On Search Guide Phrase Compilation for Recommending Home Medical Products

Gang Luo
IBM T.J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, USA
luog@us.ibm.com

***Abstract* — To help people find desired home medical products (HMPs), we developed an intelligent personal health record (iPHR) system that can automatically recommend HMPs based on users' health issues. Using nursing knowledge, we pre-compile a set of "search guide" phrases that provides semantic translation from words describing health issues to their underlying medical meanings. Then iPHR automatically generates queries from those phrases and uses them and a search engine to retrieve HMPs. To avoid missing relevant HMPs during retrieval, the compiled search guide phrases need to be comprehensive. Such compilation is a challenging task because nursing knowledge updates frequently and contains numerous details scattered in many sources. This paper presents a semi-automatic tool facilitating such compilation. Our idea is to formulate the phrase compilation task as a multi-label classification problem. For each newly obtained search guide phrase, we first use nursing knowledge and information retrieval techniques to identify a small set of potentially relevant classes with corresponding hints. Then a nurse makes the final decision on assigning this phrase to proper classes based on those hints. We demonstrate the effectiveness of our techniques by compiling search guide phrases from an occupational therapy textbook.**

I. INTRODUCTION

About 50% of Americans have one or more chronic conditions, and most of these people need home medical products (HMPs) to facilitate their daily activities of living [1], [14]. However, as explained in [2], consumers often are unaware of the HMPs that can help them and have difficulty in finding those HMPs for multiple reasons. For example, consumers typically have little medical knowledge and cannot come up with proper keywords to search HMP catalogs. Many physicians also are not fully informed of available HMPs because they receive little training on HMPs.

To address this problem, we recently built an intelligent personal health record (iPHR) system that can automatically recommend HMPs based on users' health issues [2], [3]. Our main idea is to introduce expert system technology and Web search technology into Web-based personal health records. This will allow search engine queries to be automatically generated using medical knowledge.

More specifically, for each health issue, iPHR stores a list of "search guide" phrases pre-compiled using disease/symptom treatment knowledge and nursing knowledge. Those phrases serve to bridge the semantic gap between the literal meaning and the underlying medical meaning of the health issue. For example, the phrase *grab bar* is pre-compiled for *Alzheimer's disease*, because Alzheimer's patients tend to need grab bars to reduce their risk of falling.

To recommend HMPs to a user, first iPHR automatically extracts a preliminary set of health issues (e.g., diseases, symptoms, surgeries) from the user's personal health record. The user can revise this set according to his preferences. Then iPHR uses a search engine and each of the search guide phrases pre-compiled for these health issues as a query to retrieve some relevant HMPs. As a last step, iPHR combines all retrieved HMPs together and returns them to the user.

To avoid missing relevant HMPs during retrieval, the compiled search guide phrases need to be comprehensive. Such compilation is a challenging task because it uses a lot of detailed nursing knowledge. The scope of nursing is extensive and covers almost every aspect of healthcare. Consequently, nursing knowledge contains numerous details that are scattered throughout many sources. These details are useful for compiling more search guide phrases, but extremely labor-intensive to handle. To make matters worse, the phrase compilation task needs to be performed continuously as nursing knowledge updates frequently.

This paper presents a semi-automatic tool to expedite the compilation of search guide phrases. We notice that the essence of the phrase compilation task is a multi-label classification problem of assigning each newly obtained search guide phrase to one or more proper nursing diagnoses. This assignment process is time-consuming due to the large search space [4]. There are many nursing diagnoses, each of which contains a large amount of detailed information on both itself and its associated nursing activities. To determine whether a search guide phrase should be assigned to a nursing diagnosis, we need to check the detailed information associated with that nursing diagnosis.

To expedite the assignment process, we use some automatic method to quickly and dramatically reduce the search space size. This leads to significant savings of human effort on checking the original large search space. More specifically, for each newly obtained search guide phrase, we first use nursing knowledge and information retrieval techniques to automatically identify a small set of potentially relevant nursing diagnoses. From the detailed information associated with each of these nursing diagnoses, we automatically identify some potentially relevant information serving as hints. A nurse would then review those hints to determine which nursing diagnoses this phrase should be assigned to.

