suffers from some limitations. (Hoebel et al, 2006) presenting Gugubarra project, Gugubarra 2.0 which is an improvement of Gubugarra 1.0 (Mushtaq et al, 2004) (Hoebel et al, 2006) illustrated some limited issues, such as the duration while visiting a web page is somehow tricky cause we can't be sure about what the user is exactly doing in front of his screen, so in (Hoebel et al, 2006) the researchers allowed web sites owner, within the scoop of user profiling, to break down the pages into zones and ask the users for their

interests, but the majority of time users find annoying having always to answer these questions so they gave answer which don't reflect his real interests.

Social Media, blogs, posts, electronic conversation (Nicoletti et al, 2012) and emails messages are an important source of information which convoy relevant information about users. To exploit this type of data, different methods are proposed to extract semantic content, from texts. (Jusoh et al, 2009) proposed a framework to extract Named Entities (NE) using theories of probability, fuzzy grammar rules and predicates to determine the real semantic of a sentence. In our previous work (Iggui et al, 2016), we have presented a set of information extraction methods used in different fields: text summarization field and Information Retrieval (IR), we took the advantages by combining the formal power of statistics metrics, the semantic power of ontologies and techniques of NLP to better detecting user preferences, personal and professional data, etc, proposing *PBS (Profile Builder System)* for the construction of user profiles from email's job application, in e-recruitment area.

In some cases, information sources are limited such Twitter where user's tweets are limited to 140 characters (Michelson et al, 2010), blog titles (Oka et al, 2006), etc, so their semantic is difficult to detect, so researchers are using external databases such Wikipedia to enrich the user profile by disambiguating entities. (Abel et al, 2011) inspired by the observation of (Kwak et al, 2010) saying that 85% of the tweets are related to news, proposed an approach to enrich the user profiles using the Library of (Kohlschutter et al, 2010) and the web services provided by OpenCalais[13] to better extract NE, identifying organizations, events and topics of interests convoyed by the user tweets.

2.3. Management User Profiles

The diversity of systems, services and applications allows users to interact differently, allowing them to express themselves through different profiles, with, sometimes redundant information and sometimes incomplete one. To enrich users' profiles and to facilitate users' interactions with various information systems a management mechanism is required.

Management of user profiles comes to address many technical challenges using different methods including building and modeling techniques, Fig.1. One of the most important goal of this axe is ensuring interoperability between the different applications and services from different environment. **Interoperability** is the ability of two or more systems or components to exchange information and use the information that has been exchanged (Geraci, 1991). It refers to the ability to access and interpret information derived from multiple heterogeneous sources. It can be (i) syntactic, where the differences between the models are resolved at application level, or (ii) semantic, where the differences between the models are resolved at the level of knowledge (Carmagnola and Cena, 2009).
(Wagner et al, 2013) proposed an ubiquitous profile manager *eProfile*, of generic domain oriented towards distributed environments, with inter-system **interoperability**. eProfile, Fig.1, infers profile data through analysis of an entity trails (provided by a manager tracks, it returns all events held by an entity), following rules defined by the entities, and provides data inferred to be used by applications. eProfile's main characteristics are the use of trails for extracting profiles, the capability of generating dynamic profiles, the capability of managing dynamic inference rules and models, and the semantic operability of the model, since eProfile supports multiple applications. As an application example, it allows maintaining the student dynamic profiles updated, allowing a better adaptation to the ubiquitous learning environment.

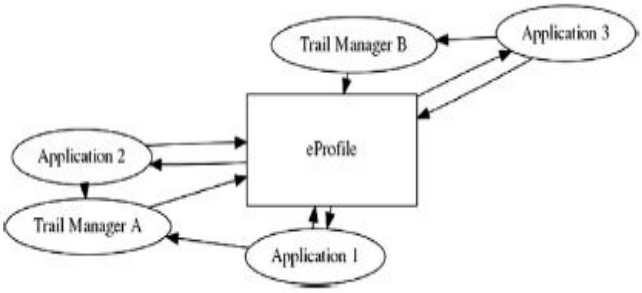

Figure 1. eProfile connected to multiple trail managers. (Wagner et al, 2013)

Some other challenges are illustrated in (Ghosh el al, 2009) (Ghosh et al, 2008) such as:

[13] http://www.opencalais.com

- *Profile sparseness known as the cold-start problem:* having users don't like to repeat their information for each service they're interacting with, in some cases, when they're interacting with a new service for the first time, this one has extremely sparse information about them.
- *User's personae* which is related to the influence of his context and environment on his behaviors and preferences.
- The *using out-of-band* data in other words, the problem of using different fragments of user's interactions or preferences from different services and applications.
- The *overload of servers* due to the storage of different profiles belonging to the same user.

Many works and tools have done and developed to meet the challenges mentioned above; such as: Google OpenSocial[14] and Yahoo Y!OS[15] .

From works we have studied, we noted different and several management techniques, such as matching, aggregation, negotiation and selection, etc, where the goal is to select the most adequate profile or a set of profiles for a given situation.

