Providing Multilingual Access to Health-Related Content

Till PLUMBAUM[a], Sascha NARR[a], Elif ERYILMAZ[a], Frank HOPFGARTNER[a],
Funda KLEIN-ELLINGHAUS[b], Anna REESE[b], and Sahin ALBAYRAK [a,c]

[a] *Technische Universität Berlin, Berlin, Germany*
[b] *Leibniz Institute for Prevention Research and Epidemiology, Bremen, Germany*
[c] *German-Turkish Advanced ICT Research Centre, Berlin, Germany*

Abstract. Finding health-related content is not an easy task. People have to know what to search for, which medical terms to use, and where to find accurate information. This task becomes even harder when people such as immigrants wish to find information in their country of residence and do not speak the national language very well. In this paper, we present a new health information system that allows users to search for health information using natural language queries composed of multiple languages. We present the technical details of the system and outline the results of a preliminary user study to demonstrate the usability of the system.

Keywords. Semantic search, prevention service, multilingual querying, health service

Introduction

Nowadays, more and more people suffer from lifestyle illnesses such as coronary heart disease, obesity or type-2 diabetes, i.e., health conditions that are caused by lack of exercise, poor eating habits and unhealthy lifestyles [7]. Woolf and Atkins [8] argue that many health conditions can be prevented though, for example by educating the people about the consequences of an unhealthy lifestyle. Large education campaigns and seminars have been funded to promote healthier lifestyles, advertising low-fat diets, weight management or physical exercises. With the increasing availability of the Internet, health care providers and governments extended the education campaign to the WWW, offering health information portals with educational texts, videos, graphical illustrations and other type of media. Although health care providers intend to provide health information services to all their clients, immigrants have been identified as vulnerable population [1] that benefit less from existing health care systems since language and cultural barriers prevent them from using existing prevention services. In order to address the issues mentioned above, we introduce a multilingual health prevention service for immigrants that guarantees personalized access to professionally-created health care content. By providing them with available information in the language of their choice, the system helps those people who have language-related difficulties in understanding their physicians. Information is adapted based on users' personal context such as pre-existing medical conditions and their location.

This paper is structured as follows. First, we discuss the use of online health information services. Following that, we introduce the system infrastructure of our system and outline the scientific challenges that it addresses. We then present an initial evaluation of the system. A discussion and conclusion is given in the last section.

Related Work

According to Morahan-Martin [6], up to 4.5% of all Internet searches are about health-related topics, indicating the significance of this topic in our life. An overview of different resources in the Web that can be used for this information gathering task is provided by Johnson et al. [3]. Generally, three types of services can be identified: (a) Professionally maintained health advice and information services where users can check their symptoms in a constantly updated database (e.g., the NHS Direct[1] service maintained by the English National Health Service) or check up on public service announcements (e.g., by the World Health Organization), (b) unsupervised sources such as Wikipedia and (c) discussions of similar cases in online forums, blog posts, or biased advertisements for specific products that can be retrieved by standard Web search engines. From a medical point of view, relying on such sources is not recommended and the consultation of a professional is highly advised. Therefore, the American Medical Library Association recommends to *"trust your physician, not a chat room"* [2]. Morahan-Martin suggests approaching this unconsidered information handling by asking physicians to point their patients to reliable health portals and to work on improving such sites, e.g., by improving search and retrieval techniques. Our works builds on her suggestion.

1. Methods

The next sections describe the main components of the health information system. The presented system builds on semantic technologies to perform the task of multilingual information supply.

1.1. The Health Ontology

In order to provide a knowledge base that can easily be maintained by a computer system, we define a simple health ontology (HO) that defines basic health concepts and their relations. The concepts in the Health Ontology are enriched with multilingual labels, e.g., the concept *obesity* is labelled with the English term "obesity", the Turkish term "Obezite" and the German expression "Fettleibigkeit". Besides, relevant documents are attached to the concept nodes. In a health care scenario, information quality is of crucial importance. False information does not only lead to a loss of trust but also may lead to serious harms. Therefore, information in the HO should be maintained by a group of experts.

---

[1] http://www.nhsdirect.nhs.uk/

1.2. Multilingual Semantic Search

As explained above, concepts in the health ontology are enriched with multilingual labels. When a user enters a search query, we process the entered text by lowercasing it and pruning unusual characters. The search terms are then used as a query to retrieve matching concept labels contained in the Health Ontology. We use a fuzzy search based on the open-source information retrieval system Lucene[2] to cover slight alterations in term surface forms. Taking the multi-lingual search query "Fettleibigkeit diyetimi nasıl ayarlamalıyım?" (Translation: "How should I arrange my diet in obesity?") as an example, the language-independent concepts General Obesity and Obesity are identified, since they have been labelled using the keyword "Fettleibigkeit". Furthermore, the concept *nutrition* is identified by the label "diyetimi".

Using the retrieved query-relevant ontology concept nodes, we employ a graph-search to find conceptually related information nodes. Different concepts from ontology classes are semantically linked with weighted edges in our ontology. Our algorithm performs a full graph search along these edges. Information nodes found during this traversal are ranked based on the proportion to the edge-weights of the path of the originating concept and anti-proportional to that path's length. Information nodes found via multiple paths receive the sum of the relevance values of all of those paths.