We implemented our techniques in a prototype tool for compiling search guide phrases. The effectiveness of this tool is evaluated by compiling search guide phrases from a popular occupational therapy textbook [5]. Our experiments show that this tool significantly improves the efficiency of the phrase compilation process.

In related work, to help nurses make nursing diagnoses, Gordon classifies all existing nursing diagnoses into eleven categories according to their functional health pattern areas [6]. Gordon's idea is to reduce cognitive strain and diagnostic errors by using consistency between organization of clinical assessment data and grouping of diagnostic categories. That approach does not apply to our case of compiling search guide phrases, because clinical assessment data is not available.

The rest of the paper is organized as follows. Section II provides some background on nursing diagnoses. Section III presents our tool for compiling search guide phrases. Section IV evaluates our tool.

II. BACKGROUND ON NURSING DIAGNOSES

In this section, we provide some background on the way iPHR uses nursing diagnoses. More details on this subject are available in [3], [7].

The North American Nursing Diagnosis Association International (NANDA-I) defines 188 *nursing diagnoses* that are clinical judgments about individual, family, or community responses to actual or potential health problems [8]. Each nursing diagnosis links to a list of *nursing activities* representing the actions that nurses, patients, and caregivers can take. In some nursing books [8], [9], this linkage is direct. In others [10], this linkage is made indirect through the use of nursing interventions. This paper uses direct linkage. (The HMP recommendation method described in [3] uses indirect linkage and can be easily modified to use direct linkage instead.)

For each *home nursing activity* (HNA) that patients and caregivers can perform at home or in the community, a nurse pre-compiles a set of search guide phrases that are stored in iPHR's knowledge base. Each health issue links to one or more nursing diagnoses [8], [9]. Using nursing diagnoses and HNAs as intermediate steps, we can link each health issue to multiple search guide phrases compiled using nursing knowledge, as shown in Fig. 1. Moreover, using the concepts of virtual nursing diagnosis and virtual HNA, we can link health issues to search guide phrases compiled from sources other than nursing knowledge (e.g., from treatment knowledge) [3].

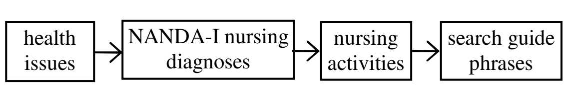

Fig. 1. Linking health issues to search guide phrases.

In a typical nursing diagnosis textbook [8], each nursing diagnosis contains the following parts in its description: definition, defining characteristics in terms of symptoms and signs, related/risk factors, and corresponding nursing activities with rationales on why they are suitable for this nursing diagnosis. Although the part on corresponding nursing activities with rationales is often several pages long, it still omits a lot of useful information. A nursing activity usually has many implementation details, such as the HMPs suitable for it. All of these details are useful for compiling search guide phrases. For instance, many search guide phrases are HMP names. Nevertheless, due to space constraints, each nursing diagnosis textbook can cover only a small portion of these details, typically in the form of a summary sentence that includes a few sample examples (e.g., HMP names, exercise names). The other missing details are scattered throughout numerous sources, including nursing care planning guides, journal articles, and textbooks in various areas (e.g., nursing specialty areas, occupational therapy, rehabilitation therapy, and speech therapy).

At present, iPHR's knowledge base includes the search guide phrases that a nurse compiled from the content of a standard nursing diagnosis textbook [8]. Although this compilation serves as a good starting point, it misses many search guide phrases not mentioned in the textbook. This significantly limits the number of relevant HMPs that iPHR can recommend.

To address this problem, we need to compile a more complete set of search guide phrases. This phrase compilation task contains two sub-tasks: phrase extraction and category assignment. In the phrase extraction sub-task, search guide phrases are extracted from the many sources (e.g., occupational therapy textbooks) that contain the missing details of the nursing activities. This sub-task requires much human effort on reading the contents of those sources, but is relatively easy to perform. In the category assignment sub-task, these extracted search guide phrases are assigned to proper nursing diagnoses so that iPHR can use these phrases in its existing HMP recommendation framework [3]. Compared to the phrase extraction sub-task, the category assignment sub-task is both more difficult to perform and much more time-consuming. The reason is as follows.

