

Rafiki: A Semantic and Collaborative Approach to Community Health-Care in Underserved Areas

Primal Pappachan*, Roberto Yus[†], Anupam Joshi*, Tim Finin*

*University of Maryland, Baltimore County, Baltimore, MD, USA

{primall, joshi, finin}@umbc.edu

[†]University of Zaragoza, Spain

ryus@unizar.es

Abstract—Community Health Workers (CHWs) act as liaisons between health-care providers and patients in underserved or un-served areas. However, the lack of information sharing and training support impedes the effectiveness of CHWs and their ability to correctly diagnose patients. In this paper, we propose and describe a system for mobile and wearable computing devices called Rafiki which assists CHWs in decision making and facilitates collaboration among them. Rafiki can infer possible diseases and treatments by representing the diseases, their symptoms, and patient context in OWL ontologies and by reasoning over this model. The use of semantic representation of data makes it easier to share knowledge related to disease, symptom, diagnosis guidelines, and patient demography, between various personnel involved in health-care (e.g., CHWs, patients, health-care providers). We describe the Rafiki system with the help of a motivating community health-care scenario and present an Android prototype for smart phones and Google Glass.

Keywords—*Collaboration in health-care, mobile health, medical diagnosis, Semantic Web, reasoning, community health-care*

I. INTRODUCTION

The lack of access to adequate care in medically underserved areas is one of the biggest challenges in health-care in both developing and developed countries [1]. Community Health Workers (CHW), who act as liaisons between patients and health-care providers in these areas, have been able to address this problem to certain extent. Examples of CHW organizations are Accredited Social Health Activist (ASHA) in India and Bangladesh Rural Advancement Committee (BRAC) in Bangladesh. These CHWs are community workers who work in local settings with basic training in health-care and are not as experienced or educated as medical professionals. By using survey techniques and health-care manuals, CHWs collect information about a patient and try to identify the possible illness.

In an evaluation of trained ASHA workers done in relation to their practices regarding child health, it was concluded that a knowledge-practice gap exists in various aspects of health-care and diagnosis [2]. Among these CHWs, 45% had received less than a secondary level education. Specifically with regards to symptoms requiring immediate referral, it was observed between 20% to 40% either did not know or feel the need to refer a child with extreme symptoms such as child with diarrhea who is unable to drink or breast feed. In the similar evaluations of CHWs done by CORE group [3] and in [4], it has been mentioned that CHWs may have difficulty

identifying key signs when multiple signs are present. But complex assessment and treatment guidelines or flowcharts are not the solution as they might make it more difficult for CHW to reach a conclusion. It is also important to limit the number of questions to collect essential information to avoid fatiguing either the CHW or patient [3]. Additionally CHWs are not used to finding information inside a text and therefore the manuals are not well suited for their work [5].

Even though CHWs link patients and health-care providers in the delivery of health services and in increasing health awareness, concerns about the access and quality of care still remain. This can be addressed by only an increased collaboration and technical assistance supplemented by continued training [6]. As identified by [7], mobile based solutions can empower CHWs in key aspects such as data collection, training and access to reference material, facilitating communication among health workers, providing job aids and decision support, supervision of CHWs, and promoting healthy behaviors in the population. The traditional mobile systems based on voice and SMS contribute in few of these areas. But despite their ease of deployment, they are not well suited to handle complex information exchange and point of care diagnosis. Joint effort among policy makers, organizations and front line workers is not new but the dynamics of collaboration among CHWs and health-care providers is a unique challenge and requires further research. As mobile technology is widespread in developing countries where other forms of communication infrastructure is either sluggish or non-existent, a mobile collaborative solution would be more effective.

We present *Rafiki*¹, a system for mobile and wearable computing devices which guides CHWs through the diagnostic process and facilitates collaboration among various personnel involved in community health-care. Rafiki identifies possible diseases based on the patient context (e.g., age, gender, location, profession, other diseases in the area). With the help of patient response, a semantic DL reasoner [8], and OWL ontologies [9] that model health-care knowledge, Rafiki suggests questions to be asked of the patient. The usage of a mobile phone based diagnostic tool can also eliminate potential human errors in identifying symptoms and recognizing patients who require attention from health-care provider. Additionally, Rafiki can facilitate the collaboration among the involved parties (i.e., CHWs, Community Health Centers, patients)

¹Rafiki is a Swahili word which means friend.

by exchanging and managing new knowledge automatically using ontologies (for example, information about outbreaks of illnesses defined by health-care providers). Through a semantic and collaborative approach, Rafiki assists CHWs with an up-to-date knowledge regarding health-care.

In summary, the main benefits of our system are:

- 1) It helps in sharing disease, patient context, and demography related facts by using semantic management of data.
- 2) It facilitates the collaboration between CHWs and health-care providers for community health-care through cloud based and Peer-to-Peer (P2P) networks.
- 3) It provides decision support to CHW by reasoning over the knowledge base consisting of ontology, instances, and rules.
- 4) It provides CHWs with a mobile platform for accessing information at point-of-care in underserved areas.

The remainder of this paper is organized as follows. In Section II, we explain the community health-care scenario using a motivating example and give an overview of the system. Section III shows the representation of knowledge managed by Rafiki. Collaboration through knowledge exchange is detailed in Section IV. Section V presents the decision support module for diagnosis. A prototype of the system for Android is presented in Section VI. Section VII surveys related works and Section VIII presents conclusions and future work.

II. OVERVIEW

In this section, we will review the current scenario of community health-care by looking at the example of ASHA which is a program for health activists in India defined as: “[...] the first port of call for any health related demands of deprived sections of the population, especially women and children who find it difficult to access health services.”. Then, we describe a motivating scenario where Rafiki can help Community Health Workers (CHWs) and finally, we describe the high-level system architecture of the system.

A. The Need for Rafiki

Through ASHA, India has one of the largest initiative of community health-care with more than 890,000 CHWs in underserved areas². Figure 1 shows the structure of the ASHA program which is similar to the structure of other community health initiatives. Notice that community health-care is group endeavor involving patients, CHWs, and Health-Care Providers (HCPs). The interaction of these participants (for example to educate CHWs in the diagnosis of common diseases) is done mainly through meetings. However, current methods of communication (bi-weekly meetings, monthly training classes, etc.) delay collaboration and bring down the quality of care provided.

