Global Telehealth 2012
A.C. Smith et al. (Eds.)

doi:10.3233/978-1-61499-152-6-10

Personal Health Records: Retrieving Contextual Information with Google Custom Search

Mahmud AHSAN, H. Lee SELDON[1] and Shohel SAYEED
Faculty of Information Science & Technology
Multimedia University, 75450 Melaka, Malaysia

Abstract. Ubiquitous personal health records, which can accompany a person everywhere, are a necessary requirement for ubiquitous healthcare. Contextual information related to health events is important for the diagnosis and treatment of disease and for the maintenance of good health, yet it is seldom recorded in a health record. We describe a dual cellphone-and-Web-based personal health record system which can include 'external' contextual information. Much contextual information is available on the Internet and we can use ontologies to help identify relevant sites and information. But a search engine is required to retrieve information from the Web and developing a customized search engine is beyond our scope, so we can use Google Custom Search API Web service to get contextual data. In this paper we describe a framework which combines a health-and-environment 'knowledge base' or ontology with the Google Custom Search API to retrieve relevant contextual information related to entries in a ubiquitous personal health record.

Keywords. Personal health records, health-related information, web service, web mining

Introduction

'Telehealth' or any kind of healthcare depends strongly on a health record in some form, whether it be paper, electronic or even oral. The value of 'integrated' health records is becoming ever more apparent, although such are still rare in most of the world. 'Availability' is another key concept linked to health records.

One way to integrate health-related information is the 'Personal Health Record' or PHR; this is a record owned and maintained by individuals, with terminology which they understand. To solve the 'availability' problem, PHRs may be Web-based; in regions where the Web is not available they may be tied to the person as a portable cellphone application. We are developing a dual cellphone-and-Web-based PHR for use in the more remote parts of Malaysia such as Borneo.

[1] Corresponding Author: Dr Lee Seldon, Faculty of Information Science & Technology Multimedia University, 75450 Melaka, Malaysia; E-mail: lee.seldon@mmu.edu.my

The information in health records is critical for people to achieve or maintain their health. It may derive from the individual himself or herself, from healthcare providers, from devices which measure health parameters and from the rest of the individual's world. It may be physical or mental signs or symptoms, diagnoses, treatments, etc. Importantly, however, a PHR may also contain health-related information not usually recorded in a medical record, such as information about the person's physical and social environments over time. This might be important not only to the individual, but also to healthcare providers.

Here we discuss the information from 'the rest of the world'. Many external factors influence our health, so knowledge of the context of a health problem is important—during an encounter a doctor often asks a patient questions like 'when did this start?', 'where were you?', 'does anybody else have something like this?', etc. The person can report to the healthcare provider about his problem, but he may not remember or properly describe the surrounding situation. Much contextual information is almost never recorded in health records—items like pollution levels, local floods, government health warnings (e.g. epidemics), work or study environments, etc.

In the past, tracking the environmental issues was tough due to scarce information, low awareness in the general public and insufficient technological support. But now we can track and get contextual information from various sources. So our target is to get relevant contextual information for a person's health events (e.g. symptoms) which are recorded in a PHR and to save the context in his PHR. We target contextual information which may be available on the Internet, i.e. which is publicly available. There are several problems to overcome when trying to integrate such information into a PHR system and our objectives are to solve these.

1. Objectives

We want to solve these problems:

- How do we identify reliable sources of relevant health-related information?
- How do we identify information which is relevant to a particular person with a particular problem (at a particular time in a particular location)?
- How do we retrieve such relevant information?
- Where and how do we store the relevant information so that it is accessible to the concerned individual?

2. Related Work

2.1. Personal Health Records (PHR)

PHRs are health records, which are owned and maintained by individuals, in contrast to 'medical records' owned and maintained by healthcare providers. IndivoHealth [1] and Microsoft HealthVault [2] are examples of Web-based PHRs. PHRs store health-related information which is entered manually or uploaded by the individual owner or by healthcare providers. This of course includes physiological measurements, including blood glucose levels, vital signs, signs and symptoms of problems, family and social history, vaccination history, etc. There are also 'smartphone' applications which allow users to track their health; these are usually targeted at people with diabetes or other specific diseases, as people with chronic conditions would likely be the most motivated to maintain an ongoing PHR.

The particular PHR system with which we work is a dual phone-and-Web-based PHR described by Seldon [3] and Seldon et al., [4]. The mobile phone component uses the standard International Classification for Primary Care vocabulary version 2e [5]; this covers most of the parameters mentioned above. For transmission to a Web-based PHR database, the name-value pairs of the mobile PHR are converted to a standard CCR (Continuity of Care Record [6]) message.

