Automatic Home Medical Product Recommendation

Gang Luo • Selena B. Thomas • Chunqiang Tang
IBM T.J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, USA
{luog, selenat, ctang}@us.ibm.com

Abstract Web-based personal health records (PHRs) are being widely deployed. To improve PHR's capability and usability, we proposed the concept of intelligent PHR (iPHR). In this paper, we use automatic home medical product recommendation as a concrete application to demonstrate the benefits of introducing intelligence into PHRs. In this new application domain, we develop several techniques to address the emerging challenges. Our approach uses treatment knowledge and nursing knowledge, and extends the language modeling method to (1) construct a topic-selection input interface for recommending home medical products, (2) produce a global ranking of Web pages retrieved by multiple queries, and (3) provide diverse search results. We demonstrate the effectiveness of our techniques using USMLE medical exam cases.

Keywords Search engine · Language model · Personal health record · Home medical product · Nursing knowledge

1. Introduction

A few major Internet companies, including Google, Microsoft, and Revolution Health, are rapidly deploying Web-based personal health records (PHRs) [1]. PHR can help address the healthcare crisis by allowing ordinary consumers to actively manage their medical records and ultimately their health through a Web interface.

Existing PHRs have limited intelligence and can fulfill only a small portion of users' healthcare needs. To improve PHR's capability and usability, we proposed the concept of an intelligent PHR (iPHR) that was originally called intelligent consumer-centric electronic medical record in [1]. iPHR introduces and extends expert system technology and Web search technology into the PHR domain with the intention of being a centralized portal. This portal automatically provides users with comprehensive and personalized healthcare information to facilitate their daily activities of living. iPHR consists of multiple components: a PHR, a medical knowledge base, an expert system, and a search engine. Using medical knowledge, the expert system converts information in the PHR into a set of "search guide information" that reflects the user's medical conditions and healthcare needs. This search guide information will serve as seeds for the search engine to retrieve personalized healthcare information.

In this paper, we study automatically identifying relevant home medical products (HMPs) that can be recommended to users based on their medical records. This is a common application of iPHR. About 50% of Americans have one or more chronic conditions, and most of these people need HMPs [2, 22]. For example, an Alzheimer's patient would benefit from the following HMPs: (1) the Passive Infrared Alarm can alert a caregiver that the patient is getting out of bed (see http://www.allegromedical.com/patient-care-c530/pir-passive-infrared-alarm-pir-alarm-only-p199727.html), (2) the Grab Bar would reduce the patient's risk of falling in the bathroom (see http://www.allegromedical.com/bathroom-assists-c517/invacare-knurled-grab-bar-p188718.html), and (3) the Door Alarm can alert a caregiver to the patient's wandering behavior (see http://www.allegromedical.com/patient-care-c530/door-alarm-p197898.html).

Despite a great need for these products, potential consumers often are unaware of the HMPs that can help their situation and have difficulty in finding them. The HMP market changes rapidly as medical knowledge and technology improve. Consumers typically have little medical knowledge and cannot come up with appropriate keywords to search HMP catalogs. In general, physicians receive little training on HMPs and are unfamiliar with the HMP market. For these reasons, automatic HMP recommendation is a function highly desired in iPHR.

A user's medical record typically contains an overwhelming amount of information. To minimize the user's effort in finding desired HMPs, iPHR first extracts a preliminary set of topics from the user's medical record. Using medical knowledge, iPHR expands these topics into a more comprehensive set of topics. Next, iPHR will use a topic-selection interface [3] to obtain user input. In this way, ordinary people with little medical knowledge can avoid the traditional keyword interface that is difficult to use. iPHR also converts each topic selected by the user into one or more high-quality queries, which are submitted to the search system to retrieve HMPs. This topic-to-query conversion uses both disease/symptom treatment knowledge and nursing knowledge to bridge the semantic gap between the literal meaning and the underlying medical meaning of a topic. For example, "significant weight loss" as a symptom means that the user would like to gain weight rather than lose weight.

Although treatment knowledge and nursing knowledge are related, they are different. The former is taught in medical schools and the latter is taught in nursing schools. These two kinds of schools are separately organized and have different curricula. Leveraging nursing knowledge in iPHR is critical because treatment knowledge is insufficient.

In the entire domain of medicine, nursing is most closely related to iPHR's goal of facilitating people's lives. Much nursing is performed at home by consumers lacking nursing knowledge. Moreover, the scope of nursing is extensive. It covers a wide range of daily activities, and includes the diagnosis, treatment, and prevention of health problems. Without using nursing knowledge, iPHR would miss many HMPs. For example, for both Alzheimer's disease and its symptoms, neither their names nor their treatment methods appear in the three HMP Web pages mentioned above. Hence, nursing knowledge is needed to find these HMPs.

One key challenge in automatic HMP recommendation is to combine and rank HMPs retrieved by ***different*** queries, e.g., differing queries related to various home nursing activities (HNAs). To address this challenge, we develop an extended language modeling method that uses both nursing knowledge and treatment knowledge to compute a global ranking of HMPs retrieved by multiple queries. We first construct several heuristic constraints [4] that any reasonable ranking formula should satisfy. Then we extend the language modeling method [5] to meet these constraints by using nursing knowledge, treatment knowledge, and the semantic properties of our application scenario, and by folding all relevant factors into a single ranking formula. Our method is further enhanced to provide diverse search results so that the top ranked Web pages do not describe redundant HMPs.

We implemented a prototype iPHR system and evaluated the effectiveness of our techniques using USMLE medical exam cases [6]. Our experiments showed that iPHR significantly outperformed the keyword-based search engine of a leading HMP shopping Web site [7]. It was also observed that user satisfaction was crucially tied to iPHR's capability of automatically forming high-quality queries based on medical records and returning diverse search results.

The rest of the paper is organized as follows. Section 2 briefly reviews some basic nursing knowledge. Section 3 describes the user interface of iPHR. Section 4 presents our algorithm for recommending HMPs. Section 5 evaluates iPHR. Section 6 discusses related work. Section 7 concludes. The appendix provides a list of symbols used in this paper.

2. Some basic nursing knowledge

In this section, we first briefly review some basic nursing knowledge and then introduce some related notations. More details on this subject are available in [8].

2.1 Standardized nursing languages

Over two decades and the efforts of thousands of nurses, the nursing informatics community has systematically organized nursing knowledge into several standardized nursing languages [9]. Among those standardized nursing languages, iPHR's knowledge base has incorporated NANDA-I nursing diagnoses and NIC nursing interventions [8], which cover the full range of the nursing domain. NANDA-I and NIC are the acronyms for North American Nursing Diagnosis Association International and Nursing Interventions Classification, respectively.

NANDA-I nursing diagnoses are clinical judgments about individual, family, or community responses to actual or potential health problems [10]. Each medical condition links to one or more nursing diagnoses [10]. *NIC nursing interventions* are treatments that can be performed to enhance patient/client outcomes [11]. Each nursing intervention includes a list of usually 10 to 30 *nursing activities* that are used to implement it [11]. Every nursing diagnosis typically links to a list of 10 or more nursing interventions [9, 11]. Using nursing diagnoses and nursing interventions as intermediate steps, we can link each medical condition to multiple nursing activities. These nursing activities represent the actions that nurses, patients, and caregivers can take to achieve desirable outcomes for this medical condition.

2.2 Weight assignment for priorities

Nursing diagnoses, nursing interventions, and nursing activities all have various levels of priorities [8]. When recommending HMPs, iPHR considers these priorities by assigning them different weights. Each medical condition M links to a set S_M of nursing diagnoses. Each nursing diagnosis $D \in S_M$ has both a weight w_D reflecting its priority and a normalized weight $n\_w_D = w_D / \sum_{E \in S_M} w_E$ reflecting its "normalized" priority. Each nursing diagnosis $D \in S_M$ links to a set S_D of nursing interventions. Each nursing intervention $I \in S_D$ has both a weight w_I reflecting its priority and a normalized weight $n\_w_I = w_I / \sum_{J \in S_D} w_J$ reflecting its "normalized" priority. Each nursing intervention $I \in S_D$ includes a set S_I of nursing activities. Each nursing activity $A \in S_I$ has both a weight w_A reflecting its priority and a normalized weight $n\_w_A = w_A / \sum_{B \in S_I} w_B$ reflecting its "normalized" priority.

