

## Research Article

# A Distributed Collaborative Platform for Personal Health Profiles in Patient-Driven Health Social Network

Ahmed M. Elmisery,<sup>1</sup> Seungmin Rho,<sup>2</sup> and Dmitri Botvich<sup>3</sup>

<sup>1</sup>*School of Information Technology, Monash University Malaysia, Malaysia*

<sup>2</sup>*Department of Multimedia, Sungkyul University, Anyang, Republic of Korea*

<sup>3</sup>*TSSG, Waterford Institute of Technology, Waterford, Ireland*

Correspondence should be addressed to Seungmin Rho; [smrho@sungkyul.ac.kr](mailto:smrho@sungkyul.ac.kr)

Received 24 October 2014; Accepted 17 November 2014

Academic Editor: Naveen Chilamkurti

Copyright © 2015 Ahmed M. Elmisery et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Health social networks (HSNs) have become an integral part of healthcare to augment the ability of people to communicate, collaborate, and share information in the healthcare domain despite obstacles of geography and time. Doctors disseminate relevant medical updates in these platforms and patients take into account opinions of strangers when making medical decisions. This paper introduces our efforts to develop a core platform called Distributed Platform for Health Profiles (DPHP) that enables individuals or groups to control their personal health profiles. DPHP stores user's personal health profiles in a non-proprietary manner which will enable healthcare providers and pharmaceutical companies to reuse these profiles in parallel in order to maximize the effort where users benefit from each usage for their personal health profiles. DPHP also facilitates the selection of appropriate data aggregators and assessing their offered datasets in an autonomous way. Experimental results were described to demonstrate the proposed search model in DPHP. Multiple advantages might arise when healthcare providers utilize DPHP to collect data for various data analysis techniques in order to improve the clinical diagnosis and the efficiency measurement for some medications in treating certain diseases.

## 1. Introduction

The raise of social networks as an effective tool for the interaction between people and as a platform for sharing their health conditions leads to the appearance of more purpose driven social networks in healthcare. Utilizing social networks as an integral part of healthcare has made a significant impact in digital healthcare and the emerging of what is referred to as health social networks (HSNs). Health social networks hold a considerable potential value for healthcare organizations [1] because they fetch people together for collaboration and collect information related to their experiences and reflections. One-third of Americans who go online try to find fellow patients similar to their health status to discuss their conditions [2] and 36% of the users utilize other users' information and opinions on social networks before making medical decisions [3]. Health social networks (HSNs) [1] were initially directed at patients but different caretakers and researchers may be able to participate in it. HSNs hold

a considerable potential value for healthcare organizations because they can be used to reach collaborators, accumulate information, and facilitate an effective partnership. However, trends in the next generation of healthcare systems demand applications that can allow prevention of diseases even before they are apparent by using advanced analytics and learning techniques [4, 5].

Health social networks can also be employed to provide real dataset regarding clinical trials. The existence of health social networks makes traditional clinical trials more efficient through the availability of large searchable online databases of patients' information which contains their health history and conditions. Pharmaceutical firms, healthcare analysts, health policy planners, and other interested parties can assess the demand and market size directly from health social network websites. To date, there are numerous paradigms for health social networks that exist on the Internet including PatientsLikeMe, DailyStrength, CureTogether, peoplejam, and OrganizedWisdom. The largest and well-known health

social network is PatientsLikeMe which was launched in 2004, and it hits a new milestone of 100,000 members as of June 2011. PatientsLikeMe and Inspire are an example of two health social networks offering access to clinical trials, selling anonymized data to pharmaceutical companies, universities, and medical research labs. As an example of low cost patient recruitment using HSNs, in May 2008, Novartis recruited clinical trial participants from PatientsLikeMe estimating that they could reduce the time required for their study of a new medicine for only a few months [6]. In another case, PatientsLikeMe was utilized to gather ALS patients for a research project and this project has managed to collect 50 DNA samples [7]. This effect might not seem high but the time and cost savings in recognizing, inspecting, contacting, and obtaining responses from relevant patients are critical.

HSNs can lead to discovering new findings that can help to understand natural history and development of various diseases by utilizing quantitative analysis tools on massive data that is gathered through various patients' communities who are continuously interacting and reporting their health conditions and medical history. For example, PatientsLikeMe has an in-house research staff which is publishing some of their healthcare research, such as their research that is related to determining the nonmotor symptoms of Parkinson's disease in younger patients [8]. HSNs are equipped with health tracking process that can be employed by patients to provide their experience and feedback to the clinical trials process including their response to the drugs. For example, patients registered in PatientsLikeMe network have noticed and suggested a set of corrections and improvements to the graphical display of the data in ALS clinical trials [9].

The next generation of HSNs is based on patient-inspired research, which is also called crowd-sourced health research. These novel HSNs emerged as experienced patients may no longer have the willingness to wait for formal research findings and medical clinical trials and can possibly fill the gap for rare diseases that do not make outstanding business cases in the existing healthcare model. The experienced patients can study and review research literature on their own and investigate new findings, tracking the results, sharing the information, and running nontraditional clinical trials with themselves. As an example, a patient registered in PatientsLikeMe, diagnosed with rapidly progressive and young-onset ALS, managed to collect information regarding other 250 patients regarding a self-experiment with lithium [10] for a research study. This patient-inspired research had found [11] preliminary results regarding the use of lithium as a therapy which does not slow the disease progression. This example highlights the power of patient-inspired research and role of patients in medical research. The ownership of that healthcare process and the concomitant controversial legal, ethical, and methodological are other issues. However, fraud and privacy breaches are likely to arise in HSNs as there are significant economic incentives for drugs and other treatments to have high patient usage statistics and favorable reputation. This requires a platform that is able to select data in a more rational and similar way to human ones only in a shorter period of time autonomously and automatically while preserving the privacy of participants.