Generally, neither PHRs nor other health records store contextual information like that mentioned above. Here we describe an approach to filling this gap.

2.2. Contextual Information for Health

Identifying relevant health-related information: It is by now well known that human health depends on various factors including the environment, lifestyle, behaviour, biological inheritance, hygiene and so on [7]. For example, it is well known that air pollution can seriously and directly affect human health. The World Health Organization (WHO) estimates that 2.7 million people die of air pollution around the world every year [8]. Similarly water pollution may be the cause of a person's sickness; the association between diseases and bacteria types has been well established [9].

In the UK, climate change might increase COPD, asthma, tick-related diseases (e.g. Lyme disease) and some food-borne diseases during the warmer summers [10]. Furthermore the highly significant association between temperature and death rates from various causes like chronic bronchitis, pneumonia, ischemic heart disease and cerebrovascular disease indicates a change in future disease prevalence [10]. Temperature increase has direct impacts in the form of heat-related mortality and afflictions like dehydration, heatstroke, and hyperthermia [10].

Many governments post health warnings on the Web. Sites such as HealthMap (http://healthmap.org/en/) also track reports of possible outbreaks of disease. A Malaysian 'environmental health' site [11] tracks reports of environmental problems in Malaysia.

2.3. Ontology-based Modelling Approach

The problem is to relate the parameters in a PHR to the information available at relevant Web sites. An ontology or taxonomy defines classes of things (information), relations among them, functions, constraints and so on [11]. We use a simple ontology of classes of health-related information and their relationships.

The Web site, myehms.com [11] follows a similar process to retrieve contextual data. The site is based on a manually derived knowledge base of health symptoms and potentially related environmental factors. The site retrieves data from mostly Malaysian News Feed Sources (RSS), which were manually discovered and screened based on several criteria. The site uses the ontology to identify keywords and phrases from RSS feeds and then uses pattern matching to classify the contextual information [13]. Their ontology is a good base, also for a more general Web search. Currently there is no obvious way to automate maintenance of the knowledge base, so it remains manual.

2.4. Web Search Engines

Retrieving (only) relevant information from the Web presents another problem [14]. Even the most experienced searching technique finds it difficult to identify and retrieve relevant information from the worldwide web (WWW) [15]. Contextual information retrieval is not a new idea, but it has distinct challenges when compared to general information retrieval [16]. The Google Custom Search API [17] allows websites and programs to retrieve and display search results from Google Custom Search programmatically.

The remainder of this paper discusses a framework for retrieving and finding contextual, health-related information from the WWW using ontology and the Google Custom Search API.

3. Framework

Our proposed framework has five agents: Context Processor Agent, Cache Agent, Google Search Agent, Classification Agent and Notification Agent. Figure 1 represents our framework at a glance.