Consider a newly extracted search guide phrase f. Assigning f to one or more proper nursing diagnoses is a multi-label classification problem. It requires very high classification accuracy, and thus cannot be solved by existing automatic classification methods [11]. The challenge of this problem resides in the large search space. The number of classes (nursing diagnoses) is large: 188. For each nursing diagnosis D, we need all parts of its description to determine whether f should be assigned to it. For example, if f is the name of a HMP that can be used for a particular purpose, and the rationale for a nursing activity corresponding to D mentions that purpose, then f should be assigned to D.

A typical nursing diagnosis textbook is about 1,000 pages thick [8], [9]. No nurse can remember everything in the descriptions of all 188 nursing diagnoses [4]. As a result, for each newly extracted search guide phrase, a nurse may need to review the entire nursing diagnosis textbook [8] to perform the category assignment sub-task. This is not only time-consuming but also error-prone. The nurse can easily miss proper nursing diagnoses due to omissions during the reading.

To make matters worse, the phrase compilation task needs to be performed continuously. This is because nursing knowledge updates frequently. Also, each year many new HMPs enter the market as medical knowledge and technology improve. Since many search guide phrases are HMP names, this introduces a large number of new search guide phrases not covered in existing medical and nursing literature.

In the category assignment sub-task, we encounter difficulty only with search guide phrases compiled using nursing knowledge. The search guide phrases compiled from sources other than nursing knowledge can be directly assigned to the corresponding virtual nursing diagnoses [3]. In the rest of the paper, we focus on search guide phrases compiled using nursing knowledge.

III. A TOOL FOR COMPILING SEARCH GUIDE PHRASES

Without any external help, the phrase compilation task requires an extreme amount of human effort and is unmanageable. To make this task feasible, we have to build a tool to facilitate it. Ideally, that tool should satisfy the following three conditions. First, it can significantly reduce the amount of human effort required to perform this task. Second, it can reduce the number of errors that the human labeler makes by missing proper nursing diagnoses that a newly extracted search guide phrase should be assigned to. Third, it is reusable to fit the requirement that this task needs to be performed continuously.

Since compiling search guide phrases requires understanding human language, it is unlikely that such a tool can be completely automatic. Nevertheless, a semi-automatic tool is feasible. In the rest of the paper, we focus on the category assignment sub-task, which is the most time-consuming part of the phrase compilation task. For this sub-task, our tool uses nursing knowledge and information retrieval techniques to quickly, automatically, and dramatically reduce the search space size. In the reduced search space, a nurse makes the final decision on assigning the search guide phrase to proper nursing diagnoses. The phrase compilation task becomes manageable if the reduced search space is small enough for a human to handle manually. Moreover, to reduce the number of errors that the nurse makes by missing proper nursing diagnoses, our tool provides hints to facilitate the assignment process.

At a high level, our tool works as follows. Consider a search guide phrase f newly extracted from some source. For example, f can be a HMP name, whereas the source can be an occupational therapy textbook. Usually the source provides some sample usage scenarios of f for the reader to understand why f is relevant. These usage scenarios fall into the following two categories: (1) f can be used to directly deal with one or more health issues, and (2) f can be used for one or more other health-related purposes that indirectly address health issues. Two such purposes are positioning patients and clearing mucus. Either category of usage scenarios is used to obtain a reduced search space. Our tool will then compute the union of these two spaces as the final reduced search space, from which the nurse makes decision on assigning f to proper nursing diagnoses.

Next, we describe how to obtain a reduced search space for either of the two categories of usage scenarios.

A. ***Category 1****: The search guide phrase f can be used to directly deal with one or more health issues H_i.*

For each health issue $h \in H_i$, we find the set S_h of nursing diagnoses that it links to [8], [9]. As reviewed in Section II, this linkage information is available in iPHR's knowledge base [3]. Since the search guide phrase f can be used to deal with h, f should be assigned to one or more nursing diagnoses in S_h. The union of these sets of nursing diagnoses, $S_f = \bigcup_{h \in H_i} S_h$, becomes the reduced search space for the first category of usage scenarios. The nurse needs to check each nursing diagnosis $D \in S_f$ to see whether f should be assigned to it.