Aggregation concept is found in many works such (Ghosh et al, 2009) where the authors proposed a user profile management framework, which is a part of HP Labs, InfoPal Profile Manager (IPM) which deal with the *out-f-band profile information*. With the emergence of services exporting their data through standard APIs (Ghosh et al, 2008), it's possible to provide facilities to dynamically discovering and building a suitable user profile from profile fragments distributed across multiple services. The authors in this work use an agent to build a corpus of user profiles, maintained by different services, each one described with a specific ontology. The mediator calculate the semantic distance between the profiles of the corpus and the specification provided by the requesting service and according to a threshold profiles are used to build a user profile for this service. Being able to determine which information is the most relevant resolve the sparseness problem. In addition, IPM supports the *user profile personae*, allowing to users to store different profile personae and choose which personae to forward to a service. The disadvantage with this approach is that explicit point-to-point data sharing schemes and agreements are needed by the different services.

Within the same scoop, we found (Morikawa et al, 2005) proposing a Profile Aggregation (PA) as a node of collecting various profiles from diverse sources and managing them in the form of "a Personalized Profile" with a privacy control mechanism able to control the type of data that can be shared, with whom and in what situations, using the Resource Description Framework (RDF) (Beckett, 2004) for better semantic.

Social Media, is another interesting area, who has a great impact on people around the world. It's not uncommon that individuals create multiple profiles in several social networks (Facebook, Twitter, LinkenIn, etc), each containing partially overlapping sets of personal information. Matching those different profiles allows to create a global profile that gives a holistic view of the information of an individual. As goal, the management of user information, in (Bennacer et al, 2014) researchers proposed an algorithm to mach profile across n social network, they exploit the links between profiles and the values of the profile attributes, to determine the pairs that match. Inspired by the observations of krishnamurthy et al, who identified a set of common attributes available in 12 of the most important social network (Krishnamurthy et al, 2009) (such as names, family names, usernames) the authors in (Bennacer et al, 2014) focus on username, name, email and links to other Web pages.

In medical area, for a better quality of healthcare services, (Fengou et al, 2012) use an electronic Medical Health Record (eMHR) to deliver a personalized ubiquitous system. Different entities are related to the patient (patient's social network, healthcare center, smart home, patient's office, patients, doctors, member's family), in order to provide high quality and reliable personalized healthcare services to the patient when his health condition is critical. For example each patient is related to a group of indexed profiles related to his health condition, in a critical case, a User Healthcare Profile is created, and the profile of the most adequate healthcare center and doctor are **selected** and recommended.

In (Cherrat et al, 2015) the researchers proposed a data extraction system for user profile management based on behavior for developing a security and surveillance system for drivers. Having a strong connection between fatigue, sleep and poor driving behavior and errors that lead to accidents, the authors proposed an intelligent system to manage all bad driving behaviors to prevent accidents and dangerous situations. The authors used PCA method (Freund and Schapire, 1996), and Viola and Jones algorithm (Viola and Jones, 2001) for detecting objects in digital images, special sensors to monitor the driver and the vehicle, and web services to put in touch the drivers with police officers.

[14] Google OpenSocial, code.google.com/ opensocial.
[15] Yahoo Y!OS, developer.yahoo.com/yos

In Cloud environment, a Knowledge-as-a-Service, according to (Li et al, 2009) is indispensable to human learning, working and living, and Cloud computing, as a new emerging technology, is providing all kinds of scalable services enabling the creation, utilization/reutilization and the sharing of personal knowledge as proposed by (Tsui et al, 2011) a Cloud-Based Personal Knowledge Management (CBPLM). (Jeong et al, 2012) within the scoop of mobile cloud computing propose a context-aware-based intelligence platform for managing resource using context-aware information. They propose an information aggregation process using a generic ontology. They also use time and location constraints for resource recommendation.

3. Discussion

We have tried above, to illustrate the most important works in the user profiling area presenting each step a part (modeling, building and management user profiles). We noticed the advantages and limitations to take into account while proposing a user profiling system. We come out with some points, such as:

- The use of ontologies is indispensable due to their semantic power, but we have to say also that a user profile is more than a set of key-words or concepts (Iggui et al, 2016), this is why we recommend to combine ontologies with statistical measures and other formal methods.
- Some of the proposed profile structures have the disadvantage of being bounded by the types of information (attributes) defined by the model.
- Taking into account the user context and its dynamicity during the profiling process, being an important source of information able to guide the results of information systems. But we have to precise that a user context is not limited only to the devices information as it's considered by some works.
- The user location, a component of the user context, is rarely taken into account while it can guide the preferences, kind of researches and even the activities of a given user, for example we're not doing the same thing in the office as in a cafe.
- User profiles information have always to be accessible for up dating, in order to maintain the freshness of data, and so to ensure a better services quality.
- Users tend to prefer unsupervised approaches, this is why the "Interoperability" is one of most important constraint to consider while modeling or managing user profiles, in order to export and to exploit the constructed profile by different services and applications. We mentioned XML as one of the standards technologies used for interoperability by several works.
- As a last point, we can't deny the fact that each user can have several profiles with different information, sometimes incomplete, in some cases complementary in some others redundant. The redundancy is a very big challenge for many researchers specially on Cloud environment where each data is replicated on several servers (*servers overload problem*).