This paper introduces a proposed platform that we called Distributed Platform for Health Profiles (DPHP) that can extract helpful datasets for clinical trials and detect fraudulent aggregators. DPHP utilizes a search model that considers multiple attributes of various data aggregators and their offered data, such as success criterion and trust rank for each aggregator beside price, type, accuracy level, anonymization level, tuples types, number of records, gathering method, and demographics for each dataset offered by such aggregator. Furthermore, DPHP facilitates a tendering process where aggregators tender their personal health data in an intelligent manner. Privacy concerns for the participants have obliged DPHP to utilize the privacy enhancing framework proposed in [12–17] in order to give the patients confidence that the usage and disclosure of their healthcare profiles and related demographic information are under their control. This work is structured as follows. In Section 2, related works are described. Section 3 briefly introduces the proposed DPHP (Distributed Platform for Health Profiles). Section 4 describes our proposed fuzzy search model and Section 5 presents a case study to illustrate this fuzzy search model on proposed platform. Section 6 concludes this paper.

## 2. Related Works

The current literature addresses the problem of exploiting social data from the perspective of knowledge sharing. In some systems, very general techniques like the ones that were exploited in the information filtering research are used to search the heterogeneous information sources with little information available about the users' needs. The users should be assisted while exploring data in social data and the system should keep track of their actions to identify their real needs in order to extract suitable data that is matching their needs. In [18, 19] a peer-to-peer approach is proposed based on the users' communities concept, where the community will have an aggregate user profile representing the group as a whole but not the individual users. Communication occurs between the individual users but not with the servers. Thus, the processing is done at the client side. Storing users' profiles on their own side and running the required processing in a distributed manner without relying on any server is another approach proposed in [20]. While those techniques are suited in dealing with large scale applications, other works have shown the need for more purpose-specific techniques to be applied in order to personalize the search process on the social data. The work in [21] describes a recommender system for VOD applications, where the structure of a movie database is exploited to customize the recommended items for the users. The system analyzes customers' selections in order to identify the items' attributes which are affecting their decisions. This information aids in filtering out the new items in order to select the items to be recommended. The work in [22] presented a system to generate labels for museum items by summarizing the information stored in the records of an external database. This information consists of unstructured natural language text, where the system exploits NLP techniques to interpret the text and then generates summaries based on the detailed domain ontology. This deep analysis

of the contents is the basis for the generation of personalized labels. Huang work [23] explores the issues related to applying extenic methods to build product's resource character, and then the system asks the users to provide the input authority with this system's resource character value for each store. Through the process of assessment, the matching procedure poses the "buyer's point of view" and then it calculates the matching preference value of each product provided by each store and provides solutions for the selected product, to facilitate a complete deal so both the consumer and producer can get their requirements.

### 3. The Proposed DPHP

The intuition behind our solution stems from enabling the individuals or groups to control the release of their personal health profiles on a core platform that will store their datasets in a nonproprietary manner to enable the usage of this data in parallel domains, so as to maximize the monetization effort where individual participants benefit from every utilization of their personal health data. However, DPHP is not fully P2P; instead it is a hybrid P2P system like Gnutella [24]. There exists a set of nodes connected to each other as seen in Figure 1. A typical application for the DPHP involves a genomic research based on biobanks. Biobanks are a type of biorepository that store biological materials like organs, tissue, blood samples, cells, and other body fluids that are containing traces of DNA or RNA. This biological information represents the key resources for a research like genomics and personalized medicine. The research groups and pharmaceutical companies can employ the data stored in the biobank for clinical trials, personalization of treatments, or research purposes. Biobanks can employ HSNs to collect genetic or health data from patients and then share it with different external parties like healthcare providers, research and government institutions, and industry. Moreover, DPHP can be utilized as data sharing platform to verify the research output of any health related analytical studies with other datasets representing another random sample of sufferers. Different research groups which carry out similar research studies can benefit from this feature. However, patients may not be willing to participate in this platform because they are concerned about the privacy of their health profiles, as the data they are going to release can be used against them if it is linked to their real identity. For example, on the basis of their health profiles, health insurance companies can prevent them from participating in specific insurance programs or certain enterprises can refuse to hire them. The emerging privacy considerations have been handled in DPHP by utilizing the collaborative privacy framework which has been proposed in [12–17] to preserve the privacy of the users' health profiles. This approach will give the participants the confidence that the disclosure risk of their health profiles is eliminated.

The basic element in the DPHP is the Expert Agent Execution Server (EAES), which is an execution environment for the expert agents that have been created by the health expert or researcher. An expert agent is instructed with the required trial along with the query needed to fetch the data to fulfill this trial. Thereafter, the expert agent is forwarded to

EAES based on the request of the health expert or researcher. The agent can reside in the EAES and acts as a mapper agent which will be responsible for forwarding its worker agents in order to relate data aggregators to fetch the data required for the trial. There also exists a set of Aggregator Service Discovery (ASDs) which is responsible for maintaining the information regarding different data aggregators.

*3.1. System Components.* As illustrated before, a high level architecture for the DPHP was depicted in Figure 1. DPHP consists of different nodes that are connected through the Internet (it can be a private network as well). DPHP essentially creates a virtual private network even when an underlying network infrastructure is the public Internet. Each aggregator acts as a gateway for gathering anonymized patients' health profiles from different health social networks. As the patient's consent is essential in this process, he/she is notified once the data collection is started. HSN can give certain benefits (like money, prizes, gift brochures, etc.) for the users who have a sustainable rate in participation within each data collection request. A detailed explanation of different nodes is as follows.

*ASD (Aggregator Service Discovery).* An ASD is an entity in DPHP that is responsible for maintaining information about the aggregators. The information about the aggregators should include the domain names, IP addresses, and data catalogues. The information about related aggregators can be provided when a health expert tells ASD the kind of data required for the trial in hand. When only a few aggregators are active, one ASD can be utilized for serving such a small group. However, when more aggregators are deployed, a set of ASDs should be distributed in different zones in order to attain a load balancing for the serving of different data collection requests.

*EAES (Expert Agent Execution Server).* EAES is a server in DPHP that is provided for the registered health experts in order to host their expert agents that are equipped with the required trials and queries to search for the data needed for each of these trials. Based on the health expert's searching criteria, the expert agent will forward in parallel a pool of worker agents to the relevant aggregators, which in turn will return the required data for the trial. Sandboxing and logging techniques can be utilized to protect both of the execution server and expert agents from malicious attacks.