Some nursing activities must be performed by health providers in hospitals, whereas others are *home nursing activities* (HNAs) that patients and caregivers can perform at home (occasionally also in hospitals). This paper focuses on HNAs because of iPHR's consumer-centric view. A medical condition M often links to many HNAs. For instance, Alzheimer's disease links to over 200 nursing interventions and thousands of HNAs. Since many HNAs have corresponding HMPs, typically a large number of HMPs are applicable to a medical condition.

3. User interface

The user interface of iPHR's HMP recommendation function consists of a topic-selection input interface and an output interface.

3.1 Input interface

To recommend HMPs, iPHR needs to know the user's diseases, symptoms, and other medical conditions and healthcare needs. The input interface gathers this information by asking the user to make selections [3] and to input free text in three Web pages sequentially: a disease Web page, a symptom Web page, and an "other information" Web page. Often a HMP is used to treat a symptom (e.g., back pain) that can be caused by multiple diseases. Therefore, the user's input about symptoms is needed because the HMP description may mention only symptoms and not diseases. The collected information forms multiple topics reflecting the user's medical conditions and healthcare needs. The user can identify those topics that are most important to him.

3.1.1 Diseases

iPHR uses a disease Web page to obtain a list of diseases that the user currently cares about. In general, the user can have both chronic diseases and acute diseases. iPHR contains the user's PHR as one of its components and maintains a list of chronic diseases, $L_{chronic}$, of the user. Each time the user visits her doctor, a new entry is added into her iPHR, with zero, one, or more diseases S_d recorded in this new entry. For each disease $d \in S_d$, iPHR automatically adds d into $L_{chronic}$ if d is chronic. At any time, the user can delete from $L_{chronic}$ those chronic diseases that have already been cured. The user's current acute diseases L_{acute} are automatically obtained from the latest entry of iPHR, whereas we assume that the acute diseases in the other entries of iPHR either have already been cured or are no longer the user's focus. By combining $L_{chronic}$ with L_{acute}, we obtain a list $L_{d\_current}$ of current diseases of the user.

Fig. 1 shows the disease Web page, which displays all diseases in the list $L_{d\_current}$. The user can select the diseases that she cares about. She can also add into text fields other diseases not in $L_{d\_current}$, e.g., some acute diseases that are in the earlier entries of the medical record, but have not been cured. The diseases selected and entered by the user form a list L_{d_care}. For each disease $d \in L_{d\_care}$, the user can indicate whether d is highly important to her.

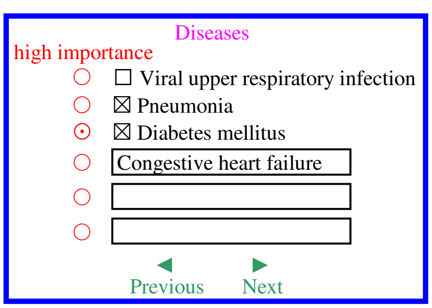

Fig. 1 The disease Web page in the input interface.

3.1.2 Symptoms

iPHR uses a symptom Web page to obtain a list of symptoms that the user currently cares about. iPHR has a knowledge base K_b that contains all diseases and their associated symptoms [12] using the standard names in [13]. For each disease in the list L_{d_care}, all possible symptoms are automatically obtained from K_b. These symptoms are merged into a list $L_{s\_possible}$ and displayed on the symptom Web page. Difficult symptom names are annotated with layman terms. For example, the symptom "hemoptysis" is explained as "coughing up blood." The user can select the symptoms that she cares about. She can also add into text fields other symptoms not in $L_{s\_possible}$. The symptoms selected and entered by the user form a list L_{s_care}. For each symptom $s \in L_{s\_care}$, the user can indicate whether s is highly important to her.

3.1.3 Other Information

iPHR uses an "other information" Web page to obtain additional medical conditions and healthcare needs that the user currently cares about. iPHR automatically provides a list $L_{c\_current}$ of current medical conditions of the user, which is constructed using the user's information in the medical record, including occupation, recent surgeries, pregnancy, breastfeeding, kids, medication, and medical equipment (e.g., wheelchair). For the medications that the user is taking, a medication database [29] is used to facilitate the task of automatically classifying them into different categories (eye drop, liquid, pill, or inhaled).

The medical conditions in the list $L_{c\_current}$ are displayed on the "other information" Web page. The user can select the medical conditions that she cares about. She can also add into text fields other medical conditions and healthcare needs not in $L_{c\_current}$. The additional medical conditions and healthcare needs selected and entered by the user form a list L_{c_care}. For each medical condition or healthcare need $c \in L_{c\_care}$, the user can indicate whether c is highly important to her.

To prevent the user from forgetting to input relevant information, we provide a reminder list of medical conditions and healthcare needs at the bottom of the "other information" Web page: "Please input missing information reflecting your medical conditions and healthcare needs. Such information includes, but is not limited to, planned activities, hobbies, computer usage, and medical equipment usage."

3.2 Output interface

The output interface of iPHR's HMP recommendation function displays the HMPs recommended by iPHR in one of two formats. The default format is the sequential order presentation traditionally used by search engines. By clicking a button, the user can switch to the alternative hierarchical format that has explicitly marked medical meanings [14]. Either format has its distinct advantages and disadvantages and is complementary to the other.

Let L_t denote the list of topics of concern by the user, which includes the items in the disease list L_{d_care}, the symptom list L_{s_care}, and the list L_{c_care} of medical conditions and healthcare needs. iPHR can recommend a large number of HMPs. These HMPs are related to various topics that have explicit medical meanings. In the sequential output interface, the HMPs recommended for all those topics are mixed together. The user can quickly see the HMPs recommended for various topics. However, she cannot easily find all HMPs recommended for a single topic.

The hierarchical output interface uses nursing knowledge and treatment knowledge to organize all recommended HMPs into a three-level hierarchy as shown in Fig. 2: the first level for topics, the second level for nursing interventions, and the third level for HMPs. The categories without any corresponding HMP are omitted from the hierarchy. Within a category, the user can easily find all HMPs recommended for it. However, navigation is required to view the HMPs recommended for another category. Moreover, HMPs in different categories can overlap, and hence the user can run into repeated results, which is undesirable.

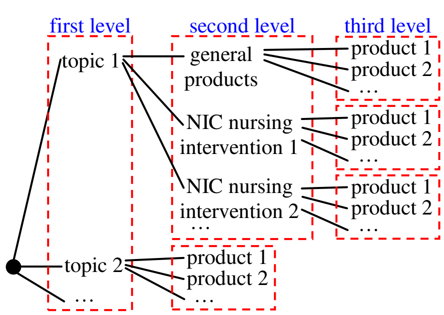

Fig. 2 The hierarchical output interface of iPHR's HMP recommendation function.

The first level of the hierarchy sorts all topics in descending order of their importance that the user specifies in the input interface. The second level of the hierarchy has several features that improve usability. First, each nursing intervention is accompanied by its definition in layman terms [11] so that ordinary users without much nursing knowledge can easily understand it. Second, for a topic T that is a medical condition, the "general products" category includes the HMPs retrieved by T's search guide information that is compiled from sources other than nursing knowledge, e.g., from T's name and treatment knowledge (see Section 4.1 below for details). Third, for a topic that is not a medical condition, all retrieved HMPs are listed directly at the second level and the third level does not exist.

4. Algorithm for recommending home medical products

Our algorithm for recommending HMPs consists of four steps. In step 1, we use an expert system equipped with a medical knowledge base to convert the original information in the medical record into a set of "search guide information" that reflects the user's medical conditions and healthcare needs. In step 2, with the help of the search guide information, the Web search engine retrieves relevant HMPs. In step 3, we use an extended language modeling method to combine and rank HMPs retrieved by different queries, while taking into account various relevant factors. In step 4, the search results are diversified and returned to the user.

4.1 Step 1: Obtaining search guide information

As mentioned in Section 3, the input interface collects a list L_t of topics of concern by the user. For each topic $T \in L_t$, T's name can be used as a query to retrieve HMPs directly. This straightforward approach, however, is often ineffective because a semantic gap can exist between T's name and T's underlying medical meaning. For instance, if T is a symptom, T's underlying medical meaning is to treat T, which can be in contrary to T's literal meaning. In this case, the keyword query of T cannot properly retrieve HMPs that are used to treat T. For example, consider a user who has the symptom "significant weight loss." She would like to gain weight rather than lose weight. The description of a HMP used for this purpose can include either "address unintentional weight loss" or "for general weight gain." As a second example, consider a user who has the symptom "chills." She would like to keep herself warm rather than become chilled. The description of a HMP used for this purpose is likely to include "help keep warm."

