In *International ICSC Congress on Intelligent Systems and Applications (ISA'2000), Special Session on Knowledge Based Systems in Medicine, December 11-15 2000, Sydney.*

INTELLIGENT HEALTHCARE INFORMATION DISSEMINATION FEATURING ELECTRONIC MEDICAL RECORD PROFILING AND CUSTOMISED INFORMATION COMPOSITION

Syed Sibte Raza Abidi & Alwyn Goh
Health Informatics Research Group
School of Computer Sciences
Universiti Sains Malaysia
11800 Penang, Malaysia.
Email: sraza@cs.usm.my

Abstract

The featured Internet-based healthcare information dissemination system was designed based on the premise that focussed accessibility is a central component in the empowerment of individuals so as to able to exercise sound judgement with respect personal health-maintenance. This requirement necessitates a re-evaluation and re-design of the information dissemination process so as to emphasise: (1) customisation based on individual-specific environmental and healthcare requirements, and (2) proactive just-in-time delivery combining both *push* (system-initiated) and *pull* (user-initiated) oriented mechanisms. Our customisation model features Electronic Medical Record (EMR) based profiling in conjunction with personalised information composition drawing from a generic document base covering a wide spectrum of healthcare information of interest to the general population. The resultant Personalised Healthcare Information (PHI) package—covering both long-term and immediate health-maintenance requirements—would subsequently be periodically refreshed based on: (1) additional EMR-related information, (2) healthcare information updates of potential interest as ascertained from individual profiles, and (3) Web-mediated online consultation sessions; the last of which concludes with users being provided a dynamically generated and up-to-date information dosage. PHI delivery can also be executed via email, thereby ensuring the timely availability of high-quality information without over-dependence on user interactivity with the system.

1. Introduction

The current healthcare paradigm—with its emphasis on wellness maintenance at the individual and community levels—is predicated on the patient (or person) empowerment and individual ownership of most healthcare-related processes [1]. Individuals would certainly be better equipped to make 'informed' judgements about personal health maintenance if they were given easier and more focussed access to healthcare information [2] [3]. Easier accessibility—at least measured in terms of information volume—is seemingly addressed by the proliferation of healthcare information Web-sites over the Internet [4]. This approach does, however, lead to difficulties on the part of the non-specialist individual who has to (a) meticulously sift through volumes of healthcare information; (b) make 'value' judgements about the validity and relevancy of the available healthcare information; and (c) pro-act to find pertinent healthcare information. Rather, for maximum efficacy of healthcare information, what is needed is a web-based (healthcare) information dissemination system that provides the following value-added features:

- *Personalised Information*: Typical healthcare web-sites provide generic information targeting a wide audience. But, for maximum impact, personalised healthcare information for each individual should be 'dynamically' compiled from generic information resources, to specifically address his/her current health needs. In essence, personalised healthcare information should take into account the individual's (i) chronic (long term) and episodic (short term) healthcare needs—medical characteristics, (ii) healthcare objectives, (iii) educational needs, (iv) literacy level, and (v) social and community factors.
- *Just-in-time Information:* Up-to-date information should be pro-actively delivered, at regular intervals, to the individual—i.e. instead of the individual 'pulling' the information it should be periodically 'pushed' to the individual over the Internet.

The rationale for the featured *Personalised Healthcare Information Delivery System* (PHIDS) stems from the obvious value of disseminating the *right* information at the *right* instance [1] [5]. From a functional viewpoint, PHIDS is an intelligent system with a WWW-enabled information dissemination model based on: (1) proactive monitoring of an individual's health profile vis-à-vis his/her Electronic Medical Record (EMR); (2) personalisation of generic healthcare information with respect various individual-specific and situation-specific health issues, and (3) delivery based on both pull (Web) and push (email) mechanisms; all of which are intended to enhance patient empowerment. PHIDS leverages elements from Artificial Intelligence (AI), Databases and Internet technologies; and draws on information contained in: (1) subscriber EMR's encoded in ICD10 format; (2) healthcare information documents encoded as Extensible Mark-up Language (XML) files; and (3) constraint rules attached with healthcare information documents

2. Information Personalisation: The State of Affairs

Information personalisation, in the context of web-based information dissemination, can be regarded as the process of using an individual profile—either acquired explicitly by asking the user his/her preferences or implicitly from the individual's longitudinal information compiled over a period of time—to dynamically generate personalised information content. In this scenario, two individuals, each with a different profile, will experience a unique information package, conforming to their specific needs.