*SAC (Security Authority Center).* SAC is a trusted third party in DPHP that is responsible for generating certificates for all aggregators and managing them. Additionally, SAC is responsible for making security assessment on those authorized aggregators according to the attack and feedback reports which are collected from the participants and the health experts. Thereafter, SAC submits periodic reports to ASD in order to reflect the updates in the trust ranks of registered aggregators.

*SMA (Success Management Authority).* SMA is the authority within DPHP that is responsible for assessing the success

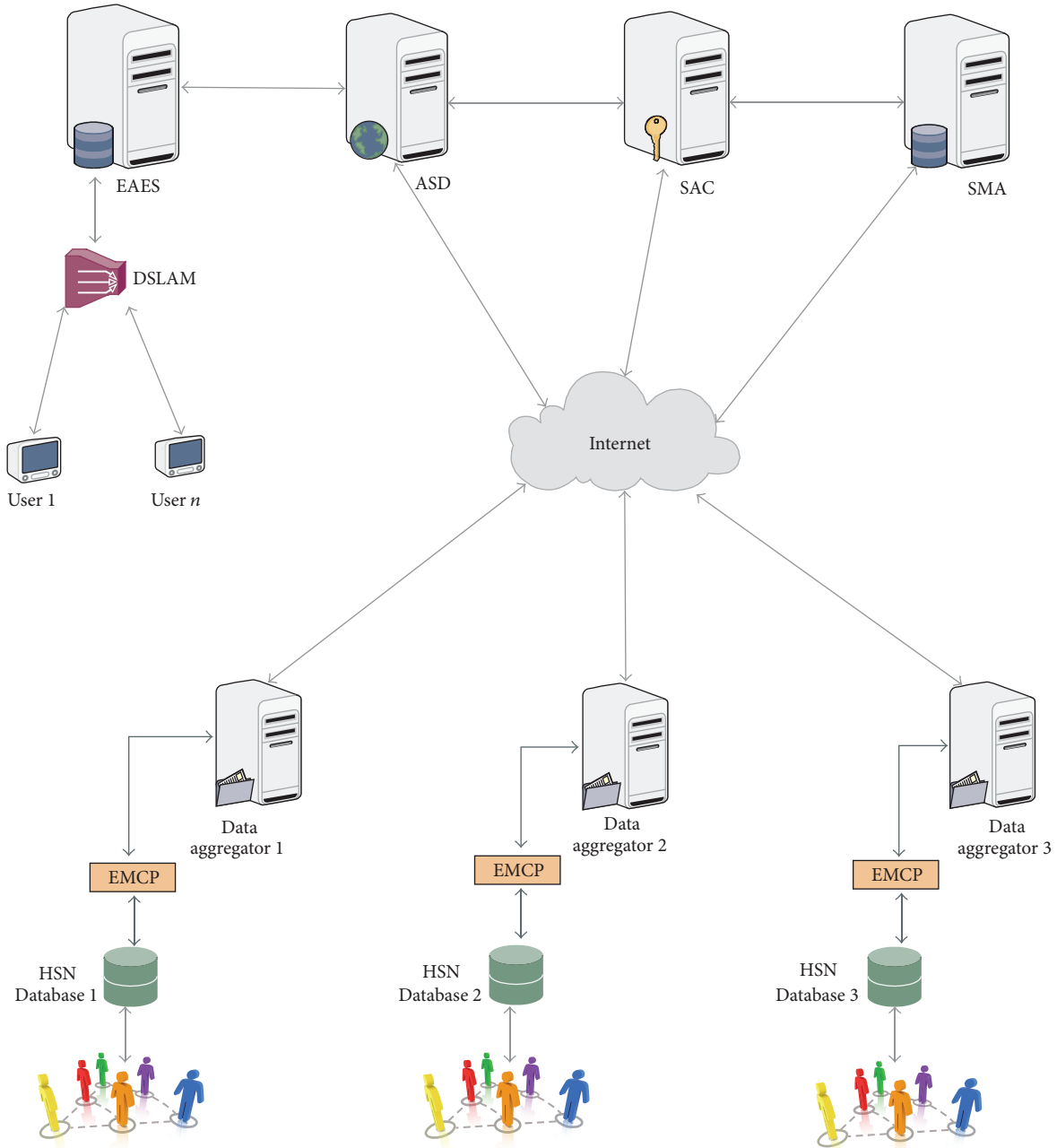


FIGURE 1: An overview of DPHP.

criterion for all aggregators. When an aggregator cheating occurs, a health expert can report this to the SMA. After investigation, the success criterion of this aggregator will be downgraded and this in turn diminishes its revenues and the credibility of the data collected from this aggregator. On the other hand, the successful processes will help to amend the success criterion for each aggregator.

*Health Expert.* The beneficiary of the DPHP could be a registered expert patient or a researcher running a trial for his/her own. Moreover, the health expert could be a medical

research institute or pharmaceutical company enrolled with any EAES before utilizing the facility of submitting task agents and collecting data using the DPHP. The health expert can utilize DPHP to search for specific data that is needed for his/her research or trial through an expert agent hosted on EAES. Additionally, the payment for the extracted data is also done through the EAES using a secure e-payment system. Finally, the health expert is also responsible for sending appeals to the SMA for any aggregator cheating that may occur during the trial and/or data collection which is difficult to be detected before the payment. If the cheating is true,

the aggregator's success criterion will be degraded, which will result in decreasing the number of worker agents that are being forward there.

3.2. *The Search Workflow in DPHP.* Based on the proposed framework, the process of enabling the selection and collecting numerous datasets from various aggregators can be described as follows.

(1) *Health Expert Requirement Elicitation.* The health expert selects an ASD where he/she has registered as a user in order to create an expert agent. Thereafter, he/she inputs the query for selecting the dataset that is required for the trial in hand. Moreover, he/she specifies the properties related to the extracted datasets such as price, type, accuracy level, anonymization level, tuples types, number of records, gathering methods, and demographics. Finally, he/she also determines the attributes for the potential aggregators, such as the trust rank and success criterion.