To bridge the semantic gap, we use expert system technology, nursing knowledge, and treatment knowledge to provide semantic translation from topics to their underlying medical meanings. The results are the search guide information for the search system. For each topic $T \in L_t$, the HMP search for T is performed using T's search guide information to increase the chance that the retrieved HMPs can reflect T's underlying medical meaning.

Some search guide information is obtained using treatment knowledge as follows. For each symptom s in the knowledge base K_b, a medical professional pre-compiles a set S_{s_p} of phrases and stores S_{s_p} in K_b. Each phrase in S_{s_p} represents one method of treating s. For each disease d in K_b, a medical professional pre-compiles a set S_{d_p} of phrases and stores S_{d_p} in K_b. Each phrase in S_{d_p} represents one method of treating d. In our prototype, these phrases are obtained from the information that physicians provide on the WebMD [15] and Healthline [12] Web sites. For example, losing weight helps alleviate the pain associated with bone spurs, and elevating the head during sleep helps treat the gastroesophageal reflux disease.

Besides treatment knowledge, nursing knowledge is also used to obtain more comprehensive search guide information. For each HNA A, a nurse pre-compiles a set S_A

of phrases as A's HMP search guide information and stores S_A in the knowledge base K_b. Each phrase in S_A provides one way of retrieving HMPs related to A. For example, for the HNA "use a bed alarm to alert caretaker that individual is getting out of bed," one phrase is compiled: "bed alarm." As another example, for the HNA "provide visible handrails and grab bars," two phrases are compiled: "handrail" and "grab bar."

For each medical condition $M \in L_t$, much search guide information is obtained using the HNAs linked to M. First we find the set S_M of nursing diagnoses linked to M. For each nursing diagnosis $D \in S_M$, we find the set S_D of linked nursing interventions. For each nursing intervention $I \in S_D$, we use S_I to denote the set of its HNAs. Then the HMP search guide information of these HNAs are merged into a set $R_M = \bigcup_{D \in S_M, I \in S_D, A \in S_I} S_A$ as part of M's search guide information compiled using nursing knowledge.

For each topic T in the list L_t, its search guide information G_T is a set of phrases. There are four possible cases:

(1) If T is a disease d, T is also a medical condition M. G_T includes d's name, the set S_{d_p} of phrases compiled for d using treatment knowledge, and the set R_M of phrases compiled for M using nursing knowledge.

(2) If T is a symptom s, T is also a medical condition M. G_T includes the set S_{s_p} of phrases compiled for s using treatment knowledge. If no disease in L_t has s, G_T also includes the set R_M of phrases compiled for M using nursing knowledge. Otherwise, R_M is not included in G_T because the HNAs for a disease d already include the HNAs for d's symptoms.

(3) If T is a medical condition M (e.g., recent hip surgery), but neither a disease nor a symptom, G_T includes T's name and the set R_M of phrases compiled for M using nursing knowledge.

(4) If T (e.g., preparation for an upcoming fishing activity) is not a medical condition, G_T contains a single phrase: T's name. In this case, we leave it for future work to compile other types of search guide information, e.g., sunburn prevention for the fishing activity.

The complete set of search guide information for all topics in L_t is $G = \bigcup_{T \in L_t} G_T$. For each phrase in G, its synonym phrases are also included in G.

For HMP search, our experience indicates that users usually prefer high recall to high precision. This is mainly for two reasons. First, the number of retrieved HMPs is not as overwhelming as the number of returned pages for a general Web search (e.g., president election). Second, users care about their health and are willing to manually filter out irrelevant HMPs, but they do not want to miss HMPs that can be helpful to them. Based on this observation, our principle of compiling the search guide information is to focus more on completeness than on specificity. As a side effect, for certain topics, their search guide information can retrieve irrelevant HMPs. For example, consider the symptom "weight loss" whose search guide information includes the phrase "weight loss." This phrase can retrieve both desired HMPs addressing unintentional weight loss and undesired HMPs helping accelerate weight loss. A task for future work is to improve precision without sacrificing recall.

4.2 Step 2: Finding relevant HMPs

In this section, we describe how to use Web search technology and search guide information to retrieve HMPs. We build a vertical search engine by crawling Web pages from a few selected, high-quality HMP shopping Web sites.

Let C denote the collection of crawled Web pages. For the Web pages in C, we use the frequent term sequence method in [3] to drop noisy information (e.g., advertisements), perform standard pre-processing steps in Web information retrieval [3], and build an inverted index I_i using the single-term vocabulary (i.e., the set of all distinct words). I_i contains sentence IDs and its format is slightly different from that of traditional inverted indices [16]. More specifically, consider a term t. In each entry E of t's posting list in I_i, there is a document ID showing that t appears in the corresponding Web page P. In addition, we include in E the IDs of all sentences in P in which t appears. Each sentence in P has a unique ID. In determining whether a HMP is relevant to a topic, we need these sentence IDs to check whether two or more terms appear in the same sentence.

The HMP search is performed using G, the complete set of search guide information. At a high level, our search method works as follows. Each phrase in G provides one way of representing the underlying medical meaning related to the user's medical condition and can be used to retrieve some relevant HMPs. The retrieved HMPs for all phrases in G are combined together as HMPs recommended to the user.

Next, we describe the HMP search method in detail. Consider a phrase f in G, the complete set of search guide information. After stopword removal, f generally contains one or more terms (words). We treat f as a Boolean conjunctive query and use it to retrieve Web pages in which each contains all terms in f. A Web page is considered relevant only if it has at least one sentence that contains all terms in f, i.e., f is treated as a sentence level Boolean conjunctive query. This is because f has a specific medical meaning and is usually well formed, e.g., compiled by a medical professional. If not all terms in f appear in one sentence of Web page P, the HMP described in P is unlikely to be relevant to T.

For example, consider the phrase "back pain" whose corresponding topic is the back pain symptom. If the term "back" appears only once and early in Web page P', whereas the term "pain" appears only once and late in P', the HMP described in P' is unlikely to be able to treat back pain. It is not required that all terms in the phrase f appear consecutively in the same sentence in the Web page. Otherwise, many relevant HMPs will be excluded. For instance, if a Web page describes a HMP p as "great for joint and muscle stiffness, pain & more," the two terms in the phrase "joint pain" are separated in that sentence, whereas p is indeed relevant to the corresponding topic "joint pain." In short, to retrieve HMPs using f, we take the intersection of posting lists for all terms in f and check sentence IDs stored in the inverted index I_i.

For each phrase f in the complete set G of search guide information, a set W_f of Web pages can be obtained using the method described above. These Web pages are merged into a set $R_{all} = \bigcup_{f \in G} W_f$ as HMPs recommended to the user. Typically each Web page in R_{all} describes a single HMP p. The name of p can be obtained using the wrapper induction technique [17].

The above discussion assumes that the vertical search engine crawls Web pages from a single HMP shopping Web site. In practice, Web pages from different Web sites can describe the same HMP. We can proceed in one of the following two ways to avoid overwhelming the user with redundant Web pages describing the same HMP:

(1) Default choice: For every HMP, only the most highly ranked Web page is kept in the set R_{all} (see Section 4.3 for the ranking algorithm). The other Web pages in R_{all} are dropped.
(2) Alternative choice: The Web pages describing the same HMP are clustered together [14]. In the hierarchical output interface, this is achieved by adding another level into the search result hierarchy. In the sequential output interface, this is achieved by forming a two-level hierarchy.

Recently, some HMP shopping Web sites (e.g., AllegroMedical [7]) have started to allow customers to write reviews on HMPs. Links to these reviews are provided on HMP Web pages. We find that customer reviews often describe innovative usage scenarios of the products beyond those in the original product descriptions. Our HMP search engine automatically adds reviews on a HMP p into the bottom of p's Web page to expand p's description. This helps retrieve a more complete set of relevant HMPs.

4.3 Ranking HMPs

iPHR uses an extended language modeling method to rank HMP Web pages. The language modeling method [5] is traditionally used to rank documents for a single query and unsuitable for iPHR because the HMP Web pages are retrieved by multiple phrases in different topics' search guide information. Depending on the numbers of terms in the queries, the relevance scores for various queries can be on different orders of magnitude, and thus cannot be compared directly. Below, we first briefly review the language modeling method and then present our enhanced algorithm for ranking HMP Web pages.