Precursor to the generation of personalised information is the process of acquiring the individual's profile (interests, needs, usage patterns, etc). Approaches to web-mediated user profiling can be classified into direct, inductive and deductive approaches:

- *Direct* approaches rely on the individual to dictate information pertaining to one's profile. For instance, the acquisition of static profiles obtained during user registration vis-à-vis answers to static questionnaires. Content based filtering techniques, such as GroupLens, as used by WebWatcher [6] and Letizia [7], are of note here. The profiles generated give a static subjective description of the user, and are prone to 'ageing'—i.e. they do not reflect the recent changes in the individual's behaviour.
- *Inductive* approaches tend to develop the individual's profile in a passive manner by sifting through the individual's personal information (stored in accessible databases) and web-usage patterns from web-server logs i.e. acquiring the individual's preferences based on the content of the pages visited [8] [9]. Typically, data mining techniques are used to meet this end. In particular, clustering techniques applied to an ensemble of user sessions enable to develop generic profiles for a group of people (with similar interests or needs).
- Deductive approaches use pre-defined representation schemas (such as records in a database), containing labelled individual-specific information to deduce a profile for the individual. Put simply, deductive rules are applied to labelled information to both generate the profile and then to determine the terminological consistency and conceptual cohesion of the generated profile [5].

In our work, at present, information personalisation is performed by adopting the deductive approach. We assume the EMR to reflect the medical profile of an individual. Selected medical information from the EMR is used to customise the content of the healthcare information as per the medical characteristic of an individual. Later, we plan to use data mining techniques to inductively derive both profiles of individuals and also to establish correspondences between PHI content generated for a group of individuals with similar profiles.

3. Functional Overview

The overall functionality of PHIDS can be divided into three main activities (shown in Figure 1):

(1) *Generation* of an up-to-date healthcare profile based on information contained in individual-specific EMRs and characteristic information acquired via system-initiated Web-based consultations sessions.
(2) *Authoring* of a Personalised Healthcare Information (PHI) package—by systematically amalgamating multiple Topic-specific Documents—based on the recently generated individual's health profile.
(3) *Delivery* of the PHI using both pull (client-mediated) and push (system-mediated) methods.

We discuss in detail certain key issues that are central to the overall functionality of PHIDS.

3.1 Information Extraction from EMR

A typical EMR usually contains free text data, which makes the task of extracting selected information from heterogeneous EMRs a daunting research effort. For our purposes, we assume the EMR to have a pre-determined structured outlook, whereby ICD10 codes are used by healthcare practitioners to record the patient's health status. Information extraction from the EMR then involves reading the ICD10 code (for the pre-selected fields of the EMR) and mapping it to the corresponding *Topic-specific Document (TD)*—a XML document containing information about a particular medical problem. To account for heterogeneous EMR, which may use non-standard terminology, we perform terminology mapping (to the standard terminology) using the Unified Medical Language Source (UMLS) meta-thesaurus.

It is important to note that regardless of the use of structured data within a EMR, we anticipate that at times the information available with an EMR cannot directly map to the most specific TD. In such situations, we map the EMR content to a default TD, which contains generic information about the most-related medical concept/condition. Default TD's are usually selected from higher levels of the medical ontology modelled within the PHI repository.

