Personalized Web Information Retrieval

Based on

Varying Health Parameters Related to Diabetes

A. De S. Jayatilaka
General Sir John Kotelawala Defence University
e-mail: asangi234@gmail.com

S.M.A.K.D. Arunatileka
University of Colombo School of Computing Sri Lanka
e-mail: sda@ucsc.cmb.ac.lk

R.G. Premaratne
Ministry of Health, Sri Lanka
e-mail: risintha@yahoo.com

Abstract— **This paper describes an ongoing study that examines the problem of extracting personalized health information from World Wide Web. Internet has presented a much more promising channel for health information dissemination, however, there are significant challenges. Even though people frequently use Internet to find medical information, the relevancy of the information retrieved is questionable. These retrieved information is not tailored to specific individuals and are not retrieved based on individual characteristics. Personalization or tailoring of health-related content according to individual characteristics is a highly desirable goal for information retrieval research in the field of medical informatics. The strategies proposed so far, depends only on medical records and diagnosis cards. This research proposes a methodology to personalized access to diabetes related web documents by analyzing the individual health parameters, which are fluctuating over time.**

Keywords- Information Retrieval; Medical Informatics; Medical Information; World Wide Web; Personalization

I. INTRODUCTION

The introduction of the World Wide Web has given rise to enormous amount of information to be accessed in the digital form [1]. However, exponential growth of these digital documents have given rise to many new challenges, specially in terms of information retrieval. WWW is over loaded with information. Therefore, web search results suffer from accuracy and precision.

Internet has presented a much more promising channel for health information dissemination, and has been widely used as such by the general public. People frequently use the Internet to find medical information related to their health condition. However, concerns are frequently expressed regarding the relevancy of those information obtained [2]

Generally, people search in the Internet without a clear road map of what he or she is searching [3]. For example, if

someone searches for a general topic such as 'Diabetic management' in Google, over 27 million web documents are fetched. Even though, there is a lot of information in the web, the number of web documents have become unmanageable and these documents are scattered all over the web.

Personalization of health-related content according to individual characteristics, is a highly desirable goal for information retrieval research in the field of medical informatics [4,5].

This research uses diabetes as the case study. It is expected that the proposed methodology could be adopted to other diseases as well.

Diabetes is a chronic disease. Chronic diseases are diseases of long duration having generally slow progression and by far the leading cause of mortality in the world, representing 63% of all deaths [6]. Diabetes is responsible for deaths of three in five people world wide and it is predicted to be the 7th leading cause of death by year 2030 [7].

Nowadays, due to the busy life styles, people don't generally visit the doctor for regular checkups [8]. Therefore, they are not aware about their health conditions; *i.e.* whether their health conditions are deteriorating or improving over time. Most of them visit the doctor only after getting symptoms of diabetes. Therefore, diagnosis information is not always available and it is paramount important to analyze the individual health parameters over time and to use those analysis information to fetch the most relevant authentic web resources. This will help to increase the knowledge of the people regarding their present health condition, possible impacts, preventive measures thereby diverging from the risks of getting diabetes.

Personalized Information retrieval methodology which is discussed in the paper, in addition to searching and filtering information aim at delivering content to individual users. This means that two users searching information using a specific query may get different results, tailored to their

needs. For example, a person who has a very low probability of getting diabetes need to look in to prevention and not management of diabetes. Therefore, personalization of web information related to health conditions allows people to take take well informed decisions.

The paper is organized as follows. *Section 2* describes the background of the research. *Section 3* discusses the goals and objectives and *section 4* illustrates the proposed solution in detail. *Section 5* discusses the conclusion and future work.

II. BACKGROUND

Chronic diseases are diseases of long duration having generally slow progression and by far the leading cause of mortality in the world, representing 63% of all deaths (WHO 2011a). More than 80% of diabetes deaths occur in low and middle income countries [6]. Diabetes is one of the four major chronic diseases, which kill three in five people world wide and it is predicted to be the 7th leading cause of death by year 2030 [9]. It is also predicted that diabetes deaths in the world will double during 2005-2030 [7]. In Sri Lanka there has been a progressive increase in diabetic deaths in the past two decades [10].