In order to enrich the user profile, to facilitate its interactions with the different information systems (to not provide his information each time), to minimize the overloading of servers, in order to propose a solution to the cold start problem and as a fruit of our present survey, we propose in what follows a User Profiles Management System having as a goal the Construction of a Global User Profile *(UPMS-CGUP)* from a set of local user profiles belonging to social Cloud.

4. User Profiles Management, Cloud and Big Data

4.1. Cloud an Big Data

Cloud computing is a new IT paradigm, no common definition exists yet (Voas and Zhang, 2009). We adopt the definition of the National Institute of Standards and Technology (NIST) (Mell et Grance, 2011): Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

With the rise of Internet and the explosion of data sources, more and more companies are facing new challenges that were previously very rare. (Sakr et al, 2011) present a comprehensive survey about data management mechanisms, open issues and future challenges in Cloud environment. One of those problems concern the storage and the usability of sets of data, so big and often growing so fast, that the usual tools were no more adapted. (Thibault, 2011) discussed the storing data problem and the elasticity of Cloud databases by presenting a comparison between Relational Database Management System and NoSQL databases, and how this later faces the new challenges. Those new, huge, sets of data are commonly called "Big Data", these data may belong to the different services and applications, specially emerged with Cloud computing. In this section, data refers to user's information. Users

interact with different Cloud services witch ask for different information, this generates several local profiles on different applications, belonging to the same user. Taking the example of social media (Facebook, Twitter, LinkedIn, etc), a user can have different profiles on each application, with different and/or redundant information. In order to remedy, mainly, the problems of **cold start**, **interoperability**, **data storage and servers overloading**, for more data **security**, better **management of digital identities**, facilitate **authentication processes**, as well as better **satisfying the user's needs**, we propose in the following a user profile management platform.

4.2. User Profiles Management Service

The diversity of services and applications allows to users to interact with several systems a day, expressing himself differently through various profiles with redundant, sometimes, incomplete information. Managing user profiles is a very important mechanism, in order to identify the relevant information of an individual and to facilitate its interactions with the different information systems (to not provide his information each time), to better improve the different services results and for a better quality of personalization, to enrich the user's profile by relying on the different interactions of this one with different systems and applications, to reduce the number of profiles belonging to the same user, so dealing with redundancy and the servers overload problem. To deal with the cold start problem and to enable interoperability between applications we propose a UPMS (User Profiles Management Service).
Inspired by (Bennacer et al, 2014) and (Kostadinov, 2005), where semantic matching methods were proposed, we choose the social Cloud (social media) as an example, to propose a user profiles management platform having as a goal the construction of a Global User Profile. The global user profile will include all the information belonging to a given user. It will be stored at the Cloud level and thus accessible from all Cloud applications and services. The corresponding architecture of the proposed system is shown in Figure 2, using HP Cloud Architecture:

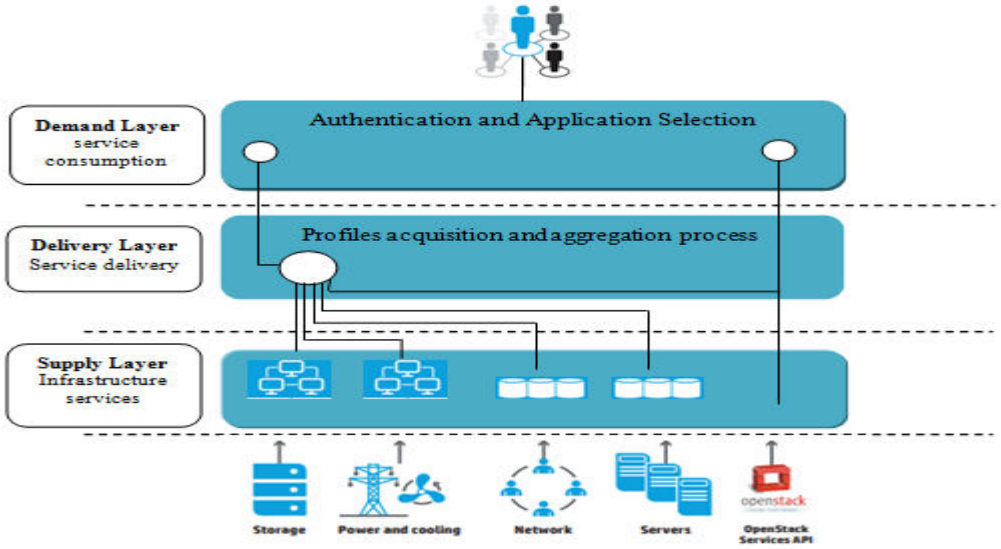

Figure 2. User Profiles Management Service (UPMS) presented on HP Cloud Architecture.

Our approach is based on the exploitation of local user's profiles on the different applications of social media. Bellow, we detail the different steps of the proposed approach:
Hypothesis: We suppose, in our case, that: All users' profiles are created with the same e-mail address and the language used is "English".