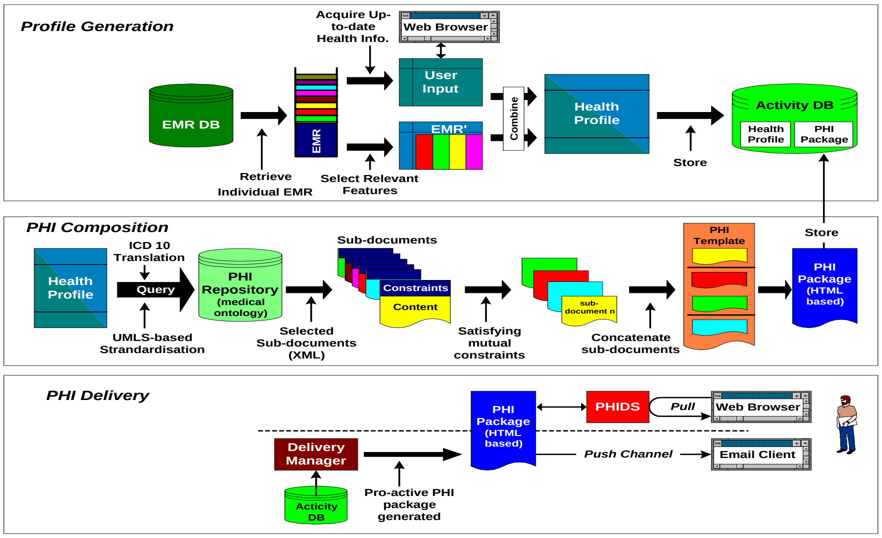

Figure 1: Overall functionality of PHIDS

3.2. Composition of the PHI Package

The PHI package, composed based on an individual's current health profile, includes information pertaining to (1) long-term clinical conditions and management regimes, (2) short-term therapy and rehabilitation associated with non-acute illness episodes, and (3) general information. The dynamically composed PHI package—a comprehensive XML-based healthcare document or brochure—is an amalgamation of multiple TDs, each with different levels of coverage and generality.

Research in the customisation of healthcare documents [10] [11] for both content and form has put forth two possible approaches. The first approach, purports the creation of a master document which comprises a large set of blocks of healthcare text (with annotations) which are selectively included or excluded for both content and form [10]. The second approach—the knowledge base approach—involves the creation of healthcare text using natural language generation techniques [12]. Here, the constituents of the text are pieces of knowledge in some knowledge representation formalism, which are then automatically selected for content but not for form. A knowledge base of rules is required for the selection of content, prepositional ordering and lexical choice—indeed a set of intractable problems.

A good example of using both techniques to generate healthcare documents for both content and form is the work by DiMarco et al [10]. Their 'generation by selection and repair' approach uses a master document that includes text selected on the basis of content which is then post-edited not only for form but also for style and coherence, i.e. polishing the document from a linguistic perspective. Note that their selection approach, though being straight forward, requires an extremely large number of bits of text to account for the content and the various modalities of the form.

Our approach for generating PHI is a hybrid of the above-mentioned approaches and can be termed as '*generation by ontological selection and constraint satisfaction*'. To cater for content we 'intelligently' select multiple (pre-authored) topic-specific *sub-documents*, which are subsequently concatenated according to a pre-defined *PHI template*. For content representation we use an ontology of medical concepts—the ontology is derived from the hierarchical representation of concepts given in ICD10 coding and the index of UMLS—where each level of the ontology refers to a unique medical concept (or topic). Each level of the medical ontology contains a unique TD corresponding to the medical concept/topic at that level. We will like to point out that the use of a medical ontology to organize the entire medical content in a healthcare information dissemination system is a way forward as it provides structure, an optimum number of medical concepts for wider coverage of issues, a common terminology to represent concepts and a hierarchical classification of medical concepts spanning from specific (as we go down the ontology) to generic (at the top of the ontology) concepts. To cater for form, we use an intelligent *constraint satisfaction* system (in conjunction with a knowledge base) to establish the suitability of the information contained in the multiple 'short-listed' TDs towards the general profile of the individual. In essence, the embedded constraint satisfaction system satisfies the mutual constraints that may arise when multiple 'focused' TDs are aggregated as a single document. For instance, say a particular TD recommends certain food in the diet but at the same time another selected TD restricts the consumption of the same food. In this case, we need the constraint satisfaction system to satisfy the mutual constraints between these two TDs and suggest what information to include and what to exclude or what information should be given a more generic level. So it may be noted that intelligent mechanisms, within our approach, are not used for the traditional polishing of the document vis-à-vis stylistic and linguistic customisations, rather for establishing the veracity of the combined PHI package as a seamless health information document.