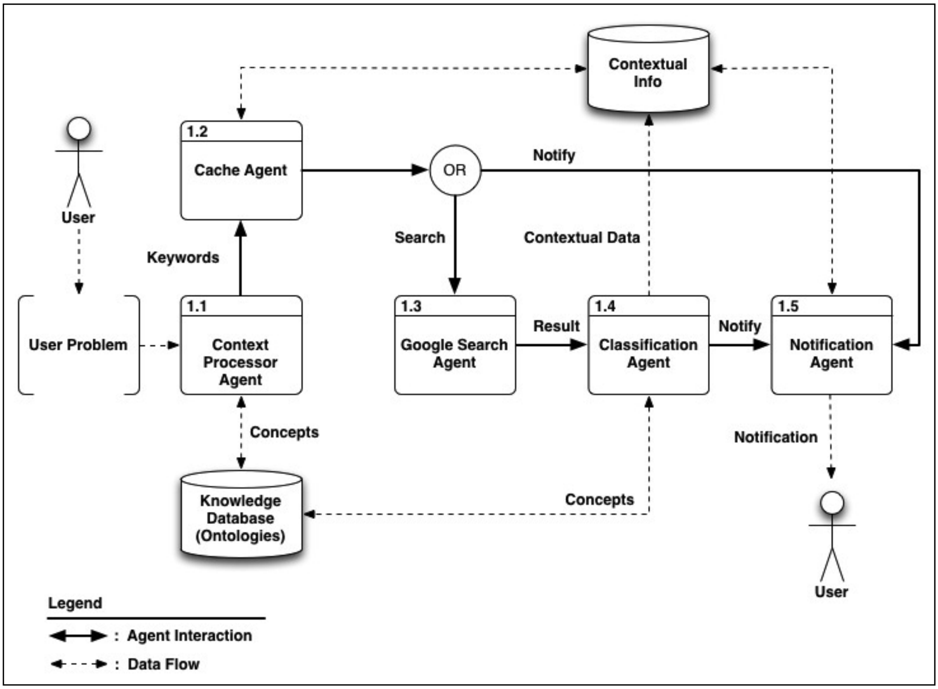

Figure 1: Proposed Framework Architecture

3.1. Context Processor Agent

A PHR user enters a problem or symptom into his cellphone PHR, then uploads the entry to his account on a Web-based PHR. For example, he might enter 'Problem> Arm> left> spots = itching;' to indicate itchy spots on his left arm.

The Context Processor Agent takes an input symptom, e.g. 'Problem> Arm> left> spots = itching;', and searches it for relevant keywords based on a Knowledge Database or ontology like the one used by myehms.com [11]. For example, if the input symptom contains 'itching' or 'itchy' and if the predefined location is set as Melaka, Malaysia, our Context Processor Agent creates a query as 'air pollution Melaka Malaysia' because itching may be a symptom of allergy, and one environmental cause for this problem could be air pollution; other queries could also be generated. Table 1 represents a sample of a Knowledge Database. The table shows not only English terms for symptoms and signs, but also corresponding Chinese and Malay language terms.

Table 1. Sample of Knowledge Database adapted from [11]

PHR Symptoms & Signs, English	Symptoms & Signs, Chinese	Symptoms & Signs, Malay	Environmental hazards, etc.	Environmental Dimension
itching, sneezing, drowsiness, nasal stuffiness, watery eyes, rash, etc.	打喷嚏-鼻腔堵塞, 呼吸困难, 鼻窦症状, 痒, 疣, 局部皮疹, 全身皮疹	bersin-hidung tersumbat, sukar bernafas, gejala sinus, kegatalan, kutil, ruam setempat, ruam menyeluruh	allergens	air pollution
nausea, vomiting, diarrhoea, abdominal pain, decreased production of red and white blood cells	恶心, 呕吐, 腹泻, 腹痛,	loya, muntah, cirit-birit, sakit perut	arsenic	water

3.2. Cache Agent

The query from the Context Processor goes to the Cache Agent, which checks to see if relevant information has already been stored in the local database ('Contextual Info' table—c.f. Figure 1, Table 3). This could happen if several people in one place report the same symptoms in a short period of time. If such information is already present, the Notification Agent will link that data to the current user and notify the user. But if the Cache Agent does not find the data locally, it will forward the query to the Google Search Agent.

3.3. Google Search Agent

The Google Search Agent uses the keywords sent by the Cache Agent to find results in the billons of web pages indexed by Google. Google's own ranking paradigm for relevance is reliable and our Google Search Agent will retrieve only the top 20 results. Figure 2 represents a pseudo- code procedure to show how Google Search Agent works.

```
GOOGLE_SEARCH_AGENT Procedure
BEGIN
    GET keyword_list
    WHILE keyword_list NOT EMPTY
    BEGIN
        GET each keword from keyword_list
        DO result = GOOGLE_SEARCH_RESULT(keyword)
        SAVE result in result_list
    END WHILE
  END
  CALL CLASSIFICATION_AGENT(result_list)

** GOOGLE_SEARCH_RESULT() is a function that use
Google Custom Search API to get search result.
```

Figure 2: Google Search Agent Procedure

Using this we can retrieve the title, summary, source and access date of the web pages by Google and can forward this report to the Classification Agent.

For example in Figure 3 we show that when we call our Google Search Agent with the query 'air pollution Melaka Malaysia' we get some relevant results. In this list items 5 and 6 look like potential contextual information, but that will finally be determined by our Classification Agent.

Rank : 4
Title : **Air Quality**
Source : http://www.doe.gov.my/files/u1/AIR%20QUALITY.pdf
Summary : **Malaysia** (Negeri Sembilan, **Melaka** and Johor) the **air quality** was also between good to moderate most of the time, with the exception of a few unhealthy days **...**

Rank : 5
Title : **Pollution** in **Malacca, Malaysia**
Source : http://www.numbeo.com/pollution/city_result.jsp?country=Malaysia&city=Malacca
Summary : Pollution in **Malacca, Malaysia**. Surveys by city **air pollution**, water pollution, greens and parks satisfaction, light and noise pollution, ...