There are two major types of diabetes. Type 1 diabetes is characterized by a lack of insulin production and type 2 diabetes results from the body's ineffective use of insulin. Type 1 diabetes is assumed to be less preventable, however, type 2 diabetes is greatly preventable. Type 2 diabetes is much more common than type 1 diabetes. Type 2 diabetes accounts for around 90% of all diabetes worldwide [7].

Diabetes and other chronic diseases, not only can end up in death, but also are major risk factors for other illnesses and result in massive socioeconomic burden [6,11].

Diabetic risk factors such as obesity, lack of exercise, unhealthy eating patterns mainly depend on the individual behavior, which can be addressed by health promotive measures.

III. GOALS & OBJECTIVES

The main aim is to provide personalized access to diabetes related web documents by analyzing the individual health parameters, which are fluctuating over time. The objectives of the research are listed down below.

A. Analyzing the information provided by people regarding their health parameters and create an initial health profile for each individual

Set of structured questions based on the risk factors of diabetes are used to get an understanding of the possible risks of the individuals towards diabetes. The questions is formed on the basis of factors such as personal profile, present & past health, family health history, lifestyle & personal habits etc... The identified health parameters are compared against the medical standards. Based on this analysis an initial health profile is constructed.

B. Update health profile over time

Some important health parameters (eg: weight, exercise level) needs to be monitored on a regular basis. This enables to derive conclusions on trends of each health parameter.

For example, whether there is a possibility for an unexplainable weight loss can be identified by analyzing factors such as the weight, level of exercise and dietary habits over time. Furthermore, there can also be life style changes which could adversely affect the individual health such as addiction to tobacco. Therefore, individual health profile is updated as new knowledge about the individual is learned.

C. Retrieve the relevant web documents based on the individual health profile

Health profile constructs the basis for "what to be searched for". Web documents that are matched with the above constructed criteria are only retrieved. The search becomes more specific and may span to many paths with the knowledge growth about the individual over time.

IV. RELATED WORK

This research spans in to two domains, constructing a patient profile using the individual health parameters which are fluctuating over time and using the constructed profile to fetch the most relevant web documents. Therefore, literature survey was carried out under the following categories.
1. Developing an individual profile based on health parameters
2. Personalized web information retrieval based on medical profile

A. Developing an individual profile based on health parameters

There are references to web-based interventions that have been successfully implemented for assessing the risk of getting diabetes [12,13] When assessing the risk of getting diabetes, individual profile is created based on the user given medical information. Based on this individual profile, risk of getting diabetes is calculated and presented to the user.

Web based interventions have also been successfully implemented in self managing of diabetes, weight management and dietary control [13,14]. The study [15], which focuses on dietary modification of adults with diabetes (type 2), is capable of generating a personalized dietary intervention program for a large group of patients.

All the research uses medical records or personal health information to derive a personal health profile for each individual.

B. Personalized web information retrieval based on medical profile

There have been several approaches aiming at linking on-line information resources with electronic patient records, but their primary focus was to support scientific or health professional information needs.

Studies which focus on existing patient health information sources and investigate approaches for linking data from a patient's medical record(s) with relevant health information on the web can be seen as the mostly related research to the proposed study [16,17].

[16] study examines problems with existing patient health information sources and investigates an approach for intergraing data from a patient's medical record(s) with relevant health information on the web. The main aim of [16] is to provide patients with simplified, customized and controlled access to web information. Data from patient medical records are extracted and linked with relevant health information on the web through a web search service.

The focus main focus is on how web information can be customized *after a diagnosis,* so that the patient will find it easy to understand the medical terms.

STEPPS (STructured Evaluated Personalized Patient Support) proposes a strategy for integration of electronic patient records with Internet health-related content and its consequent use in personalized information retrieval for patient education [17]. This study provides post-discharge support for burn patients in the Netherlands.