(2) *Aggregators Selection.* After the health expert dispatches the expert agent to the EAES, the EAES will host this expert agent in order to allow for the completion of its required task. The expert agent divides the required processing along with data query between different primary agents (PA) such that each one of them will be containing one subtask and one subquery. These primary agents will be tasked to reside within the qualified aggregators and then forward in parallel a pool of worker agents (WA) to fetch the required data. An aggregator is selected only if its trust rank and success criterion meet the same requirements specified by the health expert. The values for these attributes can be obtained from ASD, SMA, and SAC.

(3) *Datasets Assessment.* When the results are returned by all the worker agents, a second stage of assessment is taken on both properties of datasets and aggregators' trust rank and success criterion. The sorted results are presented back to the health expert by the expert agent.

(4) *Negotiation with the Successful Aggregators.* Based on the decision of the health expert, a few aggregators will be short-listed and selected for negotiation, and then the expert agent will start forwarding negotiation agents to these selected aggregators. A lot of negotiation models have been proposed and can be utilized for such process [25]. However, in this paper, we will not address this issue.

(5) *Payment for Aggregators.* With the successful results of negotiations, one or more aggregators will be favored to collect the dataset, and then an online secure payment occurs between the expert agent and each one of the selected aggregators. Different e-payment models can be utilized for this purpose such as the model proposed in [26].

(6) *Feedback from the Health Experts.* After receiving the required dataset from the selected aggregators, the health expert can evaluate the whole process or report the aggregator cheating. The success criterion of such aggregator will be

modified based on the feedback from health experts. In addition, during the whole process, in the case of the detection of any attacks from malicious hosts on the primary or worker agents [27], the expert agent at the EAES will report this to SAC, and this will lead to the deterioration of the trust rank for this aggregator. Thus, the number of agents which are being forwarded to such aggregator will be decreased, since the aggregators' selection step takes place before forwarding any of the primary agents there.

#### 4. The Fuzzy Search Model in DPHP

In our framework, we have developed a fuzzy search model that is much more powerful in search than using the conventional matching models when used for research and investigation of unfamiliar, complex, imprecise, and ambiguous cases. The proposed model can also be applied to locate multiple datasets and various aggregators based on incomplete or partially inaccurate properties; the returned results by the fuzzy search model are likely based on the subjective relevance. DPHP has easily employed software agents in order to attain parallel and distributed processing. When an expert agent is created and starts running at EAES, it retrieves from ASD a list of aggregators that offer specific datasets needed for the trail that has been specified by its health expert. Thereafter, the expert agent starts to dispatch a set of primary agents to the selected aggregators, where each primary agent forwards multiple worker agents for querying the metadata of datasets that are offered by the numerous nodes that exist within each registered HSN with a certain aggregator. This metadata involves attributes of each dataset, such as price, type, and accuracy level. Each worker agent is responsible for visiting one node within each HSN. Once all the worker agents fulfill their tasks, the primary agents send the results back to the expert agent. Suppose there are hundreds or thousands of nodes which are offering the same kind of datasets. It is unnecessary and even impossible for a health researcher or even a mobile agent to browse all of them. So it is quite necessary and reasonable for the health researcher to find a way to evaluate these nodes and get the best nodes for further investigation. This assessment process not only is compatible with the human behavior, but also can reduce the network load. Moreover, the number of datasets may be several times more than the number of aggregators, since each aggregator may provide multiple health profiles to the health expert. The health expert should evaluate these datasets and get a short list for the best of them and then negotiate with the aggregator for further benefits. The search model in DPHP explores the issues of allocating the best and most convenient aggregators to the health expert as well as assessing and refining their datasets and then returning the best datasets to the health experts. The allocation and assessment are based on a set of predefined selection criteria that are domain-specific. Additionally, as most of the real-world situations that can involve constraints may be imprecisely defined, such as recent datasets and high accuracy, additionally the common knowledge may be limited to the expert agent. The expert agent should be

autonomous enough in order to have the ability to consider such incomplete and imprecise information. In DPHP, we applied the fuzzy rules technologies that have the ability to naturally process incomplete and imprecise information to extract rational results.

Our proposed fuzzy search model has several features and advantages as it consists of two sequent and correlated stages: the first is the aggregators' selection stage and then the datasets assessment stage. The second stage is processed based on the results obtained from the first one. This model can reduce the network load that makes it suitable for an environment where the computing resources are limited. The expert agent can search more nodes and datasets based on the real-time situation and generate more reasonable results.

**4.1. Preliminaries: Fuzzy Set and Linguistic Variables.** In mathematics, a fuzzy set is different from a crisp set as each element within the fuzzy set has a degree of membership. The membership function is responsible for defining the relationship between a value in the set's domain and its degree of membership [28]. Linguistic variables [29] are variables whose values are not numbers but words or sentences in a natural or artificial language. They are used as a counterpart to the concept of numerical variables. As we mentioned earlier, we have applied fuzzy rules technologies as one of the main building blocks in our fuzzy search model. The fuzzy rule based model [30] consists of a rule base of the following form:

$$\text{if } V_1 \text{ is } A_{i1}, V_2 \text{ is } A_{i2}, \dots, V_n \text{ is } A_{in}, \text{ then } U \text{ is } B_i. \quad (1)$$

The  $V_i$ 's are the antecedent variables and  $U$  is the consequent variable. The  $A_{ij}$ 's and  $B_i$ 's are the fuzzy subsets over the corresponding variable's domain; generally, these subsets represent the linguistic variables. The fuzzy rule based model determines the consequent variable's  $U$ 's value for a given manifestation of the antecedent variables  $A_{ij}$ . This model utilizes principles from utility and fuzzy theories which make such a model straightforward and simple.

Assume a variable  $x$  is consisting of a number of attributes:

$$x = \{x_1, x_2, \dots, x_n\}. \quad (2)$$

- (1) For each attribute  $x_n$ , calculate its membership level as

$$A_i = F_i(x_i), \quad (3)$$

where  $F_i$  is a semantic function for the attribute  $x_i$ .

- (2) Calculate the units/levels of each attribute as

$$U_i = V_i(A_i), \quad (4)$$

where  $V_i$  is a transfer function that maps the attribute into prespecified values in a numerical interval that is [1, 10].