CHWs are typically high school educated (as per ASHA requirements) and undergo training to acquire the knowledge required to perform as a health activist. They use simple forms and manuals for collecting information about the patient by filling in information about symptoms and patient demographics as well as diagnosing the patient. The forms reflect the

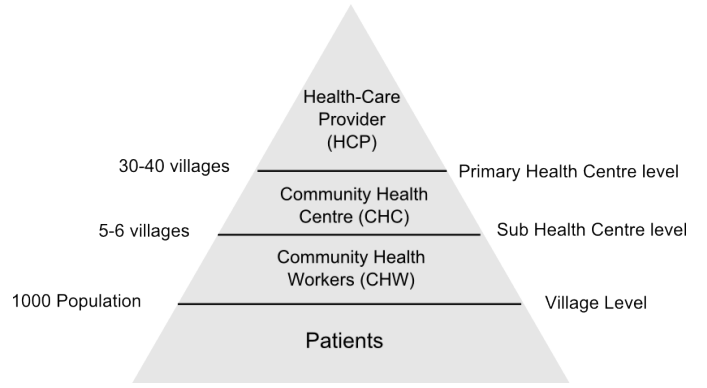


Fig. 1. Typical structure of a community health initiative. Image adapted from the National Rural Health mission (<http://nrhm.gov.in/nhm/nrhm/nrhm-framework-for-implementation.html>)

methods followed by doctors and other health-care providers in health centers and hospitals. Let us look at a possible scenario developed on the basis of the ASHA manual [10] which illustrates the need for Rafiki and points out inefficiencies of CHW program mentioned in [2].

Imagine a CHW who arrives at a house in Sirpur, a small village in the Chattisgarh state of India, after being informed of a child presenting some problems. Kumar, a six-year old boy, has been experiencing vomiting according to his parents. The CHW, following the manual, starts collecting details about Kumar and his problem using specific forms. Finally, the child is identified as having diarrhea and the CHW provides the parents with oral rehydration solution (ORS) and advises them on food intake. Even though the CHW indicated that Kumar suffers from severe dehydration, a dangerous and potentially life-threatening condition, she did not feel the need to refer the child.

Rafiki can help in this situation by reminding the CHW what questions she should ask, alerting her about possible emergency situations, and keeping her updated with the latest information related to diagnosis. In the following, we present the high-level architecture of our semantic and collaborative approach to this problem.

B. System Overview

Rafiki’s primary goal is to guide CHWs through the diagnostic process. To achieve this goal, it utilizes information about correlations between diseases, symptoms, and patient context. Rafiki obtains this information from various sources such as ontological facts and definitions, patient responses, mobile device sensors, and through collaboration with other devices and personnel. The architecture of the system (see Figure 2) has the following main components:

- 1) *Knowledge Management* (Section III) is responsible for knowledge representation, reasoning over the knowledge base (KB), and also provides an interface for other modules to interact with the knowledge base. The knowledge base is based on an OWL ontology which includes class definitions and SWRL rules, as well as instances or facts. The reasoner is used to infer interesting information coming from facts gathered by the system and that has not been explicitly stated.

²<http://nrhm.gov.in/communitisation/asha/asha-data.html>

- 2) *Collaboration Management* (Section IV) handles the communications with other Rafiki devices through P2P and cloud-based networks. It also selects the knowledge that has to be shared with other Rafiki enabled devices depending on their profile.
- 3) *Decision Support* (Section V) drives towards the primary goal of the system that is, help CHWs in medical diagnosis. It gathers information about patient context and comes up with a reduced set of possible diseases using the reasoner. This component also suggests questions to be asked of the patient by querying the knowledge base.

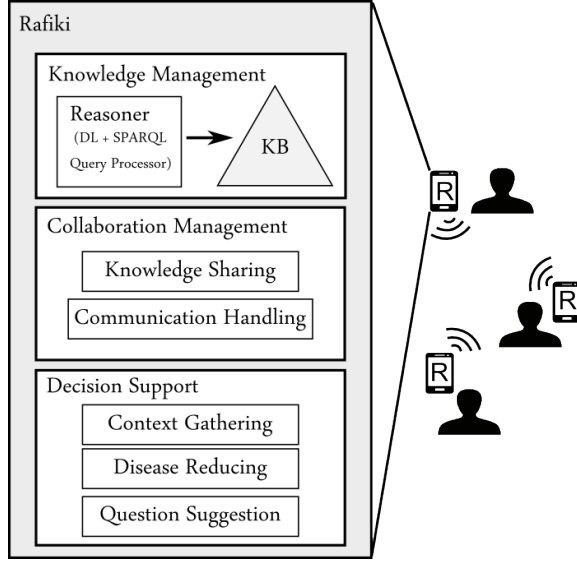


Fig. 2. High-level architecture of the system.

As Figure 2 shows, Rafiki runs on different mobile devices that can be considered nodes of the system. These devices belong to CHWs, health-care providers, and possibly patients. Apart from assisting CHWs, Rafiki can also help patients with mobile devices by providing information about their treatments as well as by gathering context information which can be further shared with CHWs. If a patient does not have a mobile device, this information is collected by the CHW by asking questions. Also, Rafiki helps authorities by providing them with demographic information about diseases gathered by CHWs while diagnosing patients.

III. KNOWLEDGE MANAGEMENT

Rafiki uses OWL ontologies to model the knowledge related to community health-care such as diseases, symptoms, questions, and patient context. By using an ontology, “an explicit specification of a conceptualization” [9], Rafiki facilitates knowledge exchange and reasoning over the knowledge. Rafiki can manage any information defined based on the ontologies. Therefore, health-care providers can disseminate new knowledge to CHWs devices quickly using this approach (see Section IV).

Rafiki’s ontology³ can be split into three main modules: diagnosis, patient, and stats. As they are part of the same

ontology, the class definitions are same across these three segments. Next we explain the main classes and properties of each of these modules:

The *diagnosis module* models the basic knowledge related to diagnosis (i.e., diseases, symptoms, and questions) that Rafiki manages (see Figure 3). Other medical ontologies, such as SNOMED [11] or the Unified Medical Language System (UMLS) [12], can be linked to extend information about diseases and symptoms. The following question types (related to diseases, symptoms, or patient context) are included in the module: 1) objective questions with positive/negative answers (for example, *Is the patient suffering from abdominal pain?*); 2) measurement questions with an associated metric for answers (for example, *Is patient’s mouth dry?* with answer options as *very dry*, *little dry* and *no*; and 3) contingency questions in which the patient is asked a filter question to determine if she needs to be asked other related questions (for example, *Is your stomach upset?* –after a positive answer the patient is asked about the kind of food recently taken). Additionally, authorities can embed action based questions which requires CHW to perform an action such as performing measurements, collecting samples, or taking pictures.