Note that when H_i contains more than one health issues, the intersection of these sets S_h ($h \in H_i$) of nursing diagnoses may not include the nursing diagnoses that the search guide phrase f should be assigned to. Hence, the above set union operation cannot be replaced by a set intersection operation. For example, consider two health issues $h1 \in H_i$ and $h2 \in H_i$. f can be used for a nursing activity $A1$ corresponding to a nursing diagnosis $D1$ that $h1$ links to. Also, f can be used for another nursing activity $A2$ corresponding to another nursing diagnosis $D2$ that $h2$ links to. $A1$ can be different from $A2$, and $D1$ can be different from $D2$. In this case, the intersection of the sets S_{h1} and S_{h2} may be empty.

B. ***Category 2****: The search guide phrase f can be used for one or more other health-related purposes S_p.*

The purposes in S_p can appear in the following parts of the description of a nursing diagnosis: definition, defining characteristics, related/risk factors, and corresponding nursing activities with rationales. For each purpose $p \in S_p$, it is unrealistic for the nurse to review all parts of the descriptions of all 188 nursing diagnoses to find the nursing diagnoses that the search guide phrase f should be assigned to. Instead, we treat p as a query and use information retrieval techniques to quickly reduce the search space size. If necessary, the nurse can come up with p's synonyms that are also used as queries.

The concrete method is as follows. We treat all parts of the description of a nursing diagnosis as a document. All 188 nursing diagnoses form a document set. In this document set, we perform Porter stemming [12] to reduce multiple words derived from a root word to the same form so that they can match with each other, drop stopwords using the SMART stopword list [13], and build an inverted index [3]. We perform stopwording and stemming on each purpose $p \in S_p$. Then we use the techniques in [3] to treat p as a sentence level Boolean conjunctive query to search this document set. In this way, we can find all occurrences of every purpose in S_p in this document set.

In most cases, the number n of such occurrences is small, e.g., $n=4$. All of these occurrences are highlighted in the sentences containing them. All such sentences and their corresponding nursing diagnoses become the reduced search space for the second category of usage scenarios. (If such a sentence belongs to the description of a nursing activity with corresponding rationale, we use the entire description rather than only this sentence to help the user understand the context.) In this reduced search space, the number of nursing diagnoses cannot be larger than n, and hence is typically small. The sentences serve as contexts and the highlights serve as hints. The nurse needs to check each nursing diagnosis in this reduced search space to determine whether the search guide phrase f should be assigned to that nursing diagnosis. This usually can be done quickly, as the nurse only needs to check a small number of hints and briefly scan their contexts to make her decision. Moreover, these hints can reduce the likelihood that the nurse misses proper nursing diagnoses due to omissions during the reading.

In practice, the number of all possible health-related purposes is large. Some purposes are frequently encountered, whereas others are not. For a frequently encountered purpose p, many search guide phrases can be used for it, whereas not all sentences containing it are actually related to it. For example, consider the symptom of *weakness*. The defining characteristics of the nursing diagnosis *defensive coping* include denial of obvious weaknesses, which has nothing to do with the symptom of *weakness*. In this case, the nurse will conduct much redundant work unnecessarily if she has to check all unrelated hints and contexts retrieved by p each time she runs into a newly extracted search guide phrase that can be used for p.

To address this problem, our tool allows the user to specify frequently encountered purposes. In the query results retrieved by such a purpose p, the user can mark which sentences containing p are unrelated to p. From then on, these sentences are excluded from the query results retrieved by p in constructing the reduced search space. In this way, the user can avoid repeatedly checking these sentences when she later encounters more search guide phrases that can be used for p.

For a nursing activity appearing in the description of a nursing diagnosis, the corresponding rationale typically mentions one or more purposes to show why this nursing activity is suitable for that nursing diagnosis. When the nurse originally compiled search guide phrases from the content of the nursing diagnosis textbook [8], some but not all of these purposes were compiled as search guide phrases [3]. The reader might wonder why not compile all of these purposes as search guide phrases. Intuitively, if all purposes S_p appear in the content of the nursing diagnosis textbook [8] and have already been compiled as search guide phrases, then those phrases may be able to retrieve all HMPs that can be retrieved by the phrase f, and hence there is no need to compile f as a search guide phrase any more. For example, suppose f is a HMP name. Each Web page about such a HMP is likely to mention (some of) the purposes S_p. If this is always the case, then the HMP Web pages retrieved by S_p should include the HMP Web pages retrieved by f.