1) **Authentication and Application Selection**: According to the choice of the user in this step, our system may be supervised or not supervised by this later:
   - ✓ *Supervised system*: in order to respect the user's privacy, the system enables the user to choose the application from which his global user profile will be constructed, and to give a relevance degree to each one.
   - ✓ *Not supervised system*: in this case, the system will consider all the sub-profiles related to the given e-mail address.
2) **Profiles Acquisition and aggregation process**: In this step the system get access to the sub-profiles belonging to the chosen applications via the corresponding APIs. Once they're recovered, a semantic alignment and filtering processes are launched.

- ✓ ***Semantic Alignment***: it consists in establishing a semantic matching between the lexicons of tow, or more, sub-profiles using ontology.
- ✓ ***Information Filtering:*** consists of selecting the relevant information (values) according to the relevance of their sources and the nature of the attribute (Mono-valued or Multi-valued).

5. Conclusion and Future Trends

This paper is one of the most complete work carried out on the user profile, indeed, we have presented a survey that deals with the three major axes of user profiling across the different environments, we have presented the most responded methods used by researchers in the fields of modeling, building and managing user profiles. We also illustrated a set of details and points to be taken into consideration during the development of a user profiling system. In the second section of this paper, we took a closer look on the user profiles management field, taking into account the emergence of the Cloud and the exponential growth of Big Data. As a result of our researches, we have presented a new profiles management system, enabling the construction by aggregation of a global user profile in the domain of social Cloud. Our system allows us to:

- Have a better interoperability using the standard technology XML.
- Have a better respond to the needs of users, by using richer user profiles in personalization and recommendation systems.
- Facilitate the user's interaction with the new applications, for example, his information may be recovered from the global user profile, so he is not obliged to enter his information over and over.
- Give a free choice to the user, to choose the contributors applications for the construction of his global profile.

The goal of this paper is more than presenting the results of our profiles management service, but to provide and make available for researchers a kind of reference in the field or user profiling area. We have illustrated through a set of works, the challenges to be faced, in addition a set of methods and technologies used for.
The results of our approach will be provided in our future paper, adding some other enriching features and taking into account other constraints, such as:

- Taking into account the user's reputation, in order to better judge the credibility of the provided information.
- The detection of false information, inspired by Artificial Intelligence, by analyzing the information communicated through his communities.
- The usage of different data model and different DBMS (Data Base Management System) by the applications of social Cloud (Cassandra by Facebook, PNUTS by Yahoo, Manhattan by Twitter) is another challenge to face in our future works; we will try to propose a generic user model for the representation of the global user profile, taking account Big Data constraints, for more ***interoperability*** between Cloud applications and services.

References

(Jusoh et al, 2009): Jusoh, S., Fawareh, H. M. Semantic Extraction from texts. International Conference on Computer Engineering and Applications IPCSIT vol.2, Singapore, 2009.

(McCallum, 2005): McCallum, A., Information extraction: distilling structured data from unstructured text, ACM *Queue, vol.* 3, no. 9, pp. 48–57, 2005.

(Alfawareh et al, 2008): AlFawareh, H. M., Jusoh, S., and Osman, W. R. S., Ambiguity in text mining. *International Conference on Computer and Communication Engineering (ICCCE'08)*, pp. 1172–1176, Kuala Lumpur, May 13-15, 2008.

(Redfearn, 2006): Redfearn, J., Text mining, JISC, pp. 1–2. 2006

(Schiaffino et al, 2009): Schiaffino, S., and Amandi, A., Intelligent User Profiling. Artificial Intelligence, LNAI 5640, pp. 193 – 216. Berlin Heidelberg , 2009.

(Soniya et al, 2013): Soniya, M., and Kalpana, B., Performance Analysis of Ranking Algorithms Using Ontology for Web Personalization. IJAIR, Vol.2, Issue 3, 2013.

(Kaczmarek et al, 2010): Kaczmarek, T., Zyskowski, D., Adam, W., and Witold, A, Information extraction from web pages for the needs of expert finding. Studies in Logic, Grammar and Rhetoric. Vol. 22(35), pp. 141-157, 2010.

(Jeong et Choi, 2012): Jeong, H., Choi, E, User Authentication using Profiling in Mobile Cloud Computing. In proceeding (2) AASRI Conference on Power and Energy Systems. Pp. 262 - 267. Hong Kong, Septembre, 2012.

(Salter et al, 2010): Satler, M., F., Romero, F., P., Menendez, V., H., Zapata, A., and Prieto, M., E, A Fuzzy Ontology Approach to represent User Profiles in E-Learning Environments. International Conference Fuzzy Systems (FUZZ), IEEE, Barcelon, July, 2010.

(Alguliev et al, 2014): Alguliev, R., Abdullaeva, F., User Profiles and Identifing User Behaviour in the Cloud Computing Environment. Universal Journal of Communications and Network, vol. 2, no. 4, pp. 63-68. 2014.

(Kanoje et al, 2014): Kanoje, S., Girase, S., Mukhopadhyay, D., User Profiling Trends, Techniques and Applications. International Journal of Advance Foundation and Research in Computer (IJAFRC), vol. 1, no. 1, 2014.