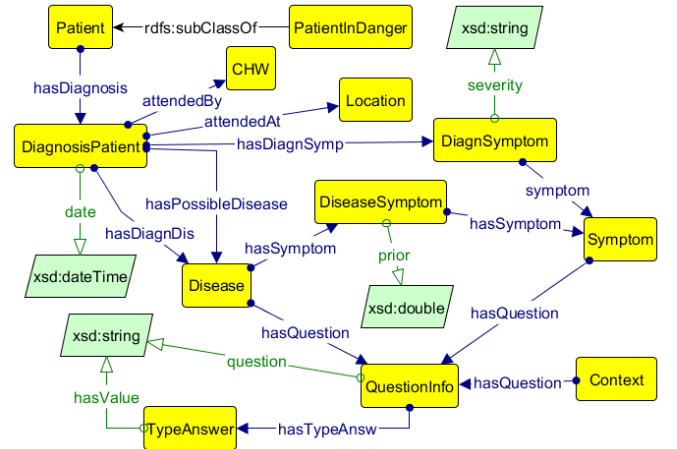


Fig. 3. Excerpt of Rafiki’s ontology: diagnosis module.⁴

The *patient module* includes all the information which is related to the patient which might be of relevance for diagnosis of the disease (see Figure 4). The primary patient information considered by Rafiki are location and patient demographic. By using these, Rafiki can extract other secondary pieces such as statistical and demographic information related to diseases and symptoms as well as factors related to location such as weather or season. Any other piece of patient information that could be interesting in the future can be added by incorporating the necessary classes and properties to this ontology.

The *stats module* models the aggregated health-care information related to various regions (see Figure 5). The goal of this ontology is to disseminate knowledge discerned by health-care providers as we will explain in Section IV. Notice that some classes, properties, and instances defined for locations

³The ontology is available at <http://ebiquity.umbc.edu/ontologies/rafiki/v1/rafikiOnt.owl>.

⁴The notation employed for the figure follows the Graffoo specification (<http://www.essepuntato.it/graffoo>): yellow boxes for classes, green boxes for datatypes, blue arrows for object properties, green arrows for data properties, and black arrows for predicates.

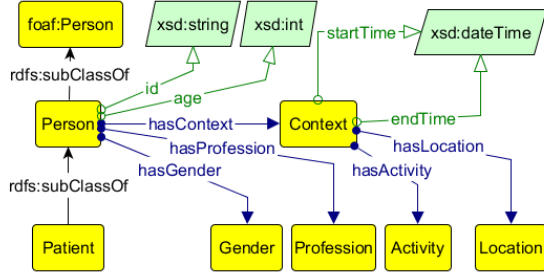


Fig. 4. Excerpt of Rafiki's ontology: patient context module.

in the Rafiki ontology extend entities from DBpedia [13], the semantic entry point to Wikipedia. This way, they are linked to other interesting information such as their geographic location following the W3C recommendation for geotagged RDF⁵.

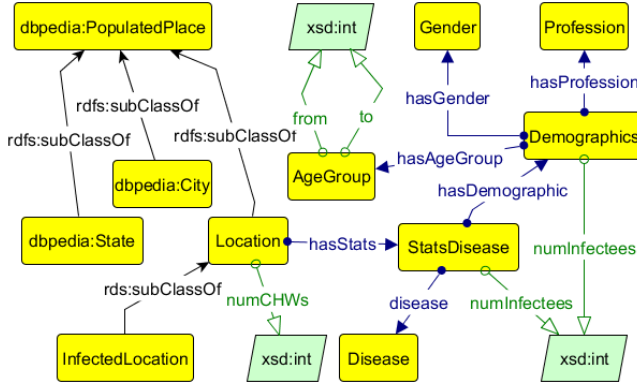


Fig. 5. Excerpt of Rafiki's ontology: stats module.

The use of an ontology to model Rafiki's knowledge enables the system to infer interesting information by using a semantic DL reasoner [8]; a software able to infer logical consequences from a set of facts. For example, suppose that six cases of diarrhea have been diagnosed in the village of Sirpur in Chattisgarh in the last month by various CHWs and the health-care providers in the area have come up with the following guideline: "if more than five patients have been diagnosed in a region with the same disease, the region should be classified as a potential infected region". This guideline has been added to Rafiki by the following Semantic Web Rule Language (SWRL) rule:

```
Location(?l) ^ hasStats(?l,?s) ^ Disease(?d) ^
disease(?s,?d) ^ numInfectees(?s,?num) ^
swrlb:greaterThan(?num, 5)
→ InfectedLocation(?l)
```

The reasoner can classify any region that fulfills the requirements as a *raf:InfectedLocation* automatically when facts about the region are created. In addition, using a local reasoner on the device is especially interesting in this scenario as Internet access is not guaranteed and the personal identifying information used in the reasoning process will remain on the device. The feasibility of running semantic reasoners on current mobile devices has been shown in [14].

The Rafiki application running on each user device (i.e., patients, CHWs, authorities) uses the ontology described above as a default. However, depending on the type of user, the information stored on her device may vary both in form of classes, properties, and instances (facts). For example, Rafiki on a patient device will contain facts related to her location, demography that will be shared with CHWs during the diagnosis. Also, Rafiki on an authority device will contain statistical facts generated by aggregating the knowledge gathered by different CHWs.

IV. COLLABORATION AND KNOWLEDGE EXCHANGE

Collaboration is an important aspect of community health-care in underserved areas [5]. The goal of collaboration in Rafiki is the knowledge exchange between various participants involved in community health-care to enrich their local knowledge bases. The exchange of knowledge among devices through wireless communications can add important benefits to mobile systems as shown in [15]. So, collaboration helps in:

- 1) Updating the local knowledge of Rafiki on the mobile device of each CHW.
- 2) Sharing the patient diagnosis and demography related facts with other CHWs and health-care providers.
- 3) Relaying patient health record related information to health-care providers.

To achieve these goals, Rafiki manages communications among devices and selects the appropriate knowledge that they should exchange.

A. Interactions in Community Health-Care

Each of the participant exchange information with others (see Figure 6) using Rafiki. However, the kind of information exchanged and the approach used may vary according to the participants involved:

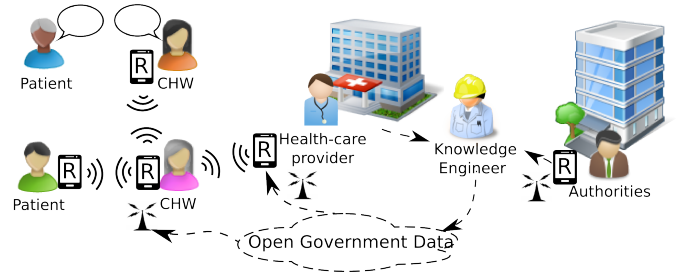


Fig. 6. Collaboration among the main participants of the system.

Patient and CHW: Patients share information related to diseases, symptoms, and their context with the CHW as answers to the questions asked in the diagnostic process (see Section V). However, in recent years we are witnessing initiatives to create affordable smart devices for developing countries and therefore, Rafiki can handle scenarios where both the patient and the CHW are equipped with a smart device. Rafiki manages the communication and exchange of information between those devices when they are in short communication range (leveraging P2P networks). Thus, CHW can save time by avoiding the first phase of diagnosis (see Section V-A).