In summary, all these studies are based on static patient profiles and they only focus on the diagnosis information on the Electronic health Records (EHR). In contrast, our study mainly focuses on people who are still not diagnosed with diabetes and tries to assist them in finding relevant knowledge from web documents based on their health profiles.

V. METHODOLOGY

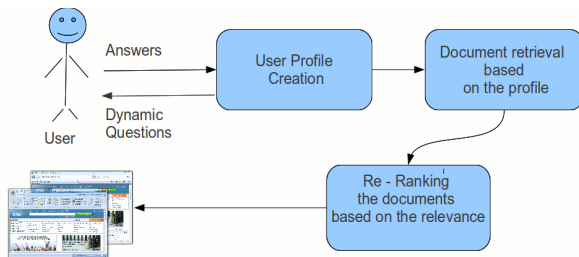Overview of the proposed methodology is depicted in the *figure 1.*

Figure 1: The methodology for personalized web document retrieval based on health parameters

A. User Profile Creation

The user profile is created based on the responses obtained for a set of dynamic questions. These question request information on the following facts.

- Test report results
- Screening tests
- Checking for symptoms (e.g. urine passing frequency)
- Checking for complications (e.g. numbness in the hand and the feet)

As illustrated in *figure 1*, if user has already taken test reports, then analysis is done on the test results. If there are no test results then a set of questions (which are used as a screening test) is used. This screening test will determine whether the person needs to move forward in the questionnaire or not.

Finally symptoms and the complications are checked for the people who have been identified as having an average or high risk in getting diabetes according to the screening test.

Due to the complex nature of diabetes, there can be people who demonstrate diabetes complications before demonstrating the symptoms. Therefore, both complications and symptoms are checked for each person who is having an average or high risk in getting diabetes according to the screening test.

Based on the medical parameters the people are clustered in to three main categories; low risk of getting diabetes, average risk of getting diabetes and high risk of getting diabetes. A simplified view of this categorization is shown in the *figure 2*. Each user profile is mapped to in to the given scale. Furthermore, each user profile, holds information about the unfavorable health parameters. This information is vital in latter stages when retrieving personalized web documents and ranking the retrieved documents based on the relevance.
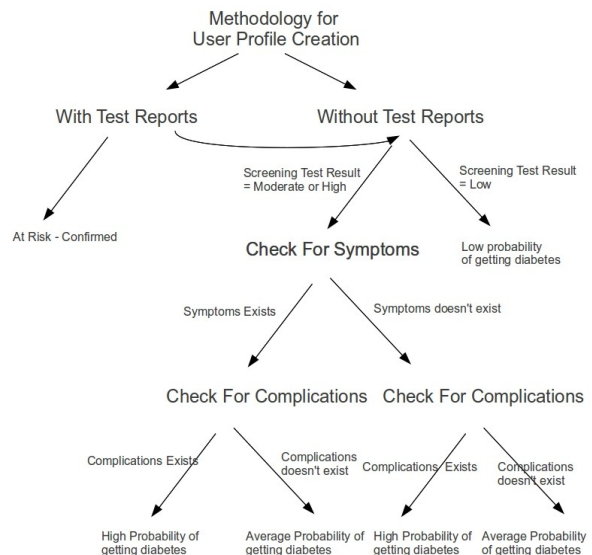

Figure 2: User Profile Creation

B. Document retrieval based on the medical profile

The user profile creates the basis for what needs to be search for. The important documents in the Internet are clustered in to three main categories based on their content.

1. Health promotion and Prevention
2. Documents related to controlling of diabetes (eg. Dietary Control, Exercising schedules etc...)
3. Treatment

The documents are retrieved based on two aspects which are;

1. The risk category
2. The unfavorable risk parameters

The important points are extracted from each document and plotted in a multidimensional space. Each point correspond to,

- Health Issue
- Description about the health issue
- document category
- Relevancy for the category

- Other parameters as necessary

Using the user profile, point "p" on this space is derived and based on the certainty of the health condition radius "r" is identified. We draw a hypothetical sphere on this space where center is p with the radius r. All the documents with in the sphere can be considered relevant.