4.3.1 Background on language modeling

Language modeling [5] with Dirichlet smoothing [18] is a state-of-the-art method for ranking documents. Due to its superior performance and solid mathematical foundation, this method has attracted much attention in recent years. Assuming that all documents in a collection C have the same prior probability of being relevant to a query Q, this method uses the following formulas to compute the conditional probability of a document $D_o \in C$ given Q:

$$p(D_o \mid Q) = p(Q \mid D_o)\, p(D_o) / p(Q) \propto p(Q \mid D_o), \quad (1)$$

$$p(Q \mid D_o) = \prod_{q \in Q} p(q \mid D_o), \quad (2)$$

$$p(q \mid C) = c(q, C) / \mid C \mid, \quad (3)$$

$$p(q \mid D_o) = [c(q, D_o) + u \times p(q \mid C)] / (\mid D_o \mid + u). \quad (4)$$

Here, $c(q, D_o)$ is query term q's frequency in D_o, $c(q, C)$ is q's frequency in C, $|D_o|$ is the length of D_o in the number of terms, and $|C|$ is the length of C in the number of terms. u is a predetermined constant. Typically, as suggested in Zhai and Lafferty [18], $1000 \le u \le 10000$. Formula (1) uses Bayes' rule. Formula (2) assumes that all query terms are independent of each other given D_o. Formula (3) estimates $p(q|C)$, the most likely probability of generating q without looking into the content of D_o. Formula (4) uses a Dirichlet prior to avoid having zero probabilities, where the hyper-parameter of the Dirichlet distribution for q is estimated as $u \times p(q|C)$. All documents in C are ranked according to $p(D_o|Q)$, or equivalently $p(Q|D_o)$ that we term the *ranking probability*.

In iPHR, the HMP Web pages are retrieved by multiple phrases in different topics' search guide information. In general, the number of contained terms can vary significantly from one phrase to another. Formula (2) computes the ranking probability $p(Q|D_o)$ as the product of multiple numbers, one for each query term ($p(q|D_o)$). Consequently, the ranking probabilities of the Web pages retrieved by different phrases are frequently on different orders of magnitude. It would make equally important topics incomparable if such probabilities are used to rank the HMPs retrieved for various topics. This problem cannot be solved by computing $p(Q) = \prod_{q \in Q} p(q)$ and using $p(D_o|Q)$ to rank the Web pages retrieved by different phrases, as such computed $p(D_o|Q)$ is still proportional to the product of multiple numbers, one for each query term ($p(q|D_o)/p(q)$).

4.3.2 Heuristic ranking constraints

To help us derive the ranking formula, it is beneficial to consider the heuristic ranking constraints that any reasonable ranking formula should satisfy [4]. These constraints are necessary, but not the only sensible constraints. Rather, they are used to verify that the derived ranking formula is consistent with our intuitions. To properly compare the HNAs linked to the topics in the list L_t, we treat each HNA, rather than each term, as a semantic unit. Each HNA is represented by the phrases in its HMP search guide information. Traditional ranking methods count terms, whereas our ranking method counts HNAs.

We first introduce some notations. Let P, $P1$, and $P2$ denote HMP Web pages, $score_P$ denote P's relevance score, $M1$ and $M2$ denote medical conditions that are also topics and hence have weights, $D1$ and $D2$ denote nursing diagnoses, $I1$ and $I2$ denote nursing interventions, and A, $A1$, and $A2$ denote HNAs. All of these HNAs, nursing interventions, and nursing diagnoses link to some medical conditions in the list L_t. Each time all terms of a phrase in A's HMP search guide information appear in one sentence in HMP Web page P, we count it as one occurrence of A in P. We also count it as one occurrence of the nursing intervention I containing A, one occurrence of the nursing diagnosis D linked to I, as well as one occurrence of the medical condition linked to D. $c(A, P)$ is the number of A's occurrences in P. $n_a(P)$ is the length of P measured in the HNA semantic unit, i.e., the total number of HNAs' occurrences in P by counting multiplicity. These HNAs are arbitrary ones and not limited to the HNAs linked to the topics in the list L_t. Recall that n_w stands for normalized weight reflecting normalized priority. Also, the user specifies topic importance in the input interface. A topic T's importance is represented by a weight w_T.

We define the heuristic **HNA priority constraint** as follows:
Assume $n_a(P1)=n_a(P2)$. $P1$ mentions only one HNA, $A1$, which links to $M1 \in L_t$ through $I1$ and $D1$. $P2$ mentions only one HNA, $A2$, which links to $M2 \in L_t$ through $I2$ and $D2$. $c(A1, P1)=c(A2, P2)$. $n\_w_{I1}=n\_w_{I2}$, $n\_w_{D1}=n\_w_{D2}$, and $w_{M1}=w_{M2}$. If $n\_w_{A1}>n\_w_{A2}$, then $score_{P1}>score_{P2}$. If $n\_w_{A1}=n\_w_{A2}$, then $score_{P1}=score_{P2}$.
This constraint essentially means that with all other conditions being equal, a higher relevance score should be given to the Web page in which the HNA with higher normalized priority occurs. HNAs with the same normalized priority are treated equally. This is different from the case of traditional document ranking methods where different query terms are treated unequally, e.g., by using inverse document frequencies. The underlying reason is that traditional document ranking methods handle a single query. In contrast, multiple query phrases are used in our HMP recommendation scenario. We want to make the HMP Web pages retrieved by various query phrases comparable. Three additional priority constraints can be defined in a similar way: one for nursing invention, one for nursing diagnosis, and one for medical condition.

We define the **length normalization constraint** as follows:
$\forall k>1$. Assume $n_a(P1)=k \cdot n_a(P2)$. For each HNA A linked to the topics in L_t, $c(A, P1)=k \cdot c(A, P2)$. Then $score_{P1}>score_{P2}$.
Given two Web pages with equal proportion devoted to mentioning HNAs that are linked to the topics in the list L_t, this constraint means that a higher relevance score should be given to the longer Web page. This constraint reflects users' general preference of obtaining longer Web pages, as these pages tend to describe HMPs in greater detail and hence can be understood by users more easily.

4.3.3 Ranking formula

To properly rank HMP Web pages, we extend the language modeling method [5] by using nursing knowledge, treatment knowledge, and the semantic properties of iPHR's application scenario, and by folding all relevant factors into a single formula. Our high-level idea of deriving the ranking formula is to start from language modeling with Dirichlet smoothing [18] and make appropriate adjustments to satisfy the heuristic ranking constraints mentioned above. The resulting method is called the *extended language modeling method for multiple queries*. In Section 4.4, this method is enhanced to provide diverse search results. In both this section and Section 4.4, we focus on the sequential output interface. In the hierarchical output interface, the HMP Web pages within a category can be ranked in a similar way.

Next, we describe this method in detail. We first assume that all topics in the list L_t are medical conditions and only nursing knowledge is used in compiling the search guide information. In our HMP recommendation scenario, we have a conceptual query Q_c representing the user's need. To satisfy the priority constraints, we need to differentiate high-priority HNAs, nursing interventions, nursing diagnoses, and medical conditions from low-priority ones. For this purpose, we write Q_c into a disjunctive form: $Q_c = \vee_{M \in L_t,\, D \in S_M,\, I \in S_D,\, A \in S_I} (M \wedge D \wedge I \wedge A \wedge C_A)$, where S_M is the set of nursing diagnoses linked to the medical condition $M \in L_t$, S_D is the set of nursing interventions linked to the nursing diagnosis $D \in S_M$, S_I is the set of HNAs contained in the nursing intervention $I \in S_D$, and C_A is the essential content of the HNA $A \in S_I$. C_A can be regarded as A's HMP search guide information. The disjunction operator reflects the fact that the user's need is satisfied if any HNA linked to the topics in the list L_t is "hit."

Recall that each topic $T \in L_t$ has a weight w_T reflecting its importance. If the user specifies in the input interface that T is highly important, we have $w_T=w_H>1$. Otherwise, we have $w_T=1$. w_H is a predetermined constant. In our current implementation, the default value of w_H is 2. A user can adjust this value according to her preference and input. T's normalized importance is reflected by a normalized weight $n\_w_T = w_T / \sum_{U \in L_t} w_U$.