Note that each TD observes a pre-defined format together with a set of local constraints that need to be satisfied at a global level for the TD to be included in the final PHI package. Functionally, if the constraints render a particular TD to be incompatible with the other sub-documents, then we move up the ontology and select a more generic TD, re-establish its appropriateness by satisfying its constraints and finally include it (to the PHI package) if deemed appropriate.

3.3. The PHI Template

Put simply, the *PHI template* is a specification of the manner in which disparate pieces of information (i.e. TDs) are to be organised to yield a seamless, structured and readable PHI package.

From an information presentation perspective, the PHI template can be divided into three main sections:

(1) Information about long term-term clinical conditions and management regimes. We provide a generic definition of the clinical condition followed by a more specific account of the disease, details of the undergoing management regime, other treatment choices (medication and surgical), drugs information—their behaviour and side effects, day-to-day management of the disease, recovery management, do's and don'ts, lifestyle choices and other information of the kind.

(2) Information about short-term therapy and rehabilitation associated with non-acute illness episodes.

(3) General healthcare education with a wellness maintenance connotation. For instance, facts and reports about diet, fitness and exercise, pregnancy (if present), stress management, downside of smoking (if practiced), community health, vaccinations and so on.

In practice, the PHI template is a stylesheet (XSL)—a pre-defined structure comprising place-holders for imported text (in XML format) and graphics. Figure 2 gives an exemplar template, whereas Figure 3 gives the composed PHI document derived by placing text at the designated places in the template.

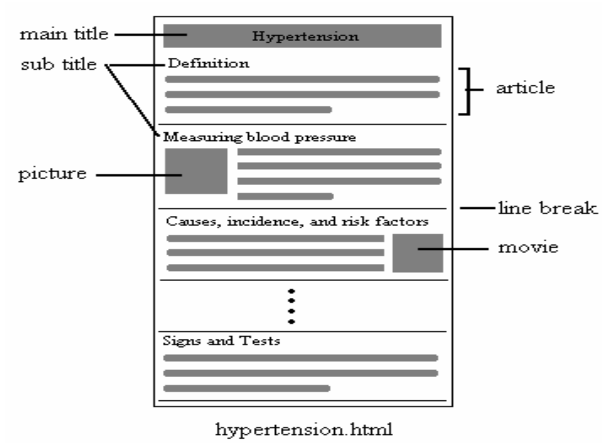

Figure 2: An Exemplar PHI template

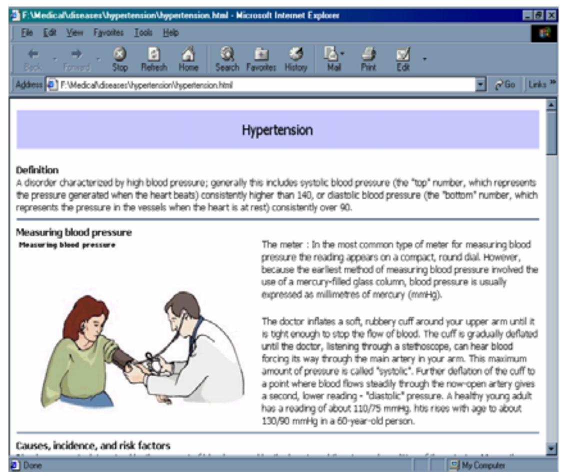

Figure 3: An exemplar HTML document derived from the PHI template

3.4. The Pro-Active Information Delivery Mechanism

We mentioned earlier that PHIDS incorporates both *push* and *pull* modes for information delivery. The operational characteristics are different for these two delivery modalities: The pull delivery mechanism represents the traditional mode of information dissemination, whereby the user initiates the information acquisition activity by accessing PHIDS via the Internet (more specifically a Web-based interactive session)—i.e. *pull* implies the user's prerogative to interact with PHIDS. On the contrary, the push delivery mechanism is enacted when PHIDS decides to disseminate information to the user, i.e. push corresponds to PHIDS's initiative to assist the user. Here, PHIDS decides to push information, via email, under the following two conditions:

- Update or addition of healthcare content that corresponds (or is deemed relevant) to the most recent health profile of an individual—the most recent health profile and the last version of the PHI package (list of XML files) delivered to an individual is stored for this purpose in the *Activity Database*.

In *International ICSC Congress on Intelligent Systems and Applications (ISA'2000), Special Session on Knowledge Based Systems in Medicine, December 11-15 2000, Sydney.*

- The EMR is updated (always by a healthcare practitioner) for any medical reason. We record the date of the EMR based on which the most recent healthcare profile is generated. If the date of the EMR is later than the date of the most recent healthcare profile it implies that the current healthcare profile has 'aged', thus initiating the generation of a recent healthcare profile and subsequently the delivery, by the push mechanism, of an up-to-date PHI package. It may be noted that the operational functionality of the push mechanism is made feasible by the implementation of a back-end *Delivery Manager*—akin to an autonomous intelligent agent—that monitors the PHI content repository and the EMR database for any changes.

4. Architecture of PHIDS

To meet the aforementioned functional objectives, PHIDS is implemented as an Intelligent Info-Structure that leverages elements from Artificial Intelligence (AI), Database (DB) and Internet technologies. Figure 4, shows the modular architecture of PHIDS. We briefly explain the modules of PHIDS vis-à-vis their functionality.

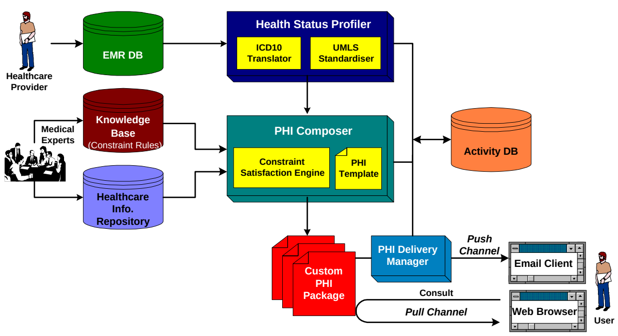

Figure 4: The architecture of PHIDS, illustrating the constituent modules and resources.

4.1. Electronic Medical Record Repository

The EMR repository contains EMR for individuals registered with PHIDS. The EMR adheres to ICD10 coding for recording patient information. Ideally, we seek global access to EMR repositories, but since we are in the preliminary stage of the project we are accessing EMRs from the on-campus medical centre over a LAN network.

4.2. Health-Status Profiler

The Health-Status Profiler (HSP) module determines the current health profile of an individual by intelligently collating information from two sources: (1) individual-specific EMRs, and (2) user-initiated Web-based consultation sessions. We introduce the concept of a *Health Status Window*—the time period over which the health profile will be generated. For all practical purposes, the health status window is the retrospective period spanning from the current date to some specified past date (typically set to 3 weeks, but can be altered by the user). In practice, HSP will refer to medical information available in the EMR, within the health status window, in order to determine the health profile.

The HSP, when invoked, proactively searches the EMR for (a) acute diseases with corresponding therapeutic regimes; and (b) episodic encounters with healthcare providers during the health status window. Terminology mapping—necessary if we presume non-terminology usage in the EMRs collected from heterogeneous sources—is performed by using the Unified Medical Language Source (UMLS) meta-thesaurus. A built-in ICD10 translator allows for the translation of ICD10 codes, available in the EMR, to standard medical terminology.