C. Re- ranking of the web documents

In document retrieval, all documents which satisfy $PD_{i<}$ r are fetched,where PD_i is the distance between user profile 'p' and document D_i. However, in presenting the retrieved web resources to the user, it is important rank them according to the relevancy. For example, if it is identified that the trend of urine passing frequency is increasing significantly for a certain user, then documents which discuses that issue needs to be presented to the user with high priority. This can be easily done by sorting the documents based on the distance with the user profile 'p'. The document which has the lowest distance with the point p in the multi dimensional vector space is the most relevant document to the user, according to his/her personal characteristics. Therefore, such documents are displayed by giving high priority.

VI. CONCLUSION AND FUTURE WORK

The Internet has become one of the most useful mediums to facilitate health promotion. [18] defines health promotion as:

> "Health promotion is the process of enabling people to increase control over, and to improve, their health."

In Sri Lanka, one third of people with diabetes, are undiagnosed [19]. They may carry the symptoms for years and diagnosed of diabetes only after getting complications of diabetes. It is identified that prevention needs to take place simultaneously at the population level and at the individual level.

In order to prevent chronic diseases such as diabetes, people need to continuously monitor themselves, should be well informed about their health conditions and should be aware of different preventive actions based on their health conditions. Empowering the potential patients, which places the responsibility of maintaining health and wellness on the hands of the people is a key for preventive medicine. Empowered people make more informed decisions and are better prepared to negotiate the demands of their environment, bringing a positive change to their lives and personal behavior.

In order for empowerment of the potential patients:

I. People need to be fully aware of how their health parameters are functioning over time and

II. They must have good knowledge regarding the diseases and preventive actions, based on their health parameters

However, both these conditions are not satisfied due to the following reasons. Nowadays, due to the busy life styles, people don't generally visit the doctor for regular checkups [20] Therefore, they are not aware of their health conditions; whether their health conditions are deteriorating or improving over time. Most of the people visit the doctor only after getting symptoms of a particular chronic disease.

People frequently use Internet to find medical information. Studies have shown that more than 80% of the adult population in the United States use the Internet to find medical information [21], but concerns are frequently expressed regarding the appropriateness, safety and accuracy of the information obtained. It has become hard for people to get hold of authentic web resources related to their health conditions because health information is scattered all over the web. Therefore, relevancy can be a barrier to on-line health communication interventions that lead to meaningful changes in behavior.

This on going study proposes a solution to the above stated problem. The methodology proposed in the paper, can retrieve documents from web based on individual medical parameters. Personalized web document search based on medical parameters will no doubt pave the way for patient empowerment.

It is expected to develop a prototype based on the proposed methodology in the future and comprehensive quantitative analysis will be carried out based on precision and recall measures to validate the proposed methodology.

REFERENCES

[1] T. Berners-Lee. Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor. London, Texere., 2000.

[2] Kavanagh, J. 2010. Caring for people with chronic conditions: How technology can support an evidence-based model in order to improve chronic care. Available at:
http://kinwahlin.wordpress.com/2010/05/18/caring-for-people-with-chronic-conditions-howtechnology-can-support-an-evidence-based-model-in-order-to-help-improve-chronic-caremicrosoft-white-paper/ [Accessed October 1, 2011].

[3] Feldman, S. & Sherman, C. 2004. The high cost of not finding information. Information Today, Inc. Available at :
http://ejitime.com/materials/IDC on The High Cost Of Not Finding information.pdf [Accessed March 12 , 2012]

[4]HUMPHREYS, B. L., 2000, Electronic health record meets digital library: a new environment for achieving an old goal. Journal of the American Medical Informatics Association, 7, 444–452.

[5] CIMINO, J. J., 1996, Linking patient information systems to bibliographic resources. Methods of Information in Medicine, 38, 122–126.

[6] WHO 2011a. WHO | Chronic diseases. *WHO | Chronic Diseases*. Available at:
http://www.who.int/topics/chronic_diseases/en/ [Accessed November 15, 2011].

[7] WHO 2011b. WHO | Diabetes. Available at:
http://www.who.int/mediacentre/factsheets/fs312/en/index.html [Accessed November 15, 2011].