(Bedekar et al, 2008) : Bedekar, M.V., Deshpande, B., Joshi, R., Web Search personalization by User Profiling. In Emerging Trends in Engineering and Technology ICETET'08, Nagpur, India, 2008.

(Ouanaim et al, 2010): Ouanaim, M., Harroud, H., Berrado, A., Boulmalf. M., Dynamic User Profiling Approach for Services Discovery in Mobile Environments. Proceedings of the 6th International Wireless Communication and Mobile Computing Conference IWCMC'10, pp. 550-554, New York, 2010.

(Bradley et al, 2000): Bradley, K., Rafter, R., Smyth, B., Case-Based User Profiling for Content Personalization. International Conference Adaptive Hypermedia and Adaptive Web-Based System, pp. 62-72, Berlin, 2000.

(Aquino and Filgueiras, 2005): Aquino, P.T.J., Filgueriras, L.V.L., User Modeling with Personas. Latin American Conference on Human-Computer Interaction CLIHC'05, pp. 277-282, Cuernavaca, Mexico, October 23-26, 2005.

(Iggui et al, 2016): Iggui, T., Nacer, H., Sklab, Y., Ait Radi, T., *Web Application for User Profiling*. International Journal of Information Systems in the Service Sector, IJISSS, vol. 8, no. 2, pp. 44-56, April – June, 2016.

(Salem and Rauterberg, 2004): Salem, B., Rauterberg. M., Multiple User Profiling Merging (MUPE): Key Challenges for Environment Awareness. Lecture Notes in Computer Science, vol. 4161, pp. 103-116. 2004.

(Liu et al, 2009): Liu, H., Salem, B ., Rauterberg, M., A Surevey on User Profile Modeling for Personalized Service Delivery Systems. Proceeding of IADIS International Conference on Interfaces and Human Computer Interaction, pp. 45-51, 2009.

(Kay, 2001): Kay, K,. User Modeling for Adaptation. User Interfaces for All: Concepts, Methods and Tools (Ed. Constantine Stephanidis), pp. 271-294, 2001.

(Kuflik et al, 2012): Kaflik, T., Kay, J., Kummerfeld, B., *Challenges and Solutions of Ubiquitous User Modeling.* In Kruger, A. and Kuflik, T. (eds.), Ubiquitous Display Environments, Congnitive Technologies, Berlin, 2012.

(Manning et al, 2008): Manning, C.D., Raghavan, P., Schutze, H., Introduction to Information Retrieval. Cambridge University Press, 2008.

(Hafez and Xu, 2013): Hafez, A. A., Xu, Y., A *Survey of User Modelling in Social Media Websites.* Computer and Information Science, vol. 6, no. 4. September, 2013.

(Alguliev et al, 2014): Alguliev, R., Abdullaeva, F., User Profiles and Identifing User Behaviour in the Cloud Computing Environment. Universal Journal of Communications and Network, vol. 2, no. 4, pp. 63-68. 2014.

(Tayari et al, 2009): Tayari, I., Thanh, N. L., Ben Amar, C., Modélisation des Etats Emotionnels par un Espace Vectoriel Multi-dimensionnel. Laboratoire d'Informatique de Signaux et Systèmes de Sophia Antipolis – UNSA – CNRS, Décembre 2009.

(Abel et al, 2011): Abel, F., Gao, Q., Houben, G.J., Tao, K., Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. 8th Extended Semantic Web Conference, ESWC 2011, pp 375-389, Greece, May 29 – June 2, 2011.

(EARL, 2006): Humaine Emotion Annotation and Representation Language (EARL), 30 June 2006.

(Gruber, 2008): Gruber, T., What is an Ontology, Encyclopedia of Database Systems, vol. 1, 2008.

(Daoud and Tamine, 2008): Daoud, M., Tamine-Lechani, L., Construction des profils utilisateurs à base d'ontologie pour une recherche d'information personnalisée. Journal of Laboratoire IRIT University of Paul Sabatier, 2008.

(Wasim et al, 2011): Wasim, M., Shahzadi, I., Ahmad, Q., Mahmood, W., Extracting and Modeling User Interests based on Social Media. Multitopic Conference (INMIC), IEEE 14th International, pp. 284-289. Karachi, 22-24 December. 2011.

(Michelson et al, 2010): Michelson, M., Macskassy, S.A.. Discovering Users' Topics of Interest on Twitter: A First Look. The 4th workshop ACM on Analytics for Noisy unstructured text Data, pp.73-80, New York, USA. 2010.

(Doan et al, 2002): Doan, A., Madhavan, J., Domingos, P., Halevy, A., Learning to Map between Ontologies on the Semantic Web, 11th International Conférence World Wide Web (WWW '02), pp. 662-673, 2002.