- (3) Calculate the overall utility of the variable  $x$  as

$$U(x) = \sum w_i U_i, \quad (5)$$

where the relative importance assigned for each attribute is represented as a normalized weight  $w_i$  such as  $\sum w_i = 1$ .

- (4) Calculate the overall membership value of the variable  $x$  as

$$U = F(U(x)), \quad (6)$$

where  $F$  is a transfer function for  $x$ . So the overall membership value for a variable  $x = \{x_1, x_2, \dots, x_n\}$  in a multidimensional space is defined as

$$U = F\left(\sum w_i V_i(F_i(x_i))\right). \quad (7)$$

**4.2. Transforming Linguistic Variables Using Semantic Function.** The semantic function is responsible for assigning each linguistic attribute into its meaning as a membership value. These values are usually represented as linguistic values, such as very clear, clear, semisanitized, sanitized, or encrypted. These functions have several features as follows.

- (i) These functions are attribute-dependent; that is, for different linguistic attributes, there may exist different levels for each category. In addition, for the attributes that can be represented as digital values, that is, price and number of attributes, the semantic functions can use these digital values directly; for the attributes that cannot be represented as digital values directly, that is, accuracy level and anonymization level, a table should be built that maps these linguistic values into digital values.
- (ii) These functions can either classify attribute values into predefined number of categories or classify them based on real-time properties of the dataset's metadata. In the first case, the health expert should specify the number of categories that he/she prefers. In the other one, the expert agent summarizes all the information that has been collected from the DPHP and then it starts to extract the standard categories based on this information. These standards are dynamic and suitable for this process only.

As the computing resource for the expert agent is limited, we have used a modified version of LLA algorithm that was proposed in [31] for the first case described above as shown in Algorithm 1. We adopted another algorithm for the latter one as shown in Algorithm 2.

**4.3. Mapping Attributes Using Transfer Function.** A Transfer function is responsible for mapping the attribute's membership levels into prespecified values in a numerical interval that is [1, 10]. DPHP makes use of a linear transfer function of the following type:

$$U(A_i^*) = \frac{\text{Max } A^* - A_i^*}{\text{Max } A^* - \text{Min } A^*} 0.9 + 1 \quad (8)$$

or

$$U(A_i^*) = \frac{A_i^* - \text{Min } A^*}{\text{Max } A^* - \text{Min } A^*} 0.9 + 1, \quad (9)$$

**Inputs**  
 Initial values:  $X_i$  ( $i = 1, \dots, n$ )  
 Number of categories:  $k$

**Outputs**  
 Clustering Results:  $Y_j$  ( $j = 1, \dots, k$ )

- (1) Select any values  $X_{i1}, X_{i2}, X_{i3}, \dots, X_{ik}$  from  $X_i$  randomly
- (2) Set an initial starting category  $Y_j = X_{ij}$  ( $j = 1, \dots, k$ )
- (3) Do until the group member is stable
  - For each  $X_i$  ( $i = 1, \dots, n$ )
    - If  $X_i \in [Y_j, Y_{j+1}]$ 
      - $D_1 = \text{Distance}(X_i, Y_j)$
      - $D_2 = \text{Distance}(X_i, Y_{j+1})$
      - If  $D_1 < D_2$  then  $X_i$  is in the cluster (category) of  $Y_j$
      - Else  $X_i$  is in the cluster (category) of  $Y_{j+1}$
      - End if
    - End If
  - End for
  - $Y_j = \text{the average of cluster } Y_j$  ( $j = 1, \dots, k$ )
  - End Do

ALGORITHM 1: Modified LLA clustering algorithm.

**Inputs**  
 Initial values:  $X_i$  ( $i = 1, \dots, n$ )  
 Fuzzy factor  $\zeta$

**Outputs**  
 Categories results:  $Y_j$  ( $j = 1, \dots, n$ )

- (1) Sort  $X_i$  by ascent or descent to  $B_i$
- (2) Set the current Categorylevel = 1
- (3) Set item number  $A$  in current category level = 1
  - For each  $B_x$  ( $x = 2, \dots, n$ )
    - $$B^* = \frac{1}{A+1} \sum_{m=x-A}^x B_m$$
    - If  $|B^* - B_x|/B^* > \zeta$  then  $B_x$  is not in this level
    - Categorylevel = Categorylevel + 1
    - $A = 0$
    - Else  $B_x$  is in this level
    - $Y_x = \text{Categorylevel}$
    - $A = A + 1$
    - End if
  - End for

ALGORITHM 2: Adopted simple categorization algorithm.

where  $A_i^*$  represents the average value of current category level. Meanwhile, DPHP uses (8) if the function is decreasing with respect to  $A_i^*$  and (9) if it is increasing.

**4.4. Fuzzy Search Model in DPHP.** In this paper, the proposed fuzzy search model is executed in three stages: input, aggregator selection, and dataset assessment.

**4.4.1. Input.** In this stage, the expert agent collects from the health expert the queries that are needed to retrieve the data which are required for the trial in hand along with the properties related to the collected datasets and the attributes for

the potential aggregators. The health expert's requirements can be further organized into "debatable" requirements and "inalienable" requirements; the "inalienable" requirements are used as the basic conditions in search stage while the "debatable" requirements can be used in the negotiation stage. Moreover, the health expert should select suitable standard categories that will be predefined in the expert agent or learn the health expert's requirements by specifying the relative weights of each attribute and/or property. Finally, the health expert should specify the selection criteria such as the number of aggregators/datasets to be selected or the selection percentage. The expert agent can select the aggregators and evaluate the candidate datasets, and then the negotiation with

the appropriate aggregators about their datasets is based on the health expert's requirements.