For each HMP Web page $P \in R_{all}$, we compute a relevance score $score_P$ according to which P is ranked. Ignoring the second- and higher- order terms, we have

$$\begin{aligned} p(Q_c \mid P) &= p(\vee_{M \in L_t,\, D \in S_M,\, I \in S_D,\, A \in S_I} (M \wedge D \wedge I \wedge A \wedge C_A) \mid P) \\ &\approx \sum_{M \in L_t,\, D \in S_M,\, I \in S_D,\, A \in S_I} p(M, D, I, A, C_A \mid P) \\ &= \sum_{M \in L_t,\, D \in S_M,\, I \in S_D,\, A \in S_I} [p(C_A \mid P, M, D, I, A) \cdot p(A \mid P, M, D, I) \cdot \\ &\qquad p(I \mid P, M, D) \cdot p(D \mid P, M) \cdot p(M \mid P)]. \end{aligned} \tag{5}$$

We make the following natural assumptions:

(1) The probability of generating the essential content C_A depends only on P. That is, $p(C_A \mid P, M, D, I, A) = p(C_A \mid P)$.

(2) The probability of selecting a HNA A depends only on the corresponding nursing intervention I and is proportional to A's weight w_A. That is,

$$p(A \mid P, M, D, I) = p(A \mid I) = w_A / \sum_{B \in S_I} w_B = n\_w_A.$$

(3) The probability of selecting a nursing intervention I depends only on the corresponding nursing diagnosis D and is proportional to I's weight w_I. That is,

$$p(I \mid P, M, D) = p(I \mid D) = w_I / \sum_{J \in S_D} w_J = n\_w_I.$$

(4) The probability of selecting a nursing diagnosis D depends only on the corresponding medical condition M and is proportional to D's weight w_D. That is,

$$p(D \mid P, M) = p(D \mid M) = w_D / \sum_{E \in S_M} w_E = n\_w_D.$$

(5) A medical condition M is a topic and hence has both a weight w_M and a normalized weight n_w_M. The probability of selecting M is independent of the Web page P and proportional to M's weight w_M. That is,

$$p(M \mid P) = p(M) = w_M / \sum_{U \in L_t} w_U = n\_w_M.$$

Under these assumptions, Formula (5) becomes

$$p(Q_c \mid P) = \sum_{M \in L_t,\, D \in S_M,\, I \in S_D,\, A \in S_I} [p(C_A \mid P) \cdot n\_w_A \cdot n\_w_I \cdot n\_w_D \cdot n\_w_M]. \quad (6)$$

Let $A \in P$ denote that HNA A occurs in Web page P. We assume that A's HMP search guide information is complete so that $p(C_A|P)=0$ if $A \notin P$. Then Formula (6) becomes

$$p(Q_c \mid P) = \sum_{M \in L_t,\, D \in S_M,\, I \in S_D,\, A \in S_I,\, A \in P} [p(C_A \mid P) \cdot n\_w_A \cdot n\_w_I \cdot n\_w_D \cdot n\_w_M]. \quad (7)$$

The above derivation considers only search guide information compiled using nursing knowledge. We introduce a few "virtual" concepts to integrate other types of search guide information into a unified framework. For each topic $T \in L_t$, let O_T denote the search guide information compiled from sources other than nursing knowledge, e.g., from T's name and treatment knowledge. We introduce the concepts of virtual HNA A_v, virtual nursing intervention I_v, and virtual nursing diagnosis D_v. O_T is regarded as A_v's HMP search guide information. A_v is included in I_v, I_v links to D_v, and D_v links to T. T is treated as a medical condition regardless of whether T is really a medical condition. In this way, Formula (7) can integrate O_T to compute $p(Q_c|P)$ using every $T \in L_t$. O_T is usually important for recommending HMPs. For example, if the name of a disease d appears in HMP Web page P, it is highly likely that P is relevant to d. To reflect this consideration, we assign the largest possible value, 1, to the normalized weights of A_v, I_v, and D_v.

The remaining work is to compute $p(C_A|P)$ for $A \in P$. According to Formulas (3) and (4), we could compute $p(C_A|P)$ as follows:

$$p(A \mid C) = c(A, C) / n_a(C), \quad (8)$$

$$p(C_A \mid P) = [c(A, P) + u \times p(A \mid C)] / [n_a(P) + u]. \quad (9)$$

Here, $c(A, C) = \sum_{R \in C} c(A, R)$ and $n_a(C) = \sum_{R \in C} n_a(R)$. However, the HNA priority constraint is violated because Formula (8) treats HNAs of the same normalized priority unequally. For example, consider the case of $n\_w_{A1}=n\_w_{A2}$ in the HNA priority constraint. We have

$$p(C_{A1}|P1) \cdot n\_w_{A1} \cdot n\_w_{I1} \cdot n\_w_{D1} \cdot n\_w_{M1} = score_{P1}$$
$$\neq score_{P2} = p(C_{A2}|P2) \cdot n\_w_{A2} \cdot n\_w_{I2} \cdot n\_w_{D2} \cdot n\_w_{M2}$$

unless $p(A1|C)=p(A2|C)$.

To treat HNAs of the same normalized priority equally, we assign the same value $p(A|C)=1/N_a$ to all HNAs, where N_a is the total number of distinct HNAs appearing in the collection C of crawled Web pages. There are several thousand ordinary HNAs [10, 11]. Both the total number of symptoms and the total number of diseases are on the order of thousands [12, 13], resulting in several thousand corresponding virtual HNAs. Consequently, an appropriate value of N_a is also on the order of thousands.

Recall that $n_a(P)$ is the total number of HNAs' occurrences in P by counting multiplicity. In general, users can input arbitrary topics in the input interface. As a result, a complete set of all possible HNAs that include virtual HNAs is unavailable. However, without such information, we cannot know the precise value of $n_a(P)$ used in Formula (9). To address this problem, we assume that HNAs occur in all Web pages at a uniform rate $r<1$ and estimate $n_a(P)$ as $|P| \times r$, where $|P|$ is the length of P in the number of terms.

By computing $p(A|C)$ and $n_a(P)$ in the way mentioned above, Formula (9) becomes

$$p(C_A \mid P) = [c(A, P) + u / N_a] / [|P| \times r + u]. \quad (10)$$

Putting Formulas (7) and (10) together, we obtain the final ranking formula with three parameters: N_a, r, and u. Their default values are 3,000, 0.2, and 3,000, respectively. It can be verified that this ranking formula satisfies the priority constraints. To verify that this ranking formula satisfies the length normalization constraint, we only need to show that for each HNA $A \in P2$ that is linked to the topics in L_t, we have $p(C_A|P1)>p(C_A|P2)$. That is, we need to show that

$$[k \cdot c(A, P2) + u / N_a] / [k \cdot |P2| \cdot r + u] > [c(A, P2) + u / N_a] / [|P2| \cdot r + u].$$

Some algebraic manipulation simplifies this to $c(A, P2) > |P2| \cdot r / N_a$. Since $A \in P2$, we have $c(A, P2) \geq 1$. As mentioned before, N_a is on the order of thousands and hence we usually have $N_a > |P2|$ whereas $r<1$. Thus, the inequality $c(A, P2) > |P2| \cdot r / N_a$ indeed holds.

4.4 Step 4: Diversifying search results

If we only use the extended language modeling method described in Section 4.3 to rank the retrieved HMP Web pages, the top-ranked Web pages can easily concentrate on a few topics rather than all topics in the list L_t. For example, the same product can be packed in various quantities and each such package is mentioned in a different HMP Web page with similar descriptions. If any of these Web pages is ranked high, the rest of these Web pages are also likely to be ranked high, whereas they provide little useful new information to the user.

In the past, studies have shown that searchers usually prefer diverse search results [19, 20]. Nevertheless, existing search result diversification methods [19, 20] are designed for the case of a single query. They cannot be directly applied to our HMP recommendation scenario that uses multiple query phrases. Ideally in our case, the first few Web pages returned should cover as many topics in the list L_t and provide as much new information as possible.

To provide diverse search results, we enhance the extended language modeling method. The set R_{all} contains $|R_{all}|$ retrieved Web pages sorted in descending order of their relevance scores. We use a constant N=1,000 to control the amount of time spent on search result diversification, re-rank the top $H = \min(N, |R_{all}|)$ Web pages in H passes, and generate one result page of twenty diverse HMPs at a time. In each pass, we pick a Web page that strikes a balance among three factors: (1) offering much new information to the user, (2) having a large relevance score, and (3) providing a balanced coverage of different topics in the list L_t, and their linked nursing diagnoses, nursing interventions, and HNAs. These three factors are combined into a single diversity score. The amount of new information contained in Web page P is measured by the dissimilarity between P and the Web pages previously returned to the user. The concrete method is as follows.