The health profile comprises information about acute and recently encountered illnesses (symptoms and treatments), lifestyle changes (if any), therapeutic and rehabilitation programs and other relevant medical information. Features of the HP are: (1) It can conduct a WWW-based consultation session with the user to validate certain aspects of his/her compiled health profile. A built-in intelligent diagnostic system intelligently asks a series of relevant questions and subsequently infers the 'most representative' health profile of the user (as shown in Figure 5); (2) It compiles only the most current and imminent healthcare needs, if any, without conferring to 'outdated' (as per the specified health status window), or previously serviced health needs; and (3) It incorporates a medical thesaurus (UMLS meta-thesaurus) to translate the non-standard vocabulary found in the EMR to the standard vocabulary used to index the healthcare information within PHIDS.

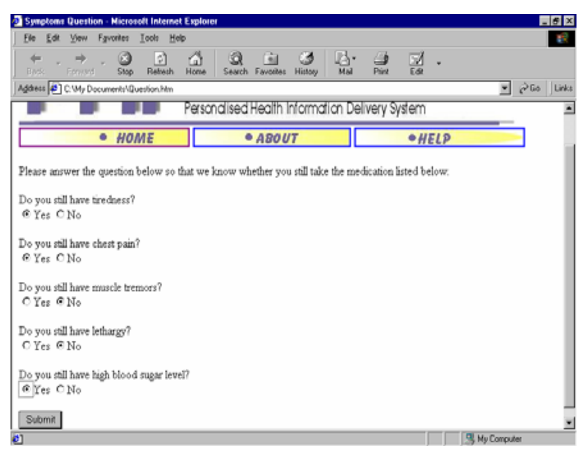

Figure 5: user-interface of the web-based consultation session with the user to verify his/her health characteristics.

4.3. Healthcare Information Repository

The *Healthcare Information Repository (HIR)* contains generic healthcare information (content) organised in a hierarchical (directory-based) structure modelling the medical ontology. At the top level, healthcare information is categorised into the following categories: *Allergy, Diseases, Drugs, Lifestyle, Symptoms, Medical Dictionary* and *Medical Vocabulary*. These categories then further expand to multiple levels, from generic to specific health topics/concepts, as per the medical ontology (as shown in Figure 6). Each level may contain TDs as XML files—pertaining to the specific topic/concept at that level—and sub-directories leading to more specific sub-documents on the same topic. It may be noted that although each TD is a unique information entity, it is related to the TDs at both higher and lower levels of the concept hierarchy.

The features of HIR are (a) the hierarchical structure manifests a taxonomy of healthcare information ranging from generic to specialised; (b) new categories (or sub-categories) of information can be added by simply creating a directory at the right level; and (c) new healthcare information can be readily added/updated by simply inserting a XML file (in the correct format) in the relevant directory.

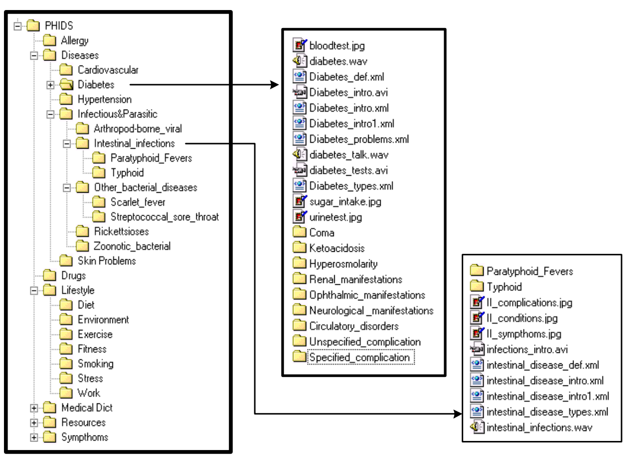

Figure 6: The hierarchical structure of the HIR, illustrating the content at each level.

