PMIR: A Personal Medical Information Recommender

Haridimos Kondylakis[a,1], Lefteris Koumakis[a], Stefan Rüping[b], Eleni Kazantzaki[a], Kostas Marias[a], Manolis Tsiknakis[a]
[a] *Computational Medicine Laboratory, FORTH-ICS, Heraklio, Crete, Greece*
[b] *Knowledge Discovery Group, Fraunhofer Institute IAIS*

Abstract. Patients today have ample opportunities to inform themselves in the internet about their disease and possible treatments. While this type of patient empowerment is widely regarded as having a positive influence on the treatment, there exists the problem that the quality of information that can be found on online is very diverse. This paper presents a platform which will empower patient with high quality knowledge about his/her condition, and will provide intelligent and personalized recommendations, according to his/her personalized preferences and medical conditions. We also introduce the EURECA project and its vision in the field of personalized medicine and show project's approach on creating a personal medical information recommender.

Keywords. Medical Information, Recommendations, Data Mining

Introduction

Medicine is undergoing a revolution that is transforming the nature of healthcare from reactive to preventive. The changes are catalysed by a new systems approach to disease which focuses on integrated diagnosis, treatment and prevention of disease in individuals. This will replace our current mode of medicine over the coming years with a personalized predictive treatment. While the goal is clear, the path is fraught with challenges. One of these challenges is the problem of the quality of information that can be found on online [1]. In addition, it may be very hard for a patient to accurately judge the relevance of some information to his own case.

The EURECA EU project aims to build an advanced, standards-based and scalable semantic integration environment enabling seamless, secure and consistent bi-directional linking of clinical research and clinical care systems which, among others, will empower patients to extract the relevant data out of the overwhelmingly large amounts of heterogeneous data and treatment information.

This paper focuses on current research activities related to the design and implementation of a Personal Medical Information Recommender (PMIR). PMIR is targeted at improving the opportunities that patients have to inform themselves in the internet about their disease and possible treatments, and providing to them personalized information and recommendations. Its goal is threefold: (1) to deliver relevant information to patients, based on their current situation as represented in their personal

[1] Corresponding Author.

healthcare record (PHR) data, (2) to ensure the quality of the presented information by giving doctors the chance to control the information that is given, and (3) to facilitate an easy uptake of the new system by minimizing the necessary manual effort. The PMIR system will be integrated into IndivoX [2] PHR as a set of applications.

Although similar approaches exist already, such as WebMD[2], MayoClinic Patient Care and Health Info[3] etc., they are not dynamically adapted according to patient's preferences of medical history, which is true in our case. The rest of this paper is structured as follows: Section 1 presents the PMIR architecture, describing the available components. Then, Section 2 concludes the paper and discusses future directions.

1. PMIR Architecture

The architecture of the system is shown on Figure 1. On the top layer, there is a PHR system through which the whole functionality of the PMIR is offered to the patient.

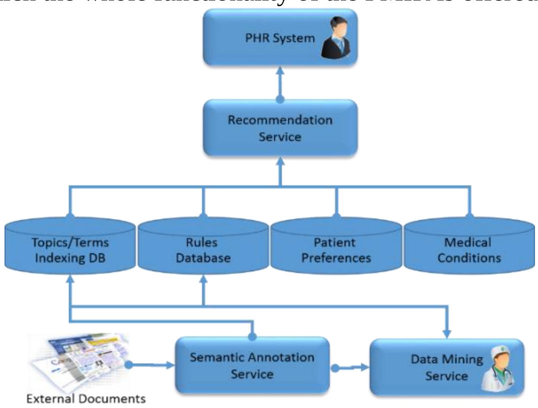

Figure 1. PMIR Architecture.

Using the PHR system, the patient, besides logging and reviewing his medical information, is able to search for relevant, high quality information using an intelligent search engine provided as an individual PHR app. The feedback presented to the user is provided by the recommendation service which considers the following modules to make the results of the query as personalized as possible:

a) *Topics/Terms Database*: This database indexes all web documents that are being searched for relevant information. Those web documents are high quality web resources (web pages, pdfs, docs etc.) selected carefully by the appropriate experts. If no relevant document is retrieved by this database then the results from HON[4] search are used instead.

b) *Patient Preferences Database*: This database contains user preferences that are acquired as the patient browses the results presented to him. His selections are logged and he is also able to rank by himself a result as relevant or irrelevant to augment similar future searches.

[2] http://www.webmd.com/

[3] http://www.mayoclinic.org/patient-care-and-health-information

[4] http://www.hon.ch/HONcode/

c) *Medical Conditions Database*: This database captures the medical conditions that the patient has logged in his PHR. Information here includes allergies, medications, problems, procedures and laboratory results.

d) *Rules Database*: This database includes rules that are used for filtering and ranking the results of the database. Besides ranking and filtering results, there exist also rules for providing automatically recommendations to the patients.

Both *Rules* and *Topics* are being generated by the Data Mining Service by processing semantically annotated external documents. Those documents can be selected by