**4.4.2. Aggregator Selection.** This stage explores the issues of selecting the appropriate and most potential aggregators to the health expert's requirement in the DPHP. Before the start of forwarding any worker agents there, this selection stage is done only over several attributes such as the success criterion, trust rank, and the type of datasets. The success criterion of each aggregator is a value that is determined based on the number of its previous successful processes and the nodes with a low price and accurate health profiles that are affiliated with it. The aggregator which is attracting large number of appropriate nodes from the HSN will get quickly a high success criterion. Those attributes for the aggregator selection stage are stored in the ASD with the domain names, IP addresses, and data catalogues for all nodes. After selection, worker agents will be forwarded to those appropriate aggregators for searching in parallel their datasets. In this stage, the processes of selecting aggregators are done in three more steps: aggregator selection, aggregator assessment, and aggregator refining.

(i) **Aggregator selection:** in this step, the expert agent queries the ASD' database using the requirements specified by the health expert in order to get the domain names and IP addresses of the correlated aggregators.

(ii) **Aggregator assessment:** in this step, the ranking of the aggregators is computed based on our fuzzy rule based model, where the overall membership function is defined as follows:

$$U = F\left(\sum w_i V_i(F_i(x_i))\right) \quad \text{where } x = \{S, T, D\}. \quad (10)$$

The variable  $x$  could be one of the following.

- (1)  $S$  denotes to the success criterion of the aggregator. The aggregator with larger number of previous successful processes and better feedback reports receives a higher value of  $S$ . For every successful process, the aggregator will receive a number of success points. Also the health expert can rate the datasets which were gained from the search process. The aggregator can get additional credit points with the positive rating or miss some credit points if the rating is negative. The information regarding the success criterion of different aggregators is maintained by the SMA.
- (2)  $T$  denotes to the trust rank of the aggregator. In DPHP, SAC is the entity which is responsible for making trust assessment on those authorized aggregators according to the attack reports obtained from various parties in DPHP. Thereafter, SAC periodically reports the updates in the trust ranks of aggregator to ASD. Higher trust rank means higher security level for the aggregator.
- (3)  $D$  denotes to the time required by the aggregator to assemble and deliver the prospective datasets. The type of datasets is quite important to the health expert. It can be long if the health expert is demanding more sophisticated datasets that will require various

TABLE 1: Time required for different datasets types.

Datasets type (per 400 records)	Numerical measurements	Pictures	Recorded signals	Other
Time (hours)	50	100	80	NA

preprocessing steps in order to be collected and prepared for the delivery. However, the size of the datasets itself is the main impact factor within the type of datasets variable. Therefore, the aggregators should have prespecified datasets types for each of the required processing scopes and only offer datasets for the health experts in these domains. In DPHP, each aggregator has a table to illustrate the time for delivering the datasets from the nodes in HSN to the health expert, such as Table 1.

Table 1 illustrates the datasets type of aggregator ABC. The required time to collect and prepare 400 records of numerical measurements is 50 hours, for pictures is 100 hours, and for recorded signals is 80 hours. Moreover, this table means that the aggregator ABC only offers datasets for the health experts from these dataset's types. If the health expert demands a dataset that the aggregator does not support such as textual data, then the value of  $D$  will be set to 0.

(iii) **Aggregator refining:** in this step, a list of aggregators addresses is returned to the health expert based on the assessment results and selection criteria that he/she has specified. The selected aggregator in this list must fulfill at least three conditions as follows:

- (a) the aggregator that is active,
- (b) the aggregator in high level,
- (c) the aggregator that has the  $D \neq 0$ .

Condition (a) ensures that the aggregator is online. Condition (b) ensures that the aggregator is "better" than the other aggregators that were not selected. Condition (c) ensures that datasets requirements for the health expert can be met at this aggregator. At the end of the aggregator selection stage, a number of aggregators are returned to the health expert, where he/she can select some/all of these aggregators in the list for a further search process.

**4.4.3. Dataset Assessment.** In this stage, datasets assessment occurs when all the worker agents send back additional information regarding the datasets, such as the price, accuracy level, anonymization level, tuples types, number of records, gathering method, and demographics. Hence another search process will be conducted again over all the gathered properties and the sorted results of appropriate datasets will be presented to the health expert. Upon the health expert decision, the expert agent can now send a new set of worker agents to a selected set of visited aggregators to negotiate for a lower price or more convenient accuracy level. According to the results, the health expert will choose one or more aggregators for data collection and payment. The datasets



assessment stage is similar to the aggregator selection stage but, instead of searching the aggregators' attributes, the search process is done over the properties of the various datasets which are offered by the selected aggregators from the previous stage. In this paper, the process of datasets assessment is carried out in two steps: datasets assessment and datasets refining.

(i) Datasets assessment: in this step, the ranking of the datasets is computed based on our fuzzy rule based model. In DPHP, the overall membership function is defined as follows:

$$U = F \left( \sum w_i V_i (F_i (x_i)) \right) \quad \text{where } x = \{P, D, W, C\}. \quad (11)$$

The variable  $x$  could be one of the following:

- (1)  $P$  denotes to the price of the datasets;
- (2)  $D$  denotes to the "number of records" within datasets;
- (3)  $W$  denotes to the anonymization level of the datasets;
- (4)  $C$  denotes to the accuracy level of the datasets.

We have predefined several standard categories with different weight for each category. The health expert can either use these predefined standard categories or customize the weight of each category based on the real-time properties of the dataset's metadata. The four standard categories that we have defined are as follows.

- (a) *The Category of Price Priority.* If the health expert takes the price as the most important factor for search and selection, he/she can select standards in this category. In this category, the price is the main impact factor to be utilized when assessing the datasets, rather than the other properties. The datasets with a lower price can get a higher score. There are three levels in this category: proportional price priority, modest price priority, and maximum price priority. Thus, within each level the relative weight of the price variable is increased gradually.
- (b) *The Category of Size Priority.* If the health expert wants to get big datasets as much as possible, such that these datasets contain a large number of records, then the "number of records" property is the most important factor for him/her. The datasets with a large number of records can get a higher score. There are three levels in this category: proportional size priority, modest size priority, and maximum size priority.
- (c) *The Category of Accuracy Priority.* In this category, the health expert prefers more accurate datasets which have been collected by experienced patients using modern and well-known medical devices. This category is suitable for healthcare providers and pharmaceutical companies, which want to perform various data analyses on the collected datasets in order to improve the clinical diagnosis and measurements for some medications in treating certain diseases, executing specific clinical trials, and/or other research purposes. There are also three levels in this category:

proportional accuracy priority, modest accuracy priority, and maximum accuracy priority.