4.4. Personalised Health Information (PHI) Composer

Profile generation is designed to allow for PHI composition based on: (1) long-term clinical conditions and management regimes, and (2) short-term therapy and rehabilitation associated with non-acute illness episodes. Individual and situation-specific composition of the PHI package subsequently proceeds with the dynamic adaptation of the collected TDs using constraint satisfaction techniques, and finally the aggregation (via stylesheeting) of the adapted XML-encoded TDs into a functionally seamless Hypertext Mark-up Language (HTML) document—the so-called PHI package.

The *PHI Composer*, comprises a *constraint satisfaction engine*—employing constraint logic programming techniques—to ensure the integrity of the dynamically authored PHI package in terms of medical correctness. We mentioned earlier that each TD includes a set of constraints which need to be satisfied to ensure that the information it contains is compatible with the rest of the selected TD. For instance, if we have two selected TDs, one on lifestyle recommends a particular diet which might have a high sugar content, whereas the other TD on diabetes maintenance prohibits the use of sugar-based and fatty substances. In this case, the two recommendations are contradictory to each other, and there is a need to determine what recommendation (information) to include and what to exclude. The constraint satisfaction engine is responsible for satisfying such constraints and it does so by simultaneously operating on all the available constraints using medical knowledge (rules) stored in the PHIDS medical knowledge base. At the conclusion of the constraint satisfaction process, a sub-set of the selected TDs—or at a more finer level a sub-set of the information contained in the TDs—is made available for composing the PHI package as per the specification of the PHI template. The composed PHI package is a functionally seamless XML document, which is subsequently converted to HTML format upon delivery (see Figure 7).

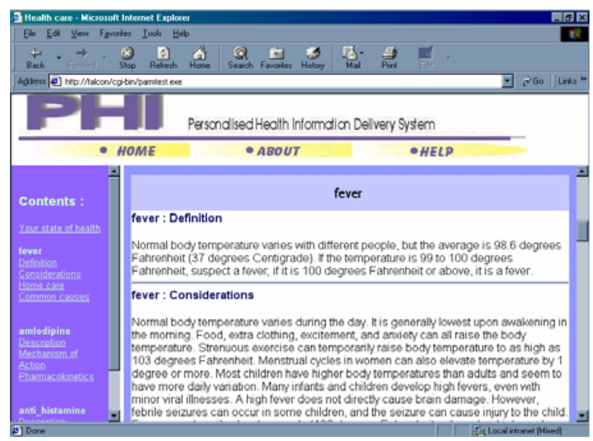

Figure 7: Screenshot of the PHI package for fever. The left hand side frame contains the navigator whereas the right hand side contains the PHI content.

An important feature of the PHI Composer is that it checks whether the PHI content has been previously delivered; if so, then depending upon the settings of the PHI Composer a decision is made whether to repeat the information or not. Note that, we do not encourage the repetitive presentation of PHI, as it might diminish the interest of the user.

4.5. PHI Delivery Modes

Delivery of PHI is made possible by two modes: (1) *Client-Motivated Mode* and (2) *System-Motivated Mode*. The functionality of each mode is as follows:

- *Client-Motivated Mode* involves the typical *Pulling* of desired information from a web-site. In this case, the user requests for a PHI package, which is then composed and sent to the user's web browser.
- *System-Motivated Mode* is the value-added and innovative delivery mode whereby up-to-date PHI is pro-actively and periodically *Pushed* to users, over the Internet to their email accounts—i.e. just-in-time PHI automatically delivered at the desktop. For registered users, PHIDS takes charge of their dynamic health needs and pro-actively compiles and delivers the 'best' PHI at scheduled intervals, based on the health status window. In this way, individuals are continuously empowered with up-to-date PHI, without the need to access health web-sites. We argue that with the implementation of the push delivery mode, PHIDS assumes the role of a *pro-active health guardian*.

5. PHIDS Implementation Details

The functional core of PHIDS is implemented using the Prolog programming language, specifically Visual Prolog version 5. Javascript is used for all web-based interfaces for information dissemination and also to collect inputs from users. The healthcare information content is developed in both XML format, with standardised XSL for style and DTD for content specification. The MS SQL server is used for all databases. Finally, we use IIS server as the web server for the delivery of PHI.