⁵<http://www.w3.org/2005/Incubator/geo/>

CHW to CHW: CHWs act as “knowledge disseminators” when communicating among themselves. The goal of this intercommunication between CHWs is to make sure that every CHW would have the latest information possible to perform a most accurate diagnosis. On one hand, the Internet infrastructure in remote areas of developing countries is unreliable or missing. On the other hand, it is unnecessary that every CHW should have to visit the community health center for updated models in their Rafiki knowledge base. So, CHWs exchange their models when they come in touch with their peers. This minimizes the time required for the spread of latest information regarding diagnostic process. The patient context, except personally identifying information, is shared with other CHWs as it can help in identifying spread of contagious diseases in nearby villages. This includes information about diagnosis (e.g., the symptoms detected and possible diseases of a patient) which can help in updating the demographic information about infectees in an area that CHWs manage, even before this information has been collected, aggregated, and published by health-care providers.

CHW and Health-Care Providers: During their rounds, CHWs collect information in relation to patient’s disease and context for diagnosis through Rafiki. This information is relayed to community health centers (during regular meetings or through the cloud) which maintain a central health record for all patients in villages under their jurisdiction. This helps to maintain an Electronic Health Record (EHR) for the patient. This information shared by CHWs in the form of RDF facts can be stored in a open source health information system like OpenMRS⁶. This type of data integration across two different systems have been explored before [16] and has advantages such as being able to provide continuity of care and querying across systems.

With help of domain experts, community health centers can analyze the information shared by CHWs and come up with diagnostic guidelines relating new symptoms or patient context to diseases. Suppose CHWs are reporting that malaria is spreading more among diarrhea patients with symptoms of high dehydration (based on symptom questions), then it would be helpful to share this information with other CHWs as a guideline. A knowledge engineer can express this information using the structure of Rafiki’s ontology in the form of a SWRL rule. Thus Rafiki can suggest CHW to make sure that patients with diarrhea are well hydrated in a region with history of malarial parasites.

Authorities to Public: The knowledge shared by Rafiki using Semantic Web technologies can be easily published by authorities (CHW organizations, hospitals, governments) following the principles of Linked Data [17]. This way, Rafiki helps them to participate in the Open Government Data (OGD) initiative that advocates making government data available to the public⁷. A knowledge engineer can extract, with help of SPARQL queries, the facts related to the diagnosis instances from CHWs and use it to populate the stats module of the ontology. This information, which is shared to Rafiki on CHW devices, can also be interesting for developers and the general

public (for example, this data can be used to create health-care applications).

B. Managing Communications

Rafiki supports two communication mechanism for knowledge exchange: 1) Internet-based approach and 2) a Peer-to-Peer (P2P) approach. The first approach expedites the dissemination of knowledge between health-care providers and CHWs as well as CHWs themselves. However, in rural and remote areas, Internet access is not guaranteed and can be unreliable. Therefore, Rafiki leverages P2P ad hoc networks using other wireless communication mechanisms (e.g., Bluetooth and WiFi) to update their knowledge base and share interesting facts.

Algorithm 1 presents the steps performed by Rafiki on a CHW’s device for updating its local knowledge. We selected the CHW device as a CHW is the primary actor in our community health-care scenario and directly communicates with various participants involved using both communication mechanisms mentioned above.

Algorithm 1 Algorithm for updating the knowledge base and exchanging information.

```

if Internet available then
    newOnt=connect to Rafiki Web service
    integrate(newOnt)
    shareDiagnosisInformationToHCP()
end if
enable Bluetooth and WiFi
devices=discover CHW devices nearby
for each devicei in devices do
    Pair with devicei
    shareDiagnosisInformationToCHW()
    for each modulei in KB do
        Exchange meta-data of modulei
        if remote version outdated then
            send(modulei,devicei)
        else
            newOnt = request(devicei.modulei,devicei)
            integrate(newOnt)
        end if
    end for
end for

```

First, if Internet connection is available, Rafiki tries to connect to the Rafiki Web service where health-care providers make accessible new knowledge for CHWs. If successful, Rafiki downloads the new ontology (classes, properties, rules, instances). When a new version of the ontology is downloaded, Rafiki integrates it with its local ontology. In the integration process, Rafiki first backs up the patient diagnostic information collected by the CHW (appropriate instances of the diagnosis module of the ontology) on her device. Then, to avoid possible inconsistencies during the integration (e.g., imagine that the infant group age definition changed), Rafiki removes all the previous knowledge in the stats and diagnosis modules of the local KB. Finally, the system incorporates the back-up to the new knowledge base.

In the next step, Rafiki enables the wireless communication mechanisms on the device to discover other nearby devices. For each such discovered device, Rafiki pairs with it by

⁶<http://openmrs.org>

⁷In the case of India, our running example along the paper, more information about the Indian Open Government Data initiative can be found in <http://data.gov.in>.

verifying its Unique Identification Number. After pairing, the meta-data (version number and time stamp) of various modules in KB is exchanged (this meta-data is modeled in the ontology through the use of annotations –using *owl:versionInfo* and *rds:comment* properties–). By comparing the meta-data, the older version of the two modules is identified and that device requests for the newer version of the module which is then integrated.

The other goal of collaboration is to exchange facts related to patient diagnosis and demography between CHW as well as relaying information to health-care providers. While the former helps in reducing the list of possible diseases (see Section V), the latter is a faster method for storing patient records centrally. As the challenges of sharing electronic health records with authorities (see *shareDiagnosisInformationToHCP* in Algorithm 1) have been explained well by previous works [18], we explain the sharing of this information among CHWs (*shareDiagnosisInformationToCHW*). First, the patient information is de-identified to make sure that no sensitive information is shared between CHWs. For this, a unique hash is generated for a patient based on the CHW id and location. Then, patient context is de-identified by going from a specific definition of the class to a more general one in the ontology (for example, aged 6 to in age group 1-12). After this step, the information is shared with other CHWs through P2P communications. This de-identification process is also performed when generating the information that authorities will publish as part of OGD.

Depending on the participants, for other types of interactions the steps will be similar or slightly different but the information exchanged will vary. For example, health-care providers can push updated modules to CHW if necessary through the Internet and patients can share their context information through Bluetooth, as explained in the previous section.

V. DECISION SUPPORT FOR DIAGNOSIS

Rafiki assists CHWs by suggesting the questions to be asked of the patient to perform her diagnosis. Similar to the diagnosis currently performed by CHWs (see Section II), Rafiki goes through a feedback driven process to identify the most likely disease as well as remedial measures related to it. In addition, as an information system, Rafiki addresses the shortcomings of the current approach (e.g., CHWs forgetting to ask important questions to the patient, or ignoring signs which requires referral of the patient). The diagnosis process (see Figure 7) starts with Rafiki gathering information about the patient. Then, this information is used to select the most likely diseases. Finally, Rafiki suggests questions related to the symptoms of these diseases to the CHW.