Rank : 6
Title : Weather Forecast: **Melaka, Malaysia** - The Weather Network
Source : http://www.theweathernetwork.com/weather/my040000
Summary : Get the most updated weather conditions and forecasts for **Melaka, Malaysia**. **...** Refers to how the outdoor **air** is expected to feel in degrees Celsius when actual **...**

Figure 3: Google Search Result Sample

3.4. Classification Agent

Based on the Google Search result and the Knowledge Database, the Classification Agent can organize the information into intuitive classifications and can store this relevant contextual information in a table in the Web-based PHR. Figure 4 shows the parts of Classification Agent—the Parser and Decision Maker. After getting the Google result list, the Parser parses the result based on the Knowledge Database (i.e. the query which was submitted), makes a standard format like Table 2 and saves it temporarily. Then the Decision Maker analyses it and decides which data are relevant according to the query, and if it finds relevant data it will save the data in the Contextual Info table.

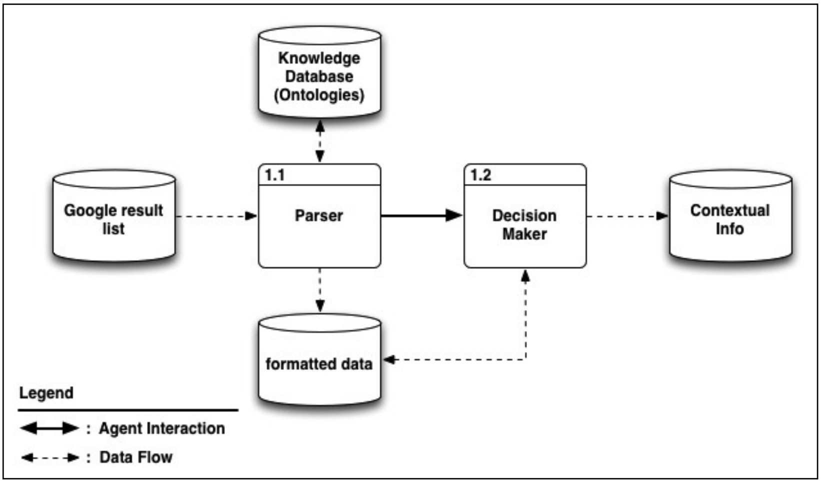

Figure 4: Classification Agent Module

Table 2: Formatted Data

Date	Google Rank	Source	Google Summary	Word 1 *air*	Word 2 *pollution*	Word N *Melaka*
2012/07/20	0.6	http://url/1	Summary	1	2	1
2012/04/08	0.8	http://url/2	Summary	1	1	0

Table 2 represents the data format used by the Parser. Here *Date* is the published date of data or Google Cache date, *Google Rank* represents the position of the data source in Google result, *Summary* is the summary of the WWW page by Google, and *'Words'* are the words in the query. So if the search query is 'air pollution Melaka' then the table will use three words and count the number of times the words appear in the Google summary. Finally, the Decision Maker will create a data index and sort the data based on its criteria. For example: if the number of occurrences of the words in the first row is four and in the second row is two, then the first row appears more relevant than the second row. Of course other criteria like publication date, Google Rank, etc. also matter to the Decision Maker. As we are using Google Search API and Google Search Engine, we do not have to worry about keywords pattern, spam or rank in the data. Google does that internally and gives us the final result. Finally, the accepted contextual data will be saved in 'Contextual Info' table as indicated in Table 3. Then the Notification Agent will link the saved contextual data with the original individual's account ID, date and time, etc.

Table 3: Contextual Information

INFO ID	**DATE** *crawl date*	**KEYWORDS** *search query*	**SUMMARY**	**URL**	**RANK**	**LOCATION** *Post code*
001	2012/11/01	air pollution melaka malaysia	Summary	http://url/1	2	75450
002	2012/12/08	water pollution kuala lumpur	Summary	http://url/2	3	43000

3.5. Notification Agent

Notification Agent creates an entry in 'User Contextual' table (Table 4) to make the relation between user and contextual information. This is a classical 'many-to-many' relation—multiple individuals with similar problems in a similar location would share the corresponding contextual information; also, one individual may find multiple relevant contextual entries for one symptom.