We form two sets: $S_{remaining}$ and $S_{returned}$. At any time, $S_{remaining}$ contains the Web pages remaining to be returned to the user, while $S_{returned}$ contains the Web pages already returned to the user. Initially, $S_{remaining}$ contains the top H Web pages with the largest relevance scores, whereas $S_{returned}$ is empty.

In the i-th ($1 \le i \le H$) pass, we compute a diversity score d_score_P for each Web page $P \in S_{remaining}$ as follows:

$$dissimilarity(P, R) = 1 - cosine\_similarity(P, R),$$

$$dissimilarity(P, S_{returned}) = \min_{R \in S_{returned}} dissimilarity(P, R),$$

$$d\_score_P = score_P \times dissimilarity(P, S_{returned}).$$

Intuitively, the larger the $score_P$ and the more dissimilar P is to the Web pages in $S_{returned}$, the larger the d_score_P will be. There are multiple ways of measuring the dissimilarity of two Web pages. Our current implementation uses one of the most popular ways: one minus their cosine similarity [16]. The Web page $P_{l\_d} \in S_{remaining}$ with the largest diversity score is moved from $S_{remaining}$ to $S_{returned}$ as the i-th Web page returned to the user. For each HNA A that occurs in P_{l_d} and links to some topic in the list L_t, appropriate discounts are given to the weights and normalized weights related to A. Specifically, suppose A is contained in the nursing intervention I, I links to the nursing diagnosis D, and D links to the topic $T \in L_t$. If A is a virtual HNA, I is a virtual nursing intervention and D is a virtual nursing diagnosis. T's normalized weight n_w_T is discounted by a constant factor s_T whose default value is 0.5. D's normalized weight n_w_D is discounted by s_D. I's normalized weight n_w_I is discounted by s_I. A's normalized weight n_w_A is discounted by s_A. s_D, s_I, and s_A are three constant factors whose default values are all 0.85. According to Formula (7), the relevance scores of the Web pages in $S_{remaining}$ depend on n_w_T, n_w_D, n_w_I, and n_w_A, and thus need to be re-computed. As a result, the more Web pages related to topic T that have been returned to the user, the less likely the next returned Web page will be related to T. A similar property exists for nursing diagnoses, nursing interventions, and HNAs.

5. Experimental results

We implemented a prototype iPHR system supporting the function of automatically recommending HMPs. We conducted experiments under a wide range of medical scenarios to demonstrate the effectiveness of our proposed techniques.

5.1 Setup

We crawled Web pages from AllegroMedical [7], the first and one of the largest HMP shopping Web sites. It sells over 42,000 HMPs covering almost every aspect of healthcare. Our experience indicates that it usually provides more detailed HMP descriptions than other HMP shopping Web sites. Below, we refer to AllegroMedical's keyword-based HMP search engine as AMSE.

We compared AMSE with two variants of iPHR using the sequential output interface. The first one is called treatment-based iPHR, which uses only treatment knowledge in compiling search guide information. The second one is called nursing-based iPHR, which uses both treatment knowledge and nursing knowledge in compiling search guide information. We used United States Medical Licensing Examination (USMLE) Step 2 CS (Clinical Skills) medical exam cases [6]. Physicians have to pass this USMLE exam to obtain their licenses for practicing medicine. Each exam case has both a sample medical record and a summary including a several-page-long, detailed description of the patient's situation. One such medical case is shown as follows:

> The patient is a 61 years old male complaining of fatigue and weakness. The patient notes that the fatigue and weakness started 6 months ago. He feels tired all day. He has poor appetite and lost 8 pounds in the last 6 months. He also complains of occasional nausea and of a vague, deep epigastric discomfort that radiates to the back. He feels sad sometimes, has lost interest in things that he used to enjoy ...

We randomly selected 30 USMLE medical exam cases as our test cases. Since USMLE covers both the typical cases and almost every aspect of daily medical practice, our random samples have a broad coverage of medical topics. Twenty-one people, twelve females and nine males, served as users. Their median age is 40. All of them are regular, ordinary Internet users without formal medical training, and hence represent iPHR's targeted users. Eighteen of them have received college education or above. Each user searched for all 30 medical cases. When users encountered difficulty in understanding medical case descriptions, a nurse was available to explain. For every medical case, each user randomly selected either AMSE, the treatment-based iPHR, or the nursing-based iPHR with equal probability, and had up to 40 minutes to search. The search session was terminated when either the user felt she had found enough relevant HMPs or time ran out, whichever came first. We allowed users to search for a relatively long time because users care about their health and often spend significant time searching medical information.

Similar to the TREC interactive track [21] that provides a standard approach for comparing the performance of various information retrieval systems, we use two sets of measures as the performance metrics for HMP search: one set is objective and the other set is subjective. The objective performance measures include the number of search result Web pages viewed and the time spent on the search process. The subjective performance measures include the number of desired HMPs found, ease of using the system, usefulness of the search results, and overall satisfaction with the system. For iPHR, the average usefulness of the returned top ten HMP Web pages is also included. A HMP Web page *P* is useful if the HMP described in *P* can help the user's medical condition, especially the highly important topics, and much of *P*'s relevant content has not been mentioned in the Web pages that are ranked higher [19] (i.e., the search results are diversified, see Section 4.4). Except for the number of desired HMPs found, all of these subjective performance measures are on a 7-point scale, with 1=low and 7=high [21]. They were obtained from a brief questionnaire that users completed after using the systems. For each objective or subjective performance measure, an average is computed for all 30 medical cases and users, and both its mean and its standard deviation are reported when appropriate. We used ANOVA [32] as the significance test. Our experiments were performed on a computer with two 3GHz processors, 2GB memory, and one 111GB disk.

5.2 Overall results

iPHR is efficient at searching HMPs. For all 30 medical cases, the average time taken by treatment-based iPHR to generate the first result page of twenty HMPs is less than one second. For nursing-based iPHR, the average time is less than two seconds. This time is longer than that taken by treatment-based iPHR, as more queries are formed using nursing knowledge and more search results are retrieved.

As will be shown in Table 2, compared to using either AMSE or the treatment-based iPHR, a user can find many more desired HMPs using the nursing-based iPHR. Due to the high quality of the search results provided by the nursing-based iPHR, a user of the nursing-based iPHR views more search results and spends more time on reading the results than a user of AMSE or the treatment-based iPHR does (see Table 1). Both differences are statistically significant.

Table 1 Objective performance measures on recommending HMPs (* means significant at <0.05 level compared to the nursing-based iPHR).

mean (standard deviation)	AMSE	treatment-based iPHR	nursing-based iPHR
number of search result Web pages viewed	30^{*} (7)	24^{*} (6)	39 (7)
time (minutes)	24^{*} (5)	18^{*} (5)	28 (6)

Table 2 Subjective performance measures on recommending HMPs (* means significant at <0.05 level compared to the nursing-based iPHR).

mean (standard deviation)	AMSE	treatment-based iPHR	nursing-based iPHR
number of desired HMPs found	11^{*} (5)	16^{*} (6)	27 (6)
ease of use	4.4^{*} (0.9)	5.4 (1.1)	5.4 (1.1)
usefulness	4.0^{*} (0.9)	5.2 (0.9)	5.1 (1.0)
satisfaction	4.2^{*} (1.0)	4.9^{*} (1.0)	5.6 (0.8)

Table 2 shows the subjective performance measures. Compared to AMSE, either variant of iPHR has advantages in all aspects. The search process in AMSE is tedious because the user needs to manually construct multiple queries, one for each topic that she cares about. Due to limited medical vocabulary and lack of medical knowledge, the user often encounters difficulties in constructing effective queries. Frequently the manually constructed queries are incomplete and cannot fully cover the principal topics described in the medical record, as users forget important issues from time to time. In contrast, iPHR automatically forms multiple queries based on the medical record and built-in medical knowledge, which allows the user to retrieve HMPs on several topics in one pass. Moreover, iPHR provides diverse search results. Consequently, users find that either variant of iPHR retrieves a larger number of desired HMPs, is easier to use, produces search results that are more useful, and is more satisfactory than AMSE. All of these differences are statistically significant.

Compared to the treatment-based iPHR, the nursing-based iPHR has advantages in two aspects. Both variants of iPHR use roughly the same interface, and hence are equally easy to use. The quality of nursing knowledge is as high as that of treatment knowledge. Thus, nursing-based iPHR's search results and treatment-based iPHR's search results have the same level of precision and are almost equally useful. Due to the use of nursing knowledge, users find that nursing-based iPHR retrieves a larger number of desired HMPs and is more satisfactory than treatment-based iPHR. Both differences are statistically significant. Overall, when using the nursing-based iPHR, it is worth spending extra time viewing more search result Web pages to know a larger number of relevant HMPs.