[8] Doherty, E. & Shevlin, M. 2005. A Picture of Men's Health; Men's Health: an examination of Attitudes and Behaviours. Available at: http://www.westernifh.org/uploads/Reports/A Picture of Mens Health - Derry Healthy Cities Report pdf_20080129113316.pdf [Accessed November 15, 2011]

[9]WHO 2011c. WHO | Noncommunicable Diseases and Mental Health. Available at: http://www.who.int/nmh/en/ [Accessed November 15, 2011].

[10] Premaratne, R., Amarasinghe, A. & Wickkramasinghe, A.R. 2005. Hospitalisation trends due to selected non-communicable diseases in Sri Lanka, 2005-2010. *Ceylon Medical Journal*, **50**(2), p.51–4.

[11]Oxford Health Alliance 2006. Chronic Disease: an economic perspective. Available at: www.oxha.org [Accessed October 1, 2011].

[12] University of Mrylank Medical System 2009. UMMS: Diabetes Type II Risk Calculator. Available at: http://www.healthcalculators.org/calculators/diabetes.asp [Accessed November 16, 2011].

[13] American Diabetes Association Diabetes Risk Test - American Diabetes Association. Available at: http://www.diabetes.org/diabetes-basics/prevention/diabetes-risk-test/ [Accessed November 16, 2011].

[14] Glasgow, R.E., Boles, S.M., McKay, H.G., Feil, E.G. & Barrera, M., Jr 2003. The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results. *Preventive Medicine*, **36**(4), pp.410–419. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12649049 [Accessed November 27, 2011].

[15] McCoy, M.R., Couch, D., Duncan, N.D. & Lynch, G.S. 2005. Evaluating an internet weight loss program for diabetes prevention. *Health Promotion International*, **20**(3), pp.221–228. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15797900 [Accessed November 27, 2011].

[15] Ramadas, A., Quek, K.F., Chan, C.K., Oldenburg, B. & Hussein, Z. 2011. Randomised-controlled trial of a web-based dietary intervention for patients with type 2 diabetes mellitus: Study protocol of myDIDeA. *BMC Public Health.* Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3123593/.

[16] Al-Busaidi, A., Gray, A. & Fiddian, N. 2006. Personalizing Web Information for Patients: Linking Patient Medical Data with the Web Via a Patient Personal Knowledge Base. *Health Informatics Journal*, **12**(1), pp.27–39. Available at: http://jhi.sagepub.com/content/12/1/27 [Accessed March 11, 2012].

[17] Doupi, P. & van der Lei, J. 2002. Towards personalized Internet health information: the STEPPS architecture. Medical Informatics and the Internet in Medicine, 27(3), pp.139–151. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12507260 [Accessed May 22, 2012].

[18] Ottawa Charter for Health Promotion 1986. WHO/HPR/HEP/95.1. WHO in Geneva. Available at: http://www.who.int/hpr/NPH/docs/ottawa_charter_hp.pdf [Accessed February 27, 2012].

[19] Katulanda, P., Constantine, G.R., Mahesh, J.G., Sheriff, R., Seneviratne, R.D.A., Wijeratne, S., Wijesuriya, M., McCarthy, M.I., Adler, A.I. & Matthews, D.R. 2008. Prevalence and projections of diabetes and pre-diabetes in adults in Sri Lanka--Sri Lanka Diabetes, Cardiovascular Study (SLDCS). *Diabetic Medicine: A Journal of the British Diabetic Association*, **25**(9), pp.1062–1069. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19183311 [Accessed January 1, 2012].

[20]Doherty, E. & Shevlin, M. 2005. A Picture of Men's Health; Men's Health: an examination of Attitudes and Behaviours. Available at: http://www.westernifh.org/uploads/Reports/A Picture of Mens Health - Derry Healthy Cities Report pdf_20080129113316.pdf

[21] Pew Internet & American Life Project. (2006). Online health search 2006. Retrieved 12 2009, from http://www.pewinternet.org/~/media/Files/Reports/2006/PIP _Online_Health_2006.pdf