1.3. Supportive User Interface

Figure 1 shows a screenshot of the graphical user interface that allows users to interact with the system. In the remainder of this section, we introduce its key features.

Entering the website, the user can log in to receive context-based search results. On top of the interface, the user can type in a search query. Since the system matches search terms with concepts in the health ontology, the search query can be formulated in different languages. In the screenshot, the user Ali has logged in and typed in the search query used in the example above: *"Fettleibigkeit diyetimi nasıl ayarlamalıyım"* (Translation: *"How should I arrange my diet in obesity?"*) Under the search box, the interface lists the concepts that the system extracted from the search query. In the example screenshot, the following concepts have been detected from the search query: Genel obezite (General obesity), Obezite (obesity), Beslenme (nutrition). Furthermore, the following context has been extracted from Ali's profile that will be considered when ranking search results: home town (Berlin) and pre-existing medical conditions (Diyabet Tip II, Type II Diabetes). Under this visualization, the interface provides a topic box containing professionally edited information for the identified concepts. Under this topic box, the search results are displayed in descending order of relevance.

Each search result offers a button that when clicked shows a view of the paths of the ontology that were traversed to find this search result. The semantic graph allows the user to understand why this search result is relevant to their search query. Additionally, the interface enables users to visually browse through the ontology concepts by expanding concept nodes and traversing semantic edges between concepts, and also finding documents attached to a concept's information nodes.

---

[2] http://lucene.apache.org/

Figure 1. Screenshot of the search interface.

2. Results

Following a demand analysis to assess immigrants' expectations and needs for a personalized preventive health assistant [4], we evaluated the development and implementation of the health assistant. Applying a qualitative research approach, we recruited a small study cohort from a local Turkish community in a mid-sized town in Germany, as well as a German control group. As suggested by Lamnek [5], we invited a homogeneous focus group of six participants with similar socio-economic backgrounds from two different districts of the town to participate in the user evaluation. The participants were asked to interact with the system in a supervised scenario and were further asked to assess the assistant with respect to its technical implementation, relevance of the presented content, usability in daily life and its potential to present health information in a structured manner.

During the user-centric evaluation of the system, the study group was separated into a German-speaking group and a Turkish-speaking group due to limited knowledge of the German language of some participants. Both groups interactively tested the health assistant prototype in a supervised user-scenario. During the test, they were asked several questions concerning the design, usability and cultural sensitivity.

Participants found the search results to be appropriate and clearly presented. Most participants found the adaptation of search results based on the context of a personal user profile useful and found that this feature helped in making an individualized search experience. However, some users were reluctant in providing certain pieces of profile information and wished to know specifically how they would be used.

Performing semantic search was found to be intuitive by a large majority of users in the study. Many users entered full health related questions, as intended, without needing any instructions. The feature to enter mixed-language queries was found to be very useful by all participants, and was considered to be of high importance for everyday use by the group. However, most users needed to be made aware of this functionality, since it was not readily apparent from the interface itself.

3. Conclusion and Discussion

In this paper, we presented a personalized health information system for immigrants. The system provides a search facility on a semantic database to assist people in finding health-oriented documents. The system is mainly designed to assist immigrants with limited knowledge of the national language used in their host country. Therefore, the system allows users to formulate search queries in their mother tongue, the host's language and in a mix of both languages. Search results can be adapted based on the users' preferred language and other personal contexts such as the users' location or pre-existing medical conditions.

The underlying technology of this system is a health ontology. Relevant information is retrieved by exploiting semantic relations between different concepts in this ontology. In order to evaluate the usability of this system, we followed a qualitative analysis scheme. Discussions in a focus group following this evaluation indicate that the system can be employed to assist immigrants to find information in their own and in the host nations' language. This study has been performed within a project with a large national health insurance provider. As future work, a promising approach to evaluate the efficiency of this system on a large scale is to incorporate it into this company's health prevention service.

References

[1] Derose, K., Escarce, J., Lurie, N.: Immigrants and health care: Sources of vulnerability. Health Affairs pp. 1258–1268 (09 2005)
[2] Fox, S., Raine, L.: Vital decisions: How internet users decide what information to trust when they or their loved ones are sick. plus a guide from the medical library association about smart health-search strategies and good websites. (2002)
[3] Johnson, P.T., Chen, J.K., Eng, J., Makary, M.A., Fishman, E.K.: A comparison of world wide web resources for identifying medical information. Academic Radiology 15(9), 1165–1172 (2008)
[4] Klein-Ellinghaus, F., Zeeb, H., Reeske, A.: Bedarf und Erwartungen an einen innovativen, technologiebasierten Gesundheitsassistenten für türkische Migrantinnen und Migranten (GeM-Projekt). Gesundheitswesen 74(A56) (2012)
[5] Lamnek, S.: Qualitative Sozialforschung. Beltz Verlag (2005)
[6] Morahan-Martin, J.M.: How Internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsyc. and Beh. 7(5), 497–510 (2004)
[7] Pollard, T.M.: Western Diseases: An Evolutionary Perspective. CUP (2009)
[8] Woolf, S., Atkins, D.: The evolving role of prevention in health care. American Journal of Preventive Medicine pp. 13–20 (04 2001)