A. Context Gathering

In the first phase of the diagnostic process, Rafiki gathers patient context information (i.e., location and demography). This process can be performed in two ways depending upon the availability of the smart device or other wearable sensors for the patient. As explained in Section IV, if the patient has a mobile device with Rafiki, it will share the facts gathered about her recent context. Otherwise, patient context information can

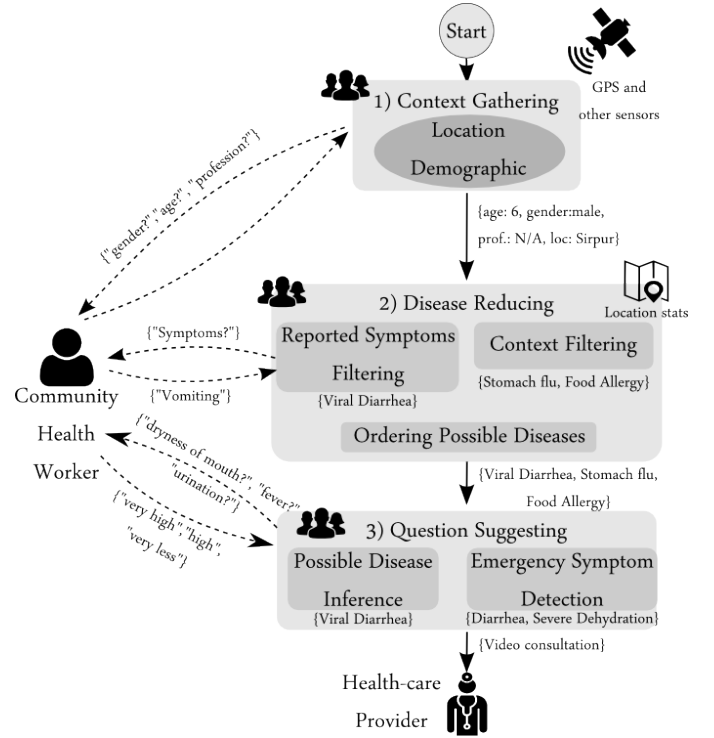


Fig. 7. Steps performed by Rafiki to help a CHW diagnosing a patient.

be gathered by the CHW device itself. The GPS sensor on the CHW device is used to find out the location of the patient. After identifying the location, it is used to fill in other pieces of information related to location (such as weather) from Web services (a list of useful Web services is included in the ontology). Also, the demographic information such as age, gender, and profession is collected through patient responses to questions modeled in the diagnosis module of Rafiki's ontology.

In our motivating scenario, Rafiki on the CHW device will obtain the location of the patient (Sirpur) and Kumar's parents will be asked about the age and gender of the kid. Therefore, the following information will be entered into Rafiki: *Patient: name=Kumar, age=6, gender=male, location=Sirpur.*

B. Reducing the Set of Diseases

The information gathered about the patient context, along with the stats module of the ontology, is used to infer the list of possible diseases. The goal of this step is to figure out the possible diseases based on others with similar context (e.g., neighbors from the same age group). Rafiki on a CHW device receives statistical information about disease in her region (as explained in Section IV) and then applies the following rule to extract the list of possible diseases based on patient context (in this example, age group):

```

Person(?p) ∧ hasLocation(?p,?l) ∧ age(?p,?ageP) ∧
InfectedLocation(?l) ∧ hasStats(?l,?s) ∧
disease(?s,?d) ∧ Disease(?d) ∧
hasDemographics(?s,?dem) ∧ hasAgeGroup(?dem,?ageG) ∧
from(?ageG,?minAge) ∧ to(?ageG,?maxAge) ∧
swrlb:lessThanOrEqual(?ageP,?maxAge) ∧
swrlb:greaterThanOrEqual(?ageP,?minAge)
→ hasPossibleDisease(?p,?d)

```

Continuing with our motivating scenario, several cases of diarrhea were detected in Sirpur and so the region is classified as *raf:InfectedLocation* (see the rule in Section III). In addition, stomach flu was diagnosed to other kids in Sirpur so it could be a diarrhea outbreak and the disease is added as a possible one for Kumar (i.e., *raf:hasPossibleDisease(Kumar,StomachFlu)*).

Other interesting information (e.g., is it the monsoon season?) can be added to this step easily by incorporating more rules or adding new antecedents to the previous rule. The use of Semantic Web techniques means that these new rules will be managed automatically by the reasoner without modifying the system (Rafiki gets the list of possible diseases by checking the *raf:hasPossibleDisease* property). Also, information like reported symptoms can be used to add more diseases to the list and reorder it. Reported symptoms can be common symptoms like fever, vomiting, or aches for which the patient might have got in touch with a CHW. For example, to add more diseases to the list based on these symptoms, the following rule is applied:

```
Person(?p)  ^ hasSymptom(?p,?s) ^
Disease(?d) ^ hasSymptom(?d,?s)
  → hasPossibleDisease(?p,?d)
```

In the case of Kumar, he was suffering from vomiting and therefore, viral diarrhea (gastroenteritis) will be added to the list of possible diseases. Then, Rafiki ranks this list according to the information gathered about reported symptoms and patient context. As an example, the SPARQL query of Figure 8 obtains from the knowledge base a ranked list of possible diseases where: 1) diseases are ranked according to the number of their reported symptoms; 2) for diseases with the same number of reported symptoms, the number of infectees for each disease with similar patient context (e.g., location and age group) is used. For Kumar, viral diarrhea is the first disease to check as he suffers from a known symptom of this disease and it is common among other children in the area. We acknowledge that some symptoms are stronger indicators than others for some diseases. By incorporating weights or probabilities (as we will explain in Section V-D), Rafiki would be able to handle this.

C. Suggesting the Diagnostic Questions

With the ranking of possible diseases ready, the goal of the CHW is to ask questions related to their symptoms in order to detect the most likely disease. This way, Rafiki retrieves questions from the diagnosis module of the ontology by executing a SPARQL query based on the disease and symptoms. The CHW inputs the patient response by selecting the answer from the available options by a button click or through voice command. We explain more details of data entry in Section VI.