The nursing-based iPHR uses three parameters for ranking: N_a, r, and u (see Section 4.3.3). It also uses four parameters for search result diversification (see Section

4.4): s_T (topic weight discount factor), s_D (nursing diagnosis weight discount factor), s_I (nursing intervention weight discount factor), and s_A (HNA weight discount factor). Suppose $s_D=s_I=s_A=s_{DIA}$. Fig. 3 shows the impacts of N_a and r on the average usefulness of the returned top ten HMP Web pages. When N_a is too small (or too large), the total number of distinct HNAs appearing in the collection C of crawled Web pages is underestimated (or overestimated). When r is too small (or too large), the total number of HNAs' occurrences in a Web page is underestimated (or overestimated). In any of those cases, the estimation errors decrease average usefulness. The safe ranges for N_a and r are [1,000, 10,000] and [0.05, 0.75], respectively.

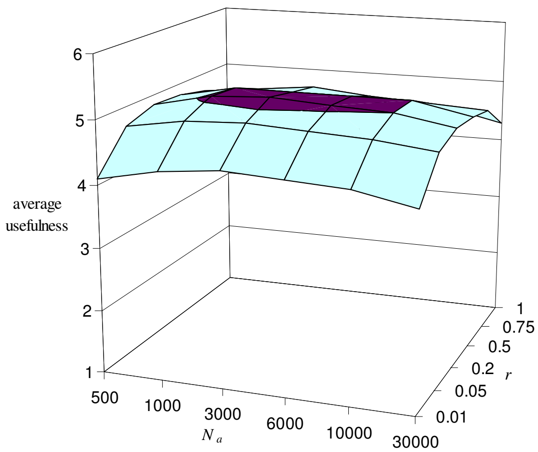

Fig. 3 Average usefulness of the returned top ten HMP Web pages vs. N_a and r.

Fig. 4 shows the impacts of u and s_T on the average usefulness of the returned top ten HMP Web pages. Within the safe range of [1,000, 10,000] suggested by Zhai and Lafferty [18], u has little effect on the average usefulness. When s_T is too small, the weights of highly important topics decrease so rapidly that those topics are insufficiently covered in the returned top Web pages. When s_T is too large, little discount is given to the weights of those topics, and hence the other topics of concern by the user are insufficiently covered in the returned top Web pages.

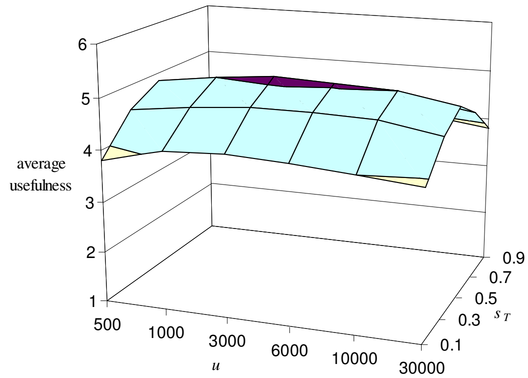

Fig. 4 Average usefulness of the returned top ten HMP Web pages vs. u and s_T.

Fig. 5 shows the impacts of the parameters s_T and s_{DIA} on the average usefulness of the returned top ten HMP Web pages. When s_T is either too small or too large, the average usefulness becomes lower. A similar property exists for s_{DIA}.

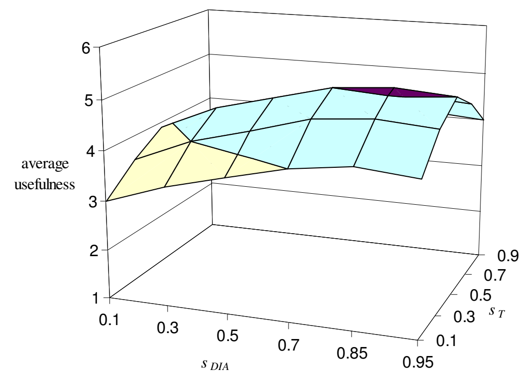

Fig. 5 Average usefulness of the returned top ten HMP Web pages vs. s_T and s_{DIA}.

Using default values of the parameters, the average usefulness is 5.1. If search result diversification is not performed, the average usefulness will drop to 2.7. In summary, the average usefulness is maximized around the default values of the parameters. Each of the parameters has a not-very-small safe range, within which the average usefulness is insensitive to parameter changes. However, if the value of a parameter is outside its safe range, the average usefulness may drop. All of our techniques are necessary to improve users' satisfaction with the system.

6. Related work

The concept of intelligent electronic medical record (EMR) was proposed more than a decade ago. However, existing intelligent EMRs are ***physician-centric*** and their intelligence is used to facilitate physicians' daily professional tasks, e.g., inputting and summarizing patient information [23]. In contrast, iPHR is ***consumer-centric*** and its intelligence is used to facilitate consumers' daily activities of living.

Cimino has built an Infobutton Manager for EMRs [30]. For each medical concept appearing in the EMR, the Infobutton Manager provides a fixed set of questions that physicians ask most often and uses manually pre-constructed queries to retrieve answers to these questions from certain resources in real time. Each answer is retrieved using one query. The Infobutton Manager does not recommend HMPs. In contrast, for each health condition, our iPHR often uses hundreds of search guide phrases simultaneously to retrieve HMPs. As a result, iPHR can retrieve a much more comprehensive set of relevant HMPs than traditional keyword search, irrespective of whether a long query or a short query is used in traditional keyword search.

Farfan et al. use ontology to facilitate keyword search in EMR [31], by incorporating the fact that many EMR standards are XML-based and have a hierarchical format. The method proposed in [31] does not apply to retrieving HMPs because HMP description has no hierarchical format.

Consumer-centric medical information retrieval is a broad, new research domain driving our long-term research. Our work in the past differs significantly from this work. Both our MedSearch system [19] and our iMed system [3, 14] focus on *disease diagnosis*, and use disease diagnosis knowledge to help users find disease information related to their medical condition. The application in this work is recommending HMPs, which differs from disease diagnosis and is addressed with techniques different from those in [3, 14, 19]. [1] proposed the general concept of iPHR and suggested that HMP recommendation could be a useful function of iPHR. This paper works out the details of the HMP recommendation algorithm. In particular, we demonstrate that leveraging nursing knowledge is crucial for high-quality HMP recommendation.

Many product search engines have been launched [24]. They use no medical knowledge and cannot automatically recommend HMPs tailored to consumers' medical conditions and healthcare needs.

Personalized search is a current trend of Web search engines [25]. Existing personalized search techniques adjust search results based on a user's search history and desktop content. Those techniques are useful for general search. However, since they leverage neither the user's medical history stored in PHR nor medical knowledge essential to obtain high-quality queries, they are less effective at providing useful, personalized healthcare information, which is the focus of iPHR.

In distributed information retrieval and meta-search engines [26], search results from multiple sources for the same query are merged together. In contrast, in our case of automatic HMP recommendation, we need to merge together the HMP Web pages retrieved for different topics (by various query phrases).

Besides iPHR, home care nurses also provide healthcare information to facilitate people's daily activities. However, these nurses are expensive to hire and have limitations due to their incomplete knowledge. The scope of nursing is so extensive that each nurse knows only a small part of it. Nevertheless, a person often has multiple medical situations and requires a wide range of healthcare information (e.g., 21% Americans have multiple chronic conditions [27]). Moreover, healthcare information keeps updating rapidly and no nurse can always keep up with the latest ones. For example, as medical knowledge and technology continue to improve, each year many new HMPs enter the market. We would expect iPHR to complement home care nurses in providing healthcare information, because its knowledge base stores a comprehensive set of nursing knowledge compiled by thousands of nurses whereas its search system can discover the latest healthcare information from the Web. For a similar reason, iPHR can provide more complete information on HMPs than any single consumer-oriented health information book (e.g., the series of books entitled "The Comfort of Home" [28]) or medical Web site (e.g., WebMD [15]).