(Han et Chang, 2002): Han. J., Chang, K.C.C., Data Mining for Web Intelligence, Computer, vol. 35, no. 11, pp. 64-70.2002

(King et al, 2007): King, J.D., Li, Y., Tao, X., Nayak, R., Mining World Knowledge for Analysis of Search Engine Content, Web Intelligence and Agent Systems,vol.5,no.3,pp.233-253, 2007

(Umamaheswari et Patil, 2012): Umamaheswari, B., Patil, P., Personalized Ontology Model – Survey. The 12th International Conference on Hybrid Intelligent System (HIS), pp. 348-353, India, 4-5 December 2012.

(Strang et al, 2004): Strang, T., Linnhoff-Popien, C., A context modeling survey. Workshop on Advanced Context Modeling, Reasoning and Management associated with the 6th International Conference on Ubiquitous Computing (UbiComp), Nottingham, 2004.

(Couto et al, 2005): Couto, R., Rocha, A., Endler, M., Evolutionary and efficient context management in heterogeneous environments. Proceeding of the 3rd International workshop on Midleware for Pervasive and Ad-hoc computing, pp. 1-7, New York, USA, 2005.

(Schmohl & Baumgarten, 2008): Schmohl R., Baumgarten, U., Context-aware computing : a survey preparing a generalized approach. International Multi-Conference of Engineers and Computer Scientists, pp. 744-750, Hong Kong, March 2008.

(Bouzeghoub et Kostadinov, 2005): Bouzeghoub M., Kostadinov D., Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils, Conférence en Recherche d'Informations et Applications Complexe de Recherche Interprofessionnel en Aérothermochimie (CORIA), pp. 201-218, France 2005.

(Kostadinov, 2005) : Kostadinov. D., Personnalisation de l'information : une approche de gestion de profils et de reformulation de requêtes. Human-Computer Interaction. Université de Versailles, France, 2007.

(Sklab et al, 2013) : Sklab, Y., Nacer, H., Tari, A. K, Modeling User Context for Pervasive Computing. Proceedings of the International Workshop on Requirements Engineering, pp. 40-46, Constantine, Algeria, 2013.

(Baldauf et al, 2007): Baldauf, M., Dustdar. S., A survey on context-aware system. International Journal of Ad Hoc and Ubiquitous Computing, vol. 2, pp.263_277, 2007.

(Chaari, 2007) : Chaari, T., Adaptation d'applications pervasives dans des environnements multi-contextes, Laboratoire d'Informatique en Image et Systèmes d'information (LIRIS), September 2007.

(Dey et al, 1999): Dey, A. K., Salber, D., Futakawa, M., et al. An architecture to support context aware applications. Technical report, ACM Symposium, http://smartech.gatech.edu/jspui/bitstream/1853/3390/1/99-23.pdf , June 1999.

(Kankainen and Parkinnen, 2001): Kankainen, T., Parkinnen, J., GUP-Graphical Presentation of User Profile. In CHI Conference on Human Factors in Computing Systems, pp. 105-106, Seattle, Washington USA, 2001.

(Jonassen et al, 1993): Jonassen, H. D., Beissner, K., Yacci, M. Structual Knowledge, Techniques for representing, conveying and Acquiring Structural Knowledge, Lawrence Eribaum Associates, New Jersey, 1993.

(Malone et al, 1987): Malone, T., Grant, K., Turbak, F., Brobst, S., Cohen, M., Intelligent information sharing systems. Communications of the ACM , vol. 30, no. 5, pp. 390-402, 1987.

(Strang et al, 2004): Strang, T., Popien, C. L., A context modeling survey. In Proceedings of the 1st International Workshop on Advanced Context Modeling, Reasoning and Management, pp. 33-40, UK, 2004

(Vanathi et al, 2011): Vanathi, B. V., Uthariaraj, R., Context aware system for smart consignment tracking. European Journal of Scientific Research, vol. 50, pp.586-596, 2011.

(Poo et al, 2003): Poo, D., Ching, B., Goh, J.M., A hybrid approach for user profiling. Proceedings of the 36th Annual Hawaii International Conference on System Sciences, vol. 7, pp. 9-13, 2003.

(Jie et al, 2010) : Jie, T.L.Y., Zhang, D., Zhang, J., A Combination Approach to Web User Profiling, ACM Transactions on Knowledge Discovery from Data, vol. 5, no. 1, Article 2, New York USA, December 2010.

(Li et al, 2014) : Li, J., Alan, R., Eduard, H., Weakly Supervised User Profile Extraction from Twitter, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics ACL, pp. 165–174, Baltimore, USA, June 23-25 2014.

(Hijikata, 2004): Hijikata, Y., Implicit user profiling for on demand relevance feedback. Proceedings of the 9th international conference on Intelligent user interfaces, pp. 198-205, Portugal 2004.

(Stuart et al, 2003): Stuart, M.E., Shadbolt, N. R., Roure, D. C. D., Capturing interest through inference and visualization: Ontological user profiling in recommender systems. In Proceedings of the 2nd international conference on Knowledge capture, pp. 62-69. USA, 2003.

(Shmueli-Scheuer et al, 2010) : Shmueli-Scheuer, M., Roitman, H., Carmel, D., Mass, Y., Konopnicki, D., Extracting User Profiles from Large Scale Data. Proceedings of the 2010 Workshop on Massive Data Analytics on the Cloud. Raleigh, North Carolina, USA. April 26, 2010.