A critical task during the diagnostic process is to detect *emergency symptoms* which requires that the patient should be immediately referred to the health-care provider. The definition of an emergency symptom in Rafiki's ontology is done based on the knowledge from experts or specific to a region. If a symptom qualifies as an emergency symptom and the patient response to the question is positive, the patient is recognized as requiring urgent attention or care. Going back to the example, severe dehydration is a sign of worry in patients, especially in children, with diarrhea as a possible disease and should

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX raf: <http://ebq.umbc.edu/ontology/rafiki/>
SELECT ?Disease (COUNT(?S) as ?Symptoms) ?Infectees
WHERE {
  ?Patient a raf:Patient;
    raf:id "42"^^xsd:string;
    raf:hasSymptom ?S;
    raf:hasLocation ?L;
    raf:hasPossibleDisease ?Disease;
    raf:hasAgeGroup ?AG.
  ?Disease a raf:Disease;
    raf:hasSymptom ?S.
  ?L a raf:Location;
    raf:hasStats ?Stats.
  ?Stats a raf:StatsDisease;
    raf:disease ?Disease;
    raf:hasDemographics ?Demographics.
  ?Demographics a raf:Demographics;
    raf:hasAgeGroup ?AG;
    raf:numInfectees ?Infectees.
}
GROUP BY ?Disease ?Infectees
ORDER BY DESC(?Symptoms) DESC(?Infectees)
```

Fig. 8. An example of SPARQL query to return the list of possible diseases for a patient, ordered by symptoms and patient context.

be immediately referred to the health-care provider. The rule capturing this guideline can be expressed as follows:

```
Person(?p) ^ hasPossibleDisease(?p,Diarrhea) ^
hasDiagnosis(?p,?dp) ^ hasDiagnSymp(?dp,?s1) ^
symptom(?s1,DryMouth) ^ severity(?s1,"Severe") ^
hasDiagnSymp(?dp,?s2) ^ symptom(?s2,Fever)
  → hasEmergencySymptom(?p,SevereDehydration)
```

Missing emergency symptoms, even after having collected information about them, is a surprisingly common and dangerous tendency among CHWs [2]. Rafiki with the help of the rules used to define these symptoms and the reasoner, can make sure that this information is not missed. At each step of diagnosis, Rafiki verifies if the patient has been classified under the *raf:UrgentCareNeededPatient* (*raf:Patient* and *raf:hasEmergencySymptom* some *raf:DiagnSymptom*) class. More complex relationships can be expressed using rules for classifying a patient under the aforementioned class. Various types of actions can be associated with such situations such as video conferencing with a health-care provider (doctor or nurse), sending an ambulance to bring in the patient, etc. By default Rafiki starts a call or videocall (if available) with the nearest health-care provider. In this scenario wearable devices such as Google Glass can be very helpful as they would enable the health-care provider to immediately start attending the patient through the CHW (see Section VI).

If the emergency symptom is not severe, for example if the patient has mild dehydration, CHW is suggested to give him a medicine (ORS in our scenario). Further on, Rafiki continues collecting more information from the patient about symptoms related to the possible diseases (identified in first phase) through questions. It is possible that some of the diseases might have a common set of symptoms and Rafiki does not repeat the same symptom question for those diseases. For example, vomiting and fever is seen both in malaria and viral diarrhea patients. Based on the number of questions positively answered for a disease, Rafiki computes a simple score which denotes the

possibility of the disease (the higher the score the more likely for that disease to be the cause). This approach is conceptually similar to the use of the Thrombolysis In Myocardial Infarction (TIMI) risk score for classifying patients with chest pain [19]. In our current prototype, since all symptoms and diseases have equi-probable priors, Rafiki simply counts the number of symptoms satisfied by using the previous SPARQL query from Figure 8. In a probabilistic approach (see Section V-D), we have to compute the maximum likelihood of a disease based on patient context and symptoms.

In some cases, measurements or analysis might be required before further continuing diagnosis. For example, Rafiki would suggest CHW to do a blood test to check for a patient with malaria as a possible disease. The diagnostic process is completed when tested positively for emergency symptoms or after asking all questions related to possible diseases. As it happens in real life community health-care, a patient could be suffering from a disease which was not included among the possible ones. In this case, Rafiki can highlight information about previous diagnosis to the same patient to the CHW and health-care providers.

D. Incorporating a Probabilistic Approach to Diagnosis

Probabilistic decision support systems like [20] have been proposed for medical diagnosis. A stochastic model based on probabilities might be more appropriate for expressing fuzzy relationships between some symptoms and diseases. Probabilities can also express relative importance of symptoms with respect to diseases which might improve the ranking of possible diseases by the decision support component. For example, while fever and dehydration are symptoms of diarrhea, a dehydrated patient is more likely to have diarrhea than the one with just fever. Additionally, uncertainties in data and knowledge either due to noise or incompleteness can be conveyed using this model.

To adapt Rafiki to perform probabilistic reasoning, the following additions will have to be made. First, information about priors of diseases have to be introduced into Rafiki's ontology; it already contains the needed properties to model these probabilities (e.g., the data property *raf:prior* that attaches a probability to a symptom and disease). For example, based on the stats model, if dehydration is seen more common in diarrhea patients then a higher prior is attached to this relationship. By using a probabilistic description logic reasoner such as [21], Rafiki would be able to reason under uncertainty. We are further investigating this issue and further discussion of this topic is out of the scope of this paper.

VI. PROTOTYPE OF THE SYSTEM

We have implemented a prototype of Rafiki focusing on the decision support and knowledge representation aspects to show the feasibility of our approach. The prototype is an Android application for smart phones and Google Glass. For the management of knowledge and ontologies in our prototype we used the OWL API⁸, an ontology API to manage OWL 2 ontologies in Java applications. We used HermiT [22] as the

semantic DL reasoner, specifically the HermiT version for Android⁹.

Mobile Application: We illustrate the prototype using the motivating scenario of Section II and Figures 9. First, Rafiki shows questions related to patient (Kumar in our scenario) context to the CHW (see Figure 9(a)). CHWs can enter this information either using the Android keyboard or by voice input by clicking on the speaker icon. We acknowledge that local languages are an issue, especially for the Indian scenario, and that currently we are limited by what Android offers. But in the future we expect that Android will support several world languages including dialects.

Then, the CHW is presented with a question to select a reported symptom, if any (see Figure 9(b)). The most common symptoms in the area with respect to the patient context are highlighted (in this case fever, vomiting, and loose motion). The video camera icon next to the symptom button will show a small video clip related to the of the symptom when clicked. After determining the most possible diseases (limited to 3 in the prototype) according to the context of the patient and the reported symptoms, Rafiki starts showing questions related to the symptoms of these diseases. For example, in the case of diarrhea (highest in the most possible diseases) one of these questions is to determine if the patient is dehydrated based on dryness of the mouth. (see Figure 9(c)).

Google Glass Application: Wearable computing devices like Google Glass has been widely used in hospitals for video consultation or health-care training. In this preliminary prototype, we are exploring its usability in community health-care scenario with visual guides and options for video conferencing. At any point during the diagnosis, CHW can click the Glass icon at the top right corner of the application to switch interfaces and communication happens through Bluetooth. Glass is used only as an interface and other components of Rafiki such as knowledge and collaboration management, decision support runs on the smartphone only. As a rudimentary example, imagine that the CHW is checking the patient's mouth with respect to dehydration. By switching to Google Glass, she can have an image of the symptom in front of her eyes without having to look away from the patient (see Figure 10(a)). Continuing with the example if the patient is found out to be suffering from severe dehydration and needs urgent care (see Figure 9(d)), Google Glass application can start a video call with a doctor or a nurse (see Figure 10(b)). Thus a first person point of view is possible for tele-consultation.

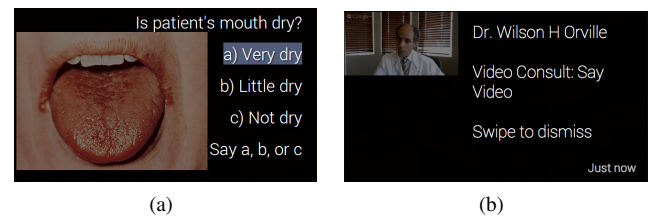


Fig. 10. Screenshots of Rafiki's prototype app for Google Glass: (a) receiving a visual clue to measure mouth dryness and (b) videocall with a doctor after detecting a patient that needs urgent care.

⁸<http://owlapi.sourceforge.net>

⁹<http://sid.cps.unizar.es/AndroidSemantic>

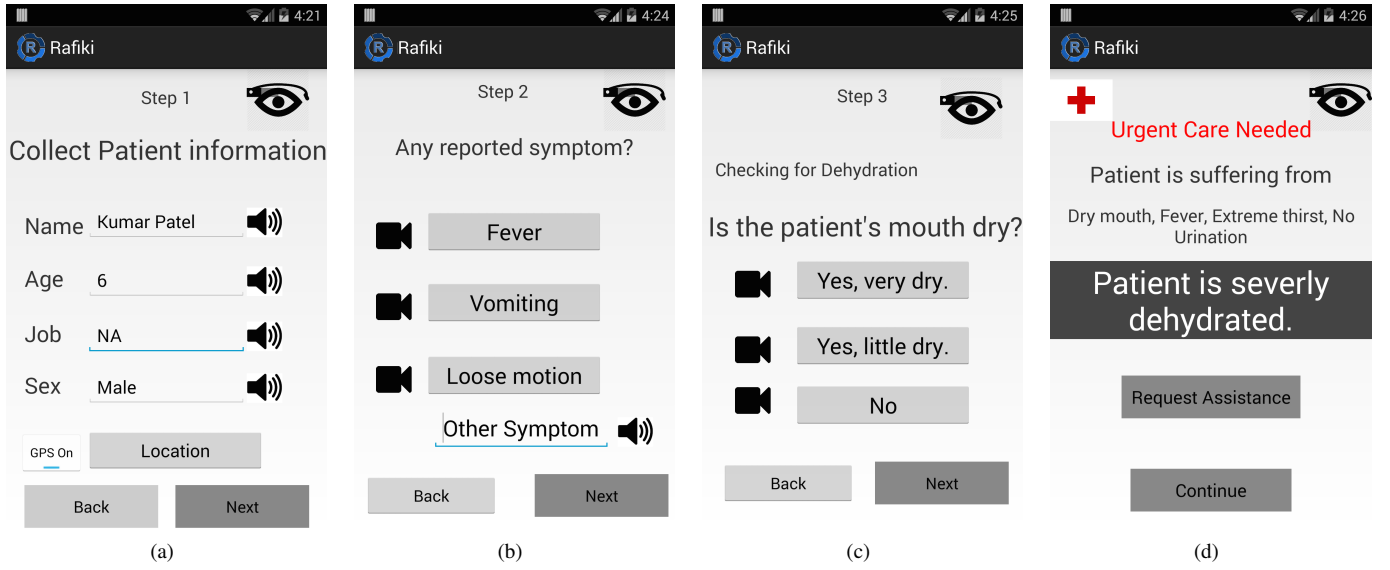


Fig. 9. Screenshots of Rafiki's prototype app for smartphones: (a) collecting the context of the patient, (b) asking about reported symptoms, (c) asking questions about the possible disease, and (d) alert for an emergency situation.

Publishing Health-care data: We have also included an example of the information collected by Rafiki's prototype that health-care providers could publish in an effort to support the Open Government Data (OGD) initiative. For this purpose, we used a tool¹⁰ developed in the University of Southampton that generates Keyhole Markup Language (KML) content from geo tagged RDF data to display this information on a map. We used as an input an RDF file which was generated by the system containing cases of diarrhea in different regions of India and extracted from the stats module of Rafiki's ontology. Then we used the resulting KML file and Google Maps API to display the information using colors depending on the amount of infectees (see Figure 11).

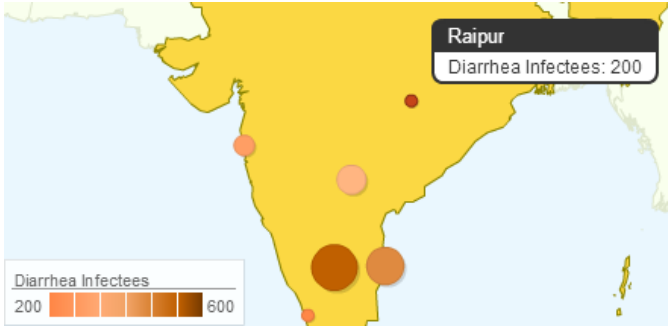


Fig. 11. Visualization of an example of data gathered by Rafiki for diarrhea infectees in India and published as OGD.