7. Conclusions

This paper uses automatic HMP recommendation as a concrete application to demonstrate that iPHR can help users obtain personalized healthcare information to facilitate their daily activities of living. We show that iPHR requires no special user training, forms high-quality queries automatically, and provides diverse and relevant search results. These features are attractive to ordinary users who have little medical background. Our experiments with a wide range of medical scenarios demonstrate that compared with the keyword-based search engine of a leading HMP shopping Web site, iPHR significantly improves user satisfaction by recommending HMPs effectively and efficiently.

Acknowledgment We thank Annemarie Credentino, Jiuxing Liu, Ying-li Tian, Jing Wang, and Debra J. Zeitlin for helpful discussions.

References

1. Luo, G., Thomas, S. B., and Tang, C., Intelligent consumer-centric electronic medical record. *Proceedings of MIE'09*, pp. 120-124, 2009.
2. *U.S. market for home care products*, 5th ed. Kalorama Information, 2007.
3. Luo, G., Design and evaluation of the iMed intelligent medical search engine. *Proceedings of ICDE'09*, pp. 1379-1390, 2009.
4. Fang, H., Tao, T., and Zhai, C., A formal study of information retrieval heuristics. *Proceedings of SIGIR'04*, pp. 49-56, 2004.
5. Ponte, J. M., and Croft, B. W., A language modeling approach to information retrieval. *Proceedings of SIGIR'98*, pp. 275-281, 1998.
6. Le, T., and Bhushan, V., *First aid for the USMLE step 2 CS (clinical skills exam)*. McGraw-Hill, 2006.
7. AllegroMedical homepage. http://www.allegromedical.com, 2009.
8. Luo, G., and Tang, C., Automatic home nursing activity recommendation. *Proceedings of AMIA'09*, pp. 401-405, 2009.
9. Johnson, M., Bulechek, G. M., and Dochterman, J. M., *et al.*, *NANDA, NOC, and NIC linkages: nursing diagnoses, outcomes, and interventions*, 2nd ed. Mosby, 2005.
10. Ackley, B. J., and Ladwig, G. B., *Nursing diagnosis handbook: an evidence-based guide to planning care*, 8th ed. Mosby, 2007.
11. Bulechek, G. M., Butcher, H. K., and Dochterman, J. M., *Nursing interventions classification (NIC)*, 5th ed. Mosby, 2007.

12. Healthline diseases & conditions A to Z directory. http://www.healthline.com/directory/diseases-and-conditions, 2009.
13. Medical symptom descriptions. http://www.wrongdiagnosis.com/lists/symptom.htm, 2009.
14. Luo, G., Intelligent output interface for intelligent medical search engine. *Proceedings of AAAI'08*, pp. 1201-1206, 2008.
15. WebMD homepage. http://www.webmd.com, 2009.
16. Manning, C. D., Raghavan, P., and Schutze, H., *Introduction to information retrieval*. Cambridge University Press, 2008.
17. Kushmerick, N., Weld, D. S., and Doorenbos, R. B., Wrapper induction for information extraction. *IJCAI* (1) 1997: 729-737.
18. Zhai, C., and Lafferty, J. D., A study of smoothing methods for language models applied to ad hoc information retrieval. *Proceedings of SIGIR'01*, pp. 334-342, 2001.
19. Luo, G., Tang, C., and Yang, H., *et al.*, MedSearch: a specialized search engine for medical information retrieval. *Proceedings of CIKM'08*, pp. 143-152, 2008.
20. Carbonell, J. G., and Goldstein, J., The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of SIGIR'98,* pp. 335-336, 1998.
21. TREC interactive track homepage. http://trec.nist.gov/data/interactive.html, 2009.
22. Hidola homepage. http://www.hidola.com, 2010.
23. Yang, K., Evens, M. W., and Trace, D., Improving the text summaries in the intelligent medical record system. *Proceedings of MAICS'06*, 2006.
24. Google product search homepage. http://www.google.com/products, 2009.
25. Teevan, J., Dumais, S. T., and Liebling, D. J., To personalize or not to personalize: modeling queries with variation in user intent. *Proceedings of SIGIR'08*, pp. 163-170, 2008.
26. Baeza-Yates, R. A., and Ribeiro-Neto, B. A., *Modern information retrieval*. ACM Press/Addison-Wesley, 1999.
27. Vogeli, C., Shields, A. E., and Lee, T. A., *et al.*, Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *JGIM* 2007; 22 (Suppl 3): 391–395.
28. Meyer, M. M., and Derr, P., *The comfort of home: a complete guide for caregivers*, 3rd ed. CareTrust Publications LLC, 2007.
29. RxList - the Internet drug index homepage. http://www.rxlist.com, 2009.
30. Cimino, J. J., Use, usability, usefulness and impact of an Infobutton manager. *Proceedings of AMIA'06*, pp. 151-155, 2006.
31. Farfan, F., Hristidis, V., Ranganathan, A., and Weiner, M., XOntoRank: Ontology-aware search of electronic medical records. *Proceedings of ICDE'09*, pp. 820-831, 2009.
32. Bickel, P. J., and Doksum, K. A., *Mathematical statistics: basic ideas and selected topics, Vol. 1*. Prentice Hall, 2001.

Appendix

List of symbols

A, $A1$, $A2$	nursing activity or HNA
A_v	virtual HNA
c	medical condition or healthcare need
C	the collection of crawled Web pages
$\|C\|$	the length of the document collection C in the number of terms
C_A	the essential content of the HNA A
$c(A, P)$	the number of HNA A's occurrences in the Web page P
$c(q, C)$	query term q's frequency in the document collection C
$c(q, D_o)$	query term q's frequency in the document D_o
d	disease
D, $D1$, $D2$	nursing diagnosis
D_o	document
$\|D_o\|$	the length of the document D_o in the number of terms
D_v	virtual nursing diagnosis
d_score_P	diversity score for the Web page P
f	phrase
G	the complete set of search guide information for all topics in L_t, the list of topics of concern by the user
G_T	search guide information of the topic T
I, $I1$, $I2$	nursing intervention
I_i	inverted index
I_v	virtual nursing intervention
K_b	knowledge base
L_{acute}	list of current acute diseases of the user
$L_{chronic}$	list of chronic diseases of the user
L_{c_care}	list of additional medical conditions and healthcare needs selected and entered by the user
$L_{c\_current}$	list of current medical conditions of the user
L_{d_care}	list of diseases selected and entered by the user
$L_{d\_current}$	list of current diseases of the user
L_{s_care}	list of symptoms selected and entered by the user
$L_{s\_possible}$	list of symptoms displayed on the symptom Web page
L_t	list of topics of concern by the user
M, $M1$, $M2$	medical condition
N	a constant to control the amount of time spent on search result diversification

N_a	the total number of distinct HNAs appearing in the collection C of crawled Web pages
$n_a(P)$	the length of the Web page P measured in the HNA semantic unit
n_w_A	normalized weight of the nursing activity A
n_w_D	normalized weight of the nursing diagnosis D
n_w_I	normalized weight of the nursing intervention I
n_w_M	normalized weight of the medical condition M
n_w_T	normalized weight of the topic T
O_T	the search guide information compiled for the topic T from sources other than nursing knowledge
p	HMP
P, P', $P1$, $P2$	Web page
$\|P\|$	the length of the Web page P in the number of terms
P_{l_d}	the Web page in $S_{remaining}$ with the largest diversity score
q	query term
Q	query
Q_c	conceptual query representing the user's need
r	the uniform rate at which HNAs occur in all Web pages
R_{all}	the complete set of retrieved HMP Web pages
R_M	the medical condition M's search guide information compiled using nursing knowledge
s	symptom
s_A	HNA weight discount factor
S_A	the set of phrases pre-compiled for the HNA A
$score_P$	the page P's relevance score
S_{d_p}	the set of phrases pre-compiled for the disease d
s_D	nursing diagnosis weight discount factor
S_D	the set of nursing interventions linked to the nursing diagnosis D
s_I	nursing intervention weight discount factor
S_I	the set of nursing activities included in the nursing intervention I
S_M	the set of nursing diagnoses linked to the medical condition M
$S_{remaining}$	the set of Web pages remaining to be returned to the user
$S_{returned}$	the set of Web pages already returned to the user
S_{s_p}	the set of phrases pre-compiled for the symptom s
s_T	topic weight discount factor
t	term
T	topic
u	a predetermined constant used in language modeling with Dirichlet smoothing
w_A	weight of the nursing activity A
w_D	weight of the nursing diagnosis D
W_f	the set of Web pages retrieved for the phrase f
w_H	a predetermined constant
w_I	weight of the nursing intervention I
w_M	weight of the medical condition M
w_T	weight of the topic T