- (d) *The Category of Balance Priority.* In this category, the health expert has no explicit preference. The weights of different properties are similar.

(ii) Datasets refining: in this step, a sorted list of all datasets is returned to the health expert based on the search process result. The health expert can select some/all of the datasets and negotiate with the aggregators about these datasets in order to attain further benefits.

## 5. A Case Study on DPHP

In this section, we will present a case study to illustrate the fuzzy search model in DPHP clearly. If we suppose a health expert wants to collect a dataset related to her research, at first, she registers at EAES and then she creates an expert agent in order to be assigned the task of collecting the required data for her research. She sets the price and accuracy as "debatable" requirements and other requirements as "inalienable" queries. She prefers a lower price than other properties, so she sets the main factor for the assessment of the datasets to be the category of price priority, where she selects modest price priority as her requirement. The health expert query is shown in Table 4.

Note: the accuracy level is a numerical value within the interval [1, 15], and it reflects the degree of the correctness and confidence for each data element within the dataset. Moreover, the selection ratio top 25% means that she only wants top 25% of the aggregators to be included in the results' list.

*5.1. Aggregator Selection.* In this stage, the worker agents perform a search process for selecting the appropriate aggregators; the selection is done only over the several attributes that are associated with the aggregators, such as a success criterion, trust rank, and type of datasets. The health expert can set the number of aggregators she needs or she can only set the percentage of the aggregators to be selected; that is, she can select the first 100 or top 25% of aggregators and then she starts forwarding the worker agents to these selected aggregators. Assume the expert agent gets a list from ASD with 100 aggregators that offer the datasets that the health expert requires. After the aggregators selection, the health expert selects top 25% of the aggregators with a better success criterion, higher trust rank, and datasets in the required type of datasets. Then the expert agent sends a set of worker agents to these aggregators in order to get detailed information regarding their offered datasets. The search results are shown in Table 2. These results were extracted based on the selection stage and the requirements that the health expert has specified. All the aggregators that have the same membership value in results were selected (aggregator ID 106 with overall membership value III was also selected). This fuzzy search model is compatible with the human behavior because all these aggregators will look the same for those that will be selected manually by the health expert.

TABLE 2: Aggregator selection results.

Aggregator ID	Success criterion	Trust rank	Dataset size	Membership value	Result
22	I	I	532	I	Yes
88	II	II	700	III	Yes
106	I	II	500	III	Yes
135	II	III	720	IV	No
174	V	IV	234	VI	No
201	V	IV	100	VII	No

TABLE 3: Datasets assessment results.

Datasets_ID	Price, accuracy	Simple categorization	Modified LLA	Final value
1	870, 2	I, II	I, I	9.51
2	970, 3	I, III	I, I	9.43
3	1008, 1	I, I	I, I	9.61
4	1420, 1	III, I	II, I	9.01
5	880, 14	I, VI	I, III	8.86
6	1200, 15	II, VI	I, III	7.53
7	1400, 15	IV, VI	III, III	6.42

TABLE 4: Sample query for the health expert.

Dataset: diabetes measurements
Owner: older male patients
Collection method: blood glucose meter
Price: <=\$1500
Accuracy level: 7
Dataset size: 500
Rating standard: modest price priority
Selection ratio: top 25%

5.2. *Datasets Assessment.* If we assume that the majority of the selected aggregators offer multiple datasets to the health expert. For example, if each aggregator offers five datasets, the health expert will get at least 60 datasets to be manually investigated further. It is impossible for the health expert to investigate 60 datasets in a short time and consume unnecessary time in the negotiation process with 12 aggregators. The health expert efforts and time should be consumed efficiently in the clinical trial on her hand. Using DPHP, the health expert should be able to select the best of datasets and then negotiate with the aggregators for further benefits. To illustrate the datasets assessment stage simply, we have used an example of 7 offers. The fuzzy factor  $\zeta$  in the simple categorization algorithm is set to be (2%, 8%) and the number of categories  $k$  in the modified LLA clustering algorithm is set to be (5, 3) for the two properties. In order to compare the results, we have used the transfer function to map the membership levels into prespecified values in a numerical interval of [1, 10]. The results for datasets assessment stage are shown in Table 3.

The results were extracted based on the real-time properties of the datasets' metadata which have been categorized into various levels. From these two tables, we can make sure that the results are more appropriate and compatible with the

health expert decision making process. The datasets in the same category have no difference to the health expert. The health expert can freely select the aggregators within any top levels for further negotiation.

## 6. Conclusions and Future Work

In this paper we present the proposed core platform which entitled Distributed Platform for Health Profiles (DPHP) that enables individuals or groups to control their personal health profiles and maximize the effort where users benefit from each usage for their personal health profiles. A fuzzy search model based on DPHP was presented and discussed in detail. The proposed model is compatible with the health expert decision making process. It aids the health expert in the selection and assessment of the appropriate datasets from a huge pool of distributed datasets that are stored in the personal profiles of health social networks. Multiple attributes and/or properties can be utilized within the proposed fuzzy search model. Clustering algorithms were employed to provide an enhanced feature in the proposed model by extracting the categories of the various properties from the real-time properties of the datasets' metadata, which aids in obtaining dynamic and realistic results for the search process. This model can reduce the network load that makes it suitable for an environment where the computing resources are limited.

Our future research agenda will include extending this model with social recommendation techniques in order to facilitate the preferences' learning for the input stage. Utilizing trust attains the success for selecting the aggregators but a possible new dimension could envision expressing this relation for each user independently without the need for a trusted third party. This would provide a more accurate representation of the trusted aggregator, not influenced as much by the dominant users in the system and business deals. Moreover, in all of the applications, users' trustworthiness is

out of interest. Considering malicious user existence would get interesting discussions to grow up.

A more thorough assessment of our model would be useful, such as case studies on a small or large scale. Furthermore, it would be appealing to investigate other innovative applications, which can be used in everyday life, with emphasize the health profiles.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2061978).