(Kullback et Leibler, 1951): Kullback, S., Leibler, R. A., On information and sufficiency. The Annals of Mathematical Statistics, vol. 22(1), pp. 79–86, 1951.

(Widyantoro et Yen, 2001): Widyantoro, D.H., Yen, J., Incorporating Fuzzy Ontology of Term Relations in a Search Engine. In BISC International Workshop on Fuzzy Logic and the Internet. Berkeley: University of California, 2001.

(Lau et al, 2009): Lau, R. Y. K., Song, D., Li, Y., Cheung, T. C. H., Hao, J. X., Toward a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning. IEEE Transactions on Knowledge and Data Engineering, vol. 21(6), pp. 800-813, 2009.

(Han and Chen, 2009): Han, L., Chen. G., A Fuzzy Clustering Method of Construction of Ontology-based User Profiles. Advances in Engineering Software, vol. 40(7), pp. 535-540, 2007.

(Sieg et al, 2007) : Sieg, A., Mobasher, B., Burke, R., Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search. IEEE Intelligent Informatics Bulletin, vol.8, no.1. November 2007.

(Morita and Shinoda, 1994) : Morita, M., Shinoda, Y., Information filtering based on user behavior analysis and best match text retrieval. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 272-281. Ireland, 1994.

(Oard and Kim, 1998): Oard, D.W., Kim, J., Implicit feedback for recommender systems. Proceedings of the AAAI Workshop on Recommender Systems. 1998.

(Xing et al, 2011): Xing, K., Zhang, B., Zhou, B., Liu, Y., School of Computer Engineering & Science Shanghai University, Shanghai, China. Behavior based User Interests Extraction Algorithm. IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing, 2011.

(Stupar et al, 2014): Stupar, G., Nad, D., Beserminji, A., Papp, I., One solution of STB Users Cloud Based Profiling. Published in Proceeding of the 4th IEEE International Conference on Comsumer Electronics (ICCE), pp. 364-366, Berlin. 7-10 September 2014.

(Hoebel et al, 2006): Hoebel, N., Kaufmann, S., Tolle, K., Zicari, R.V., The Design of Gugubarra 2.0: A Tool for Building and Managing Profiles of Web Users. Published in the proceedings of the International Conference on Web Intelligence IEEE/WIC/ACM (WI'06), pp. 317-320. Hong Kong, 18-22 December 2006.

(Hoebel et al, 2006): Hoebel, N., Kaufmann, S., Tolle, K., Zicari, R., The Gugubarra Project: Building and Evaluating User Profiles for Vistors of Web Sites, First IEEE Workshop on Hot Topics in Web Systems and Technologies, Boston, Massachusetts, USA, November 2006.

(Mushtaq et al, 2004): Mushtaq, N., Tolle, K., Werner, P., Zicari, R., Building and Evaluating Non-Obvious User Profiles for Visitors of Web Sites, IEEE Conference on E-Commerce Technology (CEC 04), San Diego, California, USA, 2004.

(Nicoletti et al, 2012): Nicoletti, M., Schiaffino, S., Godoy, D., Mining interests for user profiling in electronic conversations. Journal of Isistan Research Institute, Conicet-Unicien, University of Paraje Arroyo Seco, 2012.

(Oka et al, 2006): Oka, M., Abe, H., Kato, K., Extracting Topics From Weblogs Through Frequency Segments. Workshop of WWW'2006. Edinburgh, UK. May 22-26, 2006.

(Kwak et al, 2010): Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? The 19th ACM International Conference on WWW, pp. 591-600, New York, NY, USA, 2010.

(Kohlschütter et al, 2010): Kohlschütter, C., Fankhauser, P., Nejdl, W., Boilerplate detection using shallow text features. 3rd ACM International Conference on Web Search and Data Mining. pp. 441-450, New York, 2010.

(Geraci, 1991): Geraci, A., IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries, New Jersey, USA, 1991.

(Carmagnola & Cena, 2009): Carmagnola, F., Cena, F., User identification for cross-system personalization, Journal of Information Sciences, vol. 179, no. 1, pp. 16–32, 2009.

(Wagner et al, 2013): Wagner, A., Barbosa, J. L.V., Barbosa, D.N.F., A model for Profile Management Applied to Ubiquitous Learning Environments. In Expert Systems with Applications, vol. 41, pp. 2023–2034, 2014.

(Ghosh el al, 2009): Ghosh, R., Dekhil, M., Discovering Profiles. The 18th International Conference on World Wide Web, pp. 1233-1234, Madrid, Spain, April 20-24, 2009.

(Ghosh el al, 2008): Ghosh, R. Dekhil, M., Mashups for semantic user profiles. The 17th International Conference of World Wide Web, pp. 1229-1230, Beijing, 2008.

(Morikawa et al, 2005): Morikawa, D., Honjo, M., Yamaguchi, A., Ohashi, M., A Proposal of User Profile Management Framework for Context-Aware Service. In Proceedings of the Symposium on Applications and the Internet Workshops (SAINT-W'05) IEEE Computer Society. Italy, February, 2005.