6. Concluding Remarks

Our work was initially motivated by the Malaysian Multimedia Super Corridor (MSC) Telemedicine initiative, which articulated the necessity for PHI dissemination and population-wide education [13]. We argue that in the face of both the growing volumes of healthcare information content (at healthcare-related web-sites) and the enhanced expectations of web users for focused information, the task of healthcare information personalisation—i.e. the personalisation of content as opposed to presentation style—is no longer a value-added feature of a healthcare-related web-site rather a necessity. The featured work does, however, serve as a test-bed to evaluate the effectiveness of PHI functionality and context, in addition to system design and operational considerations for a larger-scale deployment. We are also engaged in ongoing work on related healthcare-informational issues, in particular: (1) application of medical ontologies for content classification and compilation, (2) security and privacy associated with healthcare-related transactions, and (3) intelligent technologies, in particular constraint logic programming to satisfy complex requirements; all of which would hopefully contribute towards functional enhancements in subsequent PHIDS versions.

References:

[1] S.S.R. Abidi & Z. Yusoff. *Telemedicine in the Malaysian Multimedia Super Corridor: Towards Personalised Lifetime Health Plans.* In P. Kokol et al (Eds), MIE '99, IOS Press, Amsterdam..

[2] S. Marine, R. Guard & T. Morris. *A Model for Enhancing Worldwide Personal Health and Wellness.* In B. Cesnik, A. McCray & J. Scherrer (Eds.), MedInfo'98, IOS Press, Amsterdam, 1998.

[3] M. Ball & J. Douglas. *Medical Informatics in the New Millennium.* In B. Cesnik, A. McCray & J. Scherrer (Eds.), MedInfo'98, IOS Press, Amsterdam, 1998.

[4] B. Cho, T. Yoo, B. Huh & S. Cho. Information System for a Periodic Examination and Health Promotion Center. In B. Cesnik, A. McCray & J. Scherrer (Eds.), MedInfo'98, IOS Press, Amsterdam, 1998.

[5] S.S.R. Abidi & A. Goh. *A Personalised Healthcare Information delivery System: Pushing Customised Healthcare Information Over the WWW.* In Proc. Of Medical Informatics in Europe (MIE '2000), IOS Press, Amsterdam, 2000.

[6] T. Joachims, D. Freitag and T. Mitchell. *WebWatcher: A Guide for the World Wide Web*. In proceedings of the international joint conference in AI (IJCAI'97), 1997.

[7] H. Lieberman, H. Letizia. *An Agent that Assists Web Browsing*. In proceedings of the 14th international joint conference in AI (IJCAI'95), AAAI Press, California 1995.

[8] C. Shahabi, A. Zarkesh, J. Adibi and U. Shah. *Knowledge Discovery from Users Web-page Navigation*. In proceedings of workshop on research issues in data engineering, Birmingham 1997.

[9] R. Cooley, B. Mobasher and J. Srivastave. *Data Preparation for Mining World Wide Web Browsing Patterns*. Journal of knowledge and information systems, (1) 1,1999.

[10] C. DiMarco, G. Hirst, L. Wanner and J. Wilkinson. *HealthDoc: Customising Patient Information and Health Education by Medical Condition and Personal Characteristics*. In Workshop on Artificial Intelligence in Patient Education, Glasgow, 1995.

[11] V.J. Strecher, M. Kreuter et al. *The Effects of Computer Tailored Smoking Cessation Messages in Family Practice Settings*. In Journal of Family Practice, 39(3), 1994, pp. 43-49.

[12] A. Cawsey, K. Binsted and R. Jones. Personalised Explanations for Patient Education. In Proc. 5th European Workshop in natural Language Generation, Leiden, 1995.

[13] S.S.R. Abidi, A. Goh & Z. Yusoff. *Telemedicine and Medical Informatics in the Multimedia Super Corridor: The Malaysian Vision..* In B. Cesnik, A. McCray (Eds.), MedInfo'98, IOS Press, Amsterdam.