## References

- [1] D. Giustini, "Web 3.0 and medicine," *British Medical Journal*, vol. 335, no. 7633, pp. 1273–1274, 2007.
- [2] N. Elkin, *How America Searches: Health and Wellness*, iCrossing, East Sussex, UK, 2008.
- [3] M. Levy, "Online health: assessing the risks and opportunity of social and one-to-one media," in *JupiterResearch*, vol. 2, 2007.
- [4] Y. Zhu, S. L. Keoh, M. Sloman, and E. C. Lupu, "A lightweight policy system for body sensor networks," *IEEE Transactions on Network and Service Management*, vol. 6, no. 3, pp. 137–148, 2009.
- [5] E. Lupu, N. Dulay, M. Sloman et al., "AMUSE: autonomic management of ubiquitous e-Health systems," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 3, pp. 277–295, 2008.
- [6] C. Arnst, "Health 2.0: Patients as Partners," *Business Week*, 2008.
- [7] J. Frost and M. Massagli, "PatientsLikeMe the case for a data-centered patient community and how ALS patients use the community to inform treatment decisions and manage pulmonary health," *Chronic Respiratory Disease*, vol. 6, no. 4, pp. 225–229, 2009.
- [8] P. Wicks, *Parkinson's Disease: More Non-Motor Symptoms for Younger Sufferers*, PatientsLikeMe, 2008, <http://www.patient-slikeme.com>.
- [9] P. Wicks, M. P. Massagli, C. Wolf, and J. Heywood, "Measuring function in advanced ALS: validation of ALSFRS-EX extension items," *European Journal of Neurology*, vol. 16, no. 3, pp. 353–359, 2009.
- [10] J. Kaye, L. Curren, N. Anderson et al., "From patients to partners: participant-centric initiatives in biomedical research," *Nature Reviews Genetics*, vol. 13, no. 5, pp. 371–376, 2012.
- [11] F. Fornai, P. Longone, L. Cafaro et al., "Lithium delays progression of amyotrophic lateral sclerosis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 6, pp. 2052–2057, 2008.
- [12] A. M. Elmisery and D. Botvich, "Agent based middleware for private data mashup in IPTV recommender services," in *Proceedings of the IEEE 16th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD '11)*, pp. 107–111, IEEE, Kyoto, Japan, June 2011.
- [13] A. M. Elmisery and D. Botvich, "An agent based middleware for privacy aware recommender systems in IPTV networks," in *Intelligent Decision Technologies*, J. Watada, G. Phillips-Wren, L. C. Jain, and R. J. Howlett, Eds., vol. 10, pp. 821–832, Springer, Berlin, Germany, 2011.
- [14] A. Elmisery and D. Botvich, "Enhanced middleware for collaborative privacy in IPTV recommender services," *Journal of Convergence*, vol. 2, no. 2, p. 10, 2011.
- [15] A. M. Elmisery and D. Botvich, "Privacy aware recommender service using multi-agent middleware—an IPTV network scenario," *Informatica (Ljubljana)*, vol. 36, no. 1, pp. 21–36, 2012.
- [16] A. M. Elmisery and D. Botvich, "Multi-agent based middleware for protecting privacy in IPTV content recommender services," *Multimedia Tools and Applications*, vol. 64, no. 2, pp. 249–275, 2013.
- [17] A. Elmisery and D. Botvich, "Privacy aware obfuscation middleware for mobile jukebox recommender services," in *Proceedings of the 11th IFIP Conference on e-Business, e-Service, e-Society*, IFIP, Kaunas, Lithuania, 2011.
- [18] J. Canny, "Collaborative filtering with privacy via factor analysis," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 238–245, ACM, Tampere, Finland, 2002.
- [19] J. Canny, "Collaborative filtering with privacy," in *Proceedings of the Symposium on Security and Privacy*, pp. 45–57, IEEE Computer Society, May 2002.
- [20] B. N. Miller, J. A. Konstan, and J. Riedl, "PocketLens: toward a personal recommender system," *ACM Transactions on Information Systems*, vol. 22, no. 3, pp. 437–476, 2004.
- [21] B. Raskutti, A. Beitz, and B. Ward, "A feature-based approach to recommending selections based on past preferences," *User Modeling and User-Adapted Interaction*, vol. 7, no. 3, pp. 179–218, 1997.
- [22] D. Robert, "Dynamic document delivery: generating natural language texts on demand," in *Proceedings of the 9th International Conference and Workshop on Database and Expert Systems Applications*, pp. 131–136, 1998.
- [23] P.-H. Huang, "The extenics theory for a matching evaluation system," *Computers and Mathematics with Applications*, vol. 52, no. 6-7, pp. 997–1010, 2006.
- [24] M. Ripeanu, "Clinical features of metastatic bone disease and risk of skeletal morbidity," *Clinical Cancer Research*, vol. 12, part 2, pp. 6243s–6249s, 2001.
- [25] W. Shen, Y. Li, H. H. Genniwa, and C. Wang, "Adaptive negotiation for agent-based grid computing," in *Proceedings of the Agentcities/AAMAS'02*, 2002.
- [26] M. M. Kouta, M. M. Abou Rizka, and A. M. Elmisery, "Secure e-payment using multi-agent architecture," in *Proceedings of the 30th Annual International Computer Software and Applications Conference (COMPSAC '06)*, vol. 2, pp. 315–320, IEEE, Chicago, Ill, USA, September 2006.
- [27] F. Hohl, *A Protocol to Detect Malicious Hosts Attacks by Using Reference States*, Universitätsbibliothek der Universität Stuttgart, Stuttgart, Germany, 2000.
- [28] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, pp. 338–353, 1965.
- [29] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning. I," *Information Sciences*, vol. 8, pp. 199–249, 1975.

- [30] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamical plants," *Proceedings of the Institution of Electrical Engineers*, vol. 121, no. 12, pp. 1585–1588, 1974.
- [31] A. M. Elmisery and H. Fu, "Privacy preserving distributed learning clustering of healthcare data using cryptography protocols," in *Proceedings of the 34th IEEE Annual International Computer Software and Applications Conference Workshops (COMPSACW '10)*, pp. 140–145, Seoul, Republic of Korea, July 2010.