Table 4: User and Contextual Information Relationship

USER ID	**DATE**	**CONTEXTUAL_INFO_ID**	**User Problem**
0123	2012/11/01	001 (see Table 3)	Problem> Arm> left> spots = itching

Finally, as mentioned in the Introduction, the individual who has the health problem in the first place, e.g. 'Problem> Arm> left> spots = itching;', must be notified of the outcome of the contextual information search. This person should be informed if there is a current air pollution warning in his region. An ancillary problem to solve is whether all of the results should be sent to the person concerned, or only a pointer (URL, database index, etc.). That information should then be stored in the person's cellphone PHR.

4. Conclusion

In order to achieve its goals, Global Telehealth must also serve people in more remote parts of the world. In the absence of healthcare facilities, a Personal Health Record (PHR) assumes more importance. In the absence of broadband Internet, cellphones assume more importance. So a phone-and-Web-based PHR is a practical suggestion for recording the health of individuals. Health records, whether medical or personal, seldom contain contextual information which may be relevant to health and specific health problems. This paper presents a research framework to identify and retrieve such contextual, health-related information from the Web based on a 'symptom – environment' cross-reference knowledge base and Google Custom Search. The project modifies an existing knowledge base and integrates it with Google Custom Search. We describe how, when an individual records a health problem in his/her ubiquitous PHR, the process can use the problem, together with the date and the individual's location, to search the Web for potentially related information. Such information may benefit the individual as well as healthcare providers.

References

[1] Children's Hospital Informatics Program. Indivo. 2007. http://indivohealth.org/research (accessed April 10, 2012).

[2] Microsoft. Microsoft HealthVault. 2012. http://www.microsoft.com/en-us/healthvault/ (accessed April 15, 2012).

[3] H.L. Seldon, Personal Health Records in Southeast Asia Part 1 – a Way to Computerize Healthcare? electronic Journal of Health Informatics [submitted July 2012]

[4] H.L. Seldon, H. Moghaddasi, W.J. Seo, S. JoNah Wee, Personal Health Records in Southeast Asia Part 2 - A Digital Portable Health Record. electronic Journal of Health Informatics [submitted July 2012]

[5] WICC (WONCA International Classification Committee). ICPC-2 International Classification for Primary Care. http://www.globalfamilydoctor.com/wicc/sensi.html (accessed July 1, 2011).

[6] ASTM. Standard Specification for Continuity of Care Record (CCR). E2369. 2005

[7] Lalonde MA, "New perspective on the health of Canadians," Ottawa: Health and Welfare, 1974.

[8] WHO, "WHO Guidelines for air quality," Fact sheet No. 187. 2003.

[9] P.K. GOAL, WATER POLLUTION Causes, Effects and Control. Revised 2nd Edition. New Delhi: New Age International ® Limited, 2006.

[10] M.S. Islam, T. Chaussalet, N. Ozka, E. Demir, "An Approach to Exploring the Effect of Weather Variations on Chronic Disease Incidence Rate and Potential Changes in Future Health Systems ." Computer-Based Medical Systems (CBMS), IEEE 23rd International Symposium . IEEE Conference Publications, 2010. 190-6.

[11] EHMS.com. Environmental Health Management System. 2012. http://www.myehms.com/ (accessed May 10, 2012).

[12] Urvi Shah, T. Fini, A. Joshi, R.S. Cost, J. Matfield, Information retrieval on the semantic web. New York: ACM, CIKM '02 Proceedings of the eleventh international conference on Information and knowledge management, 2002.

[13] T.C. Lau, "Analysis, Design and Implementation of a GIS-based environmental health management system." Master's Thesis, Swinburne University of Technology Sarawak Campus, Malaysia, 2012.

[14] W. Fan, M.D. Gordon, and P. Pathak, Discovery of Context-Specific Ranking Functions for Effective Information Retrieval Using Genetic Programming. Knowledge and Data Engineering, IEEE Transactions, 2004, 16(4): 523 - 527.

[15] N. O'Hanlon, Off the shelf & onto the Web: Web search engines evolve to meet challenges. Reference & User Services Quarterly, 1999, 38(3): 247.

[16] D.K. Limb, A.M. Connor, S.G. MacDonell, A framework for contextual information retrieval from the WWW, in Proceedings of the ISCA 14th International Conference on Intelligent and Adaptive Systems and Software Engineering (IASSE). Toronto, Canada, ISCA, 2005, 185- 189.

[17] Google Inc. Google Search API. 2012. https://developers.google.com/custom-search/docs/api (accessed May 30, 2012).