The above idea sounds plausible, but does not work in practice. This is because to ensure the quality of HMPs recommended by iPHR, we need to choose proper search guide phrases to avoid retrieving irrelevant HMPs as much as possible. The search guide phrase f is usually more specific than the purposes S_p. Consequently, it is often the case that S_p retrieve many irrelevant HMPs unsuitable for S_p, whereas f retrieves mostly relevant HMPs.

For example, suppose the newly extracted search guide phrase f is *bed wedge pillow*. The corresponding purpose is to *position* patients, which is suitable for the nursing diagnosis *ineffective airway clearance*. The keyword *position* will retrieve many irrelevant HMPs that cannot be used for positioning patients. One example is the illiterate plastic eye chart (see http://www.allegromedical.com/diagnostic-products-c521/illiterate-plastic-eye-chart-p193147.html). Its corresponding Web page's description contains the following sentence: "Simply *position* the patient 20 or 10 feet in front of the eye chart and test their vision." In contrast, the phrase *bed wedge pillow* almost only retrieves bed wedge pillow Web pages.

C. Pre-processing of the Nursing Diagnosis Textbook Content

The above discussion assumes that the inverted index for the document set is built using all available content in the nursing diagnosis textbook [8]. This approach is simple, but not optimal for the following reason. In the second category of usage scenarios, the search guide phrase f can be used for one or more health-related purposes. The description of each nursing diagnosis typically contains much material that cannot match with these purposes in a useful way in constructing the reduced search space. This is especially true in the "nursing activities with rationales" section of the description. Nevertheless, the words appearing in this "useless" material can accidentally match with the words appearing in these purposes. Consequently, keeping this "useless" material in the inverted index will unnecessarily increase the size of the reduced search space, and hence slow down the completion of the category assignment sub-task.

For example, the rationale for a nursing activity corresponding to the nursing diagnosis *risk for impaired parent/child attachment* contains the following sentence: "Neonatal nurses are in a unique *position* to support families' competence and confidence in caring for their infants at their own pace and to encourage the developing mother-infant relationship." Many HMPs (e.g., bed wedge pillow) are used to position patients. Although the word *position* appears in the above sentence, the search guide phrase(s) that can be used for positioning purpose should not be assigned to the nursing diagnosis *risk for impaired parent/child attachment*.

To avoid unnecessarily increasing the size of the reduced search space, a nurse pre-marks all "useless" material in the content of the nursing diagnosis textbook [8]. This material can appear as contexts for generated hints. However, it is ignored when building the inverted index for the document set, and hence does not contribute to highlights. In this way, the user of our tool can still see complete contexts. At the same time, many useless hints are excluded from the reduced search space as they are no longer generated. This pre-marking takes some time, but is worth the effort, because it needs to be performed only once, whereas the category assignment sub-task needs to be performed many times from time to time.

D. User Interface

The input interface of our tool contains two groups of textboxes for the user (usually a nurse) to input information about the usage scenarios of the newly extracted search guide phrase *f*. The first group of textboxes is for the health issues H_i in the first category of usage scenarios, one health issue per textbox. The second group of textboxes is for the health-related purposes S_p in the second category of usage scenarios, one purpose per textbox.

The output interface of our tool displays the nursing diagnoses in the final search space, which is the union of the reduced search spaces for the two categories of usage scenarios. For each such nursing diagnosis, we display the associated hints highlighted in their contexts. Also, the user can click the nursing diagnosis to see its detailed description.

IV. EXPERIMENTAL RESULTS

We implemented a prototype tool for compiling search guide phrases. To demonstrate the effectiveness of our techniques, a nurse used our tool to compile search guide phrases from a popular occupational therapy textbook [5]. We choose the area of occupational therapy because its goal is to enable people with limitations or impairments to perform daily activities of living. This goal coincides with iPHR's goal of facilitating people's daily activities by recommending proper HMPs. Moreover, the textbook [5] describes a large number of HMPs and their usage scenarios, many of which are not mentioned in the nursing diagnosis textbook [8]. Many new search guide phrases could be compiled from the textbook [5] to complement the existing search guide phrases in iPHR's knowledge base.