VII. RELATED WORK

The benefits of using mobile technology for empowering CHWs in developing countries are well known [5]. Open Data Kit (ODK) Clinic [23] is a phone-based clinical Decision Support Systems (CDSS) for health-care providers in low-resource settings. ODK clinic, which is a replacement for paper-based system, helps clinicians in managing patient

summaries electronically. The target user base is different from that of Rafiki as the former focuses on experienced users. Case management frameworks like CommCare [24] provide tools for managing diagnostic elements and tracking patient cases. While ODK Clinic and CommCare provide decision support for clinicians and CHWs, respectively, through forms specific to diseases, they do not reason over information being collected about the patient to identify potential emergencies, possible diseases, etc.

Other applications like eMocha¹¹ also use multimedia content in addition with forms for training of CHWs and educating patients about health-care. Other smart-phone based solutions (e.g. ampath¹²) are targeted towards aggregating data centrally from CHWs for maintaining health records and derivation of health statistics. All the aforementioned tools, focus on a vertical collaboration along community health-care scenario (remember Figure 1) while in Rafiki we are proposing a horizontal collaboration as well, for example among CHWs. Additionally usage of Semantic Web technologies in Rafiki makes it possible for it to take advantage of Internet of Things with low cost sensors and devices available.

On the other hand, there exists a variety of expert systems for decision support in various domains including medical diagnosis [25]. Probabilistic decision support systems for medical diagnosis, like [20], are not usually geared towards mobile systems and do not support information sharing. As explained in Section V, we have been exploring the addition of probabilistic diagnosis to Rafiki. We have used patient context to reduce the list of possible diseases, but the patient context can be extended further to include more interesting information as in [26].

¹⁰<http://graphite.ecs.soton.ac.uk/geo2kml>

¹¹<http://main.ccghe.net/content/emocha>

¹²<http://www.ampathkenya.org>

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented Rafiki, a system that guides Community Health Workers (CHWs) through the diagnostic process and facilitates the collaboration among personnel involved in community health-care. Rafiki uses Semantic Web technologies such as ontologies and reasoners to model, manage, and reason over the knowledge. In addition, Rafiki exploits different communication mechanisms to share useful information among the personnel. As a summary, the main benefit of Rafiki to community health-care scenario are:

- Access to health materials and providing of decision support for CHWs.
- Usage of patient context to reduce the number of questions asked.
- Detecting emergency situations automatically and providing on the scene assistance.
- Collection of patient health-care data through surveying.
- Increased collaboration among CHWs for overcoming limitations of infrastructure.

We have developed a prototype of Rafiki for Android and Google Glass with a limited set of diseases, symptoms and questions. As future work we would like to conduct controlled trials with organizations like ASHA in developing countries to validate our system and further explore aspects such as: learning curve of the system, interaction of CHWs with the system and, cost and time savings in patient visits.

ACKNOWLEDGMENT

We would like to thank Dr. Manjari Joshi and Dr. Michael Grasso for their technical support regarding community health-care and the diagnostic process, respectively. This research work was supported by the NSF grants 0910838 and 1228673 and the CICYT projects TIN2010-21387-C02-02, TIN2013-46238-C4-4-R, and DGA-FSE.

REFERENCES

- [1] R. F. AbuAlRub, F. El-Jardali, D. Jamal, A. S. Iblasi, and S. F. Murray, "The challenges of working in underserved areas: a qualitative exploratory study of views of policy makers and professionals," *Int. J. of Nursing Studies*, vol. 50, no. 1, pp. 73–82, 2013.
- [2] S. Shrivastava and P. Shrivastava, "Evaluation of trained accredited social health activist (ASHA) workers regarding their knowledge, attitude and practices about child health," *Rural & Remote Health*, vol. 12, no. 4, 2012.
- [3] C. G. C. Initiative, "Community case management of childhood illness," URL: <http://www.coregroup.org/our-technical-work/initiatives/ccm>.
- [4] S. Rowe, M. Olewe, D. Kleinbaum, J. McGowan, D. McFarland, R. Roachat, and M. Deming, "Longitudinal analysis of community health workers? adherence to treatment guidelines, Siaya, Kenya, 1997–2002," *Tropical Medicine & International Health*, vol. 12, no. 5, pp. 651–663, 2007.
- [5] D. Ramachandran, J. Canny, P. D. Das, and E. Cutrell, "Mobile-izing health workers in rural india," in *SIGCHI Conf. on Human Factors in Computing Systems*, 2010, pp. 1889–1898.
- [6] A. Witmer, S. D. Seifer, L. Finocchio, J. Leslie, and E. H. O'Neil, "Community health workers: integral members of the health care work force," *American J. of Public Health*, vol. 85, no. 8_Pt_1, pp. 1055–1058, 1995.
- [7] B. DeRenzi, G. Borriello, J. Jackson, V. S. Kumar, T. S. Parikh, P. Virk, and N. Lesh, "Mobile phone tools for field-based health care workers in low-income countries," *Mount Sinai J. of Medicine: A J. of Translational and Personalized Medicine*, vol. 78, no. 3, pp. 406–418, 2011.
- [8] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A practical OWL-DL reasoner," *J. of Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007.
- [9] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *Int. J. of Human-Computer Studies*, vol. 43, no. 5-6, pp. 907–928, 1995.
- [10] M. of Health & Family Welfare of the Government of India, "Accredited social health activist (ASHA) manual," URL: <http://nrhm.gov.in/communitisation/asha/asha-manual.html>.
- [11] K. A. Spackman, K. E. Campbell, and R. A. Côté, "SNOMED RT: a reference terminology for health care," in *AMIA Annual Fall Symposium*. American Medical Informatics Association, 1997, p. 640.
- [12] D. A. Lindberg, B. L. Humphreys, and A. T. McCray, "The unified medical language system," *Methods of Information in Medicine*, vol. 32, no. 4, pp. 281–291, 1993.
- [13] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia - a crystallization point for the Web of Data," *J. of Web Semantics*, vol. 7, no. 3, pp. 154–165, 2009.
- [14] R. Yus, C. Bobed, G. Esteban, F. Bobillo, and E. Mena, "Android goes semantic: DL reasoners on smartphones," in *2nd Int. Workshop on OWL Reasoner Evaluation (ORE 2013)*, 2013, pp. 46–52.
- [15] R. Yus, E. Mena, S. Ilarri, and A. Illarramendi, "SHERLOCK: Semantic management of location-based services in wireless environments," *Pervasive and Mobile Computing*, 2013.
- [16] A. F. Guidry, J. L. Walson, and N. F. Abernethy, "Linking information systems for hiv care and research in kenya," in *First ACM Int. Health Informatics Symposium*, 2010, pp. 531–535.
- [17] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data-the story so far," *Int. J. on Semantic Web and Information Systems*, vol. 5, no. 3, pp. 1–22, 2009.
- [18] R. Wu, G.-J. Ahn, and H. Hu, "Secure sharing of electronic health records in clouds," in *8th Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2012)*, 2012, pp. 711–718.
- [19] A. C. Morris, D. Caesar, S. Gray, and A. Gray, "Timi risk score accurately risk stratifies patients with undifferentiated chest pain presenting to an emergency department," *Heart*, vol. 92, no. 9, pp. 1333–1334, 2006.
- [20] B. Wemmenhove, J. M. Mooij, W. Wiegnerinck, M. Leisink, H. J. Kappen, and J. P. Neijt, "Inference in the promedas medical expert system," in *Artificial intelligence in medicine*, 2007, pp. 456–460.
- [21] P. Klinov, "Pronto: A non-monotonic probabilistic description logic reasoner," in *The Semantic Web: Research and Applications*, 2008, vol. 5021, pp. 822–826.
- [22] R. Shearer, B. Motik, and I. Horrocks, "HermiT: A highly-efficient OWL reasoner," in *Fifth Int. Workshop on OWL: Experiences and Directions (OWLED 2008 EU)*, 2008.
- [23] Y. Anokwa, N. Ribeka, T. Parikh, G. Borriello, and M. C. Were, "Design of a phone-based clinical decision support system for resource-limited settings," in *Fifth Int. Conf. on Information and Communication Technologies and Development*. ACM, 2012, pp. 13–24.
- [24] B. DeRenzi, C. Sims, J. Jackson, G. Borriello, and N. Lesh, "A framework for case-based community health information systems," in *IEEE Global Humanitarian Technology Conference (GHTC 2011)*. IEEE, 2011, pp. 377–382.
- [25] S.-H. Liao, "Expert system methodologies and applications—a decade review from 1995 to 2004," *Expert Systems with Applications*, vol. 28, no. 1, pp. 93–103, 2005.
- [26] W. Yao, J. Rolia, S. Basu, S. Singhal, and A. Kumar, "A context-aware framework for patient navigation and engagement (CANE)," in *8th Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2012)*, 2012, pp. 316–325.