(Beckett, 2004): Beckett, D., RDF/XML Syntax Specification,W3C, https://www.w3.org/TR/rdf-syntax-grammar/ February, 2004.

(Jeong et al, 2012): Jeong, H., Jeong, Y., Kang, Y., Kim, M., Choi, E., Efficient Method of Profiling on Mobile Cloud Computing. International Conference on Future Information Technology and Management Science & Engineering Lecture Notes in Information Technology, vol.14, pp. 177-182, 2012.

(Tsui et al, 2011): Tsui, E., Cheong, R.K.F., Sabetzadeh, F., Cloud-Based Personal Knowledge Management as a service (PKMaaS). Published in International Conference on Computer Science and Service System (CSSS), pp. 2152-2155, Nanjing, 27-29 June, 2011.

(Li et al, 2009): Li, L., Zheng, Y., Zheng, F., Zhong, S., Cloud Computing Support for Personal Knowledge Management, International Conference on Information Management, Innovation Management and Industrial Engineering, pp. 171-174. 2009.

(Bennacer et al, 2014): Bennacer, N., Jipmo, C.N., Penta, A., Quercini, G., Matching User Profiles Across Social Network, Springer International Publishing Switzerland, pp. 424-438. 2014.

(Krishnamurthy et al, 2009): Krishnamurthy, B., Wills, C.E.: On the Leakage of Personally Identifiable Information via Online Social Networks. The 2nd ACM Workshop on Online Social Networks, pp. 7-12, 2009.

(Weingartner et al, 2015): Weigartner, R., Brascher, G.B., Westphall, C. B., Cloud resource management: A survey on forecasting and profiling models. Journal of Network and Computer Applications, no. 47, pp. 99-106, 2015.

(Fengou et al, 2012): Fengou, M.A., Mantas, G., Lymberopoulos, D., Komninos, N., Ubiquitous Healthcare Profile Management Applying Smart Card Technoogy. International Conference on Wireless Mobile Communication and Healthcare MobiHealth: Wireless Mobile Communication and Healthcare, LNICST, vol. 83, pp. 248-255, 2012.

(Cherrat et al, 2015): Cherrat, L., Ezziyyani, M., ElMouden, A., Hassar, M., Data Extraction for User Profile Management based on Behavior. International Conference on Cloud Technologies and Applications, Morocco, 2015.

(Viola and Jones, 2001): Viola, P., Jones, M., Rapid object detection using a boosted cascade of simple features, In CVPR, pp. 511-518, USA, 2001.

(Freund and Schapire, 1996): Freund, Y., Schapire, R.E., Experiments with a new boosting algorithm, In Machine Learning, pp.148–156, 1996.

(Mell et Grance, 2011): Mell, P., Grance, T., The NIST defnition of cloud computing, National Institute of Standards and Technology, pp. 800-145, Septembre 2011.

(Voas and Zhang, 2009): Voas, J., Zhang, J., Cloud Computing: New Wine or Just a New Bottle?. Published by the IEEE Society. Computer.org/ITPro. 2009.

(Sakr et al, 2011) : Sakr, S., Liu, A., Batista, D. M., Alomari, M., A Survey of Large Scale Data Management Approaches in Cloud Environments. Published in Communication Surveys & Tutorials, IEEE, vol. 13, no. 3, pp. 311-336, 2011.

(Thibault, 2011): Thibault. D., Study and Comparison of Elastic Databases: Myth or Reality? , Master degree in science computing and engineering ,2011.

Biographies

Taous Iggui, received her Master's degree in Networks and Distributed Systems from the University of A. Mira, Bejaia, Algeria, in 2013, in the same year, she joined LIMED Laboratory(Laboratory of Medical Computing)at the same institution. As a PhD student in Computer Science, she has international publications and conferences related to User Profiling, Data Mining, Text Processing, Knowledge Management, Semantic Web, Cloud computing and Big Data.

Hassina Nacer received an engineering degree in Computer Science from ESI Algiers, a Magister degree in Computer Science from Algiers University, and a PhD thesis in distributed networks from Bejaia University. She worked as an engineer at the Algerian customs for 6 years. She has served as a teacher-researcher in the Department of Computer Science at Bejaia University since 2001. In 2015, she joined the Department of Computer Science at University of Science and Technology "USTHB", Algiers, as an Assistant Professor. She has written several papers on Semantic Web services Technology and Distributed Systems. Her research interests include Knowledge Base Systems, Distributed and Cooperative Computing, Semantic Web, Web Services and Web Services Security.

Abelkamel Tari Head of LIMED (laboratory of Medical Computing) since February 2013 and team leader of data mining. He held different positions at the university of Béjaia (Head of Computing department, Operational Rechearch department, head of doctoral school) and at the National School of Computing ESI (ex.INI), Algiers between 1986-1994. After his Diploma in Mathematics at USTHB (Algiers), he held his Master by Research in Lancaster University (England). He defended his PhD Thesis in Computer Sciences in 2008 at the university of Bejaia and his Accreditation to supervise (HDR) at ESI (ex-INI) in 2010.