The nurse extracted about 300 new search guide phrases from the occupational therapy textbook [5]. On average, for each new phrase our tool provides eight nursing diagnoses in the final reduced search space. The content of this space can be printed on two pages. The nurse spent four minutes on reviewing this content and assigning the phrase to proper nursing diagnoses. Recall that the original search space consisted of 188 nursing diagnoses and its content contained about one thousand pages. The reduced search space is much smaller than the original search space. Without our tool, the nurse stated that it would be too challenging for her to finish this phrase compilation task in any reasonable amount of time.

To give the reader a concrete feeling of the output provided by our tool, we present detailed results for the search guide phrase *playing card holder* (see http://www.allegromedical.com/daily-living-aids-c519/playing-card-holder-p192636.html). This phrase can be used to deal with the symptom of poor coordination [5, page 797]. Using the word *coordination* as a query, our tool produces the following output as the final reduced search space:

Nursing diagnosis 1: Risk for falls
Avoid restraints if at all possible ... Restrained elderly clients often have an increased number of falls, possibly as a result of muscle deconditioning or loss of ***coordination*** ...

Nursing diagnosis 2: Health-seeking behaviors
Alcoholic beverages should be avoided by individuals engaging in activities that require attention, skill, or ***coordination*** ...

Nursing diagnosis 3: Impaired physical mobility
Defining characteristics: ... ***uncoordinated*** movements ...
Techniques such as gait training, strength training, and exercise to improve balance and ***coordination*** can be very helpful for rehabilitating clients ...

Nursing diagnosis 4: Noncompliance
If the client has sensory and ***coordination*** deficits, use a medication organizer ...

Nursing diagnosis 5: Risk for suffocation
Stress water and pool safety precautions ... A child's high center of gravity and poor ***coordination*** make buckets and toilets a threat ...

Nursing diagnosis 6: Risk for trauma
... reduced muscle ***coordination*** ...

Nursing diagnosis 7: Disorganized infant behavior
... ***uncoordinated*** movement ...

Nursing diagnosis 8: Ineffective infant feeding pattern
Impaired ability of an infant to suck or ***coordinate*** the suck/swallow response ...

Nursing diagnosis 9: Impaired swallowing
Watch for ***uncoordinated*** chewing or swallowing ...

After reading the hints in the output, the nurse assigned this phrase to the nursing diagnosis *impaired physical mobility*.

ACKNOWLEDGMENT

We thank Selena B. Thomas, Chunqiang Tang, and Jing Wang for helpful discussions.

REFERENCES

[1] *U.S. Market for Home Care Products*, 5th ed., Kalorama Information, 2007.

[2] G. Luo, S. B. Thomas, and C. Tang, "Intelligent consumer-centric electronic medical record," in *Proc. MIE'09*, 2009, pp. 120-124.

[3] G. Luo, S. B. Thomas, and C. Tang, "Automatic home medical product recommendation," *JMS*, to be published, [Online]. Available: http://pages.cs.wisc.edu/~gangluo/device.pdf.

[4] NANDA International. *Nursing Diagnoses, 2009-11 Edition: Definitions and Classification*, 2nd ed., Wiley-Blackwell, 2008.

[5] M. V. Radomski and C. A. Trombly, *Occupational Therapy for Physical Dysfunction*, 6th ed., Lippincott Williams & Wilkins, 2007.

[6] M. Gordon, *Manual of Nursing Diagnosis*, 12th ed., Jones & Bartlett Publishers, 2009.

[7] G. Luo and C. Tang, "Automatic home nursing activity recommendation," in *Proc. AMIA'09,* 2009, pp. 401-405.

[8] B. J. Ackley and G. B. Ladwig, *Nursing Diagnosis Handbook: An Evidence-Based Guide to Planning Care*, 8th ed., Mosby, 2007.

[9] M. E. Doenges, *Nursing Diagnosis Manual: Planning, Individualizing & Documenting Client Care*, 3rd ed., F.A. Davis Company, 2009.

[10] G. M. Bulechek, H. K. Butcher, and J. M. Dochterman, *Nursing Interventions Classification (NIC)*, 5th ed., Mosby, 2007.

[11] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *Proc. CIKM'05,* 2005, pp. 195-200.

[12] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130-137, 1980.

[13] (2009) SMART stopword list. [Online]. Available: http://www.lextek.com/manuals/onix/stopwords2.html.

[14] (2010) Hidola homepage. [Online]. Available: http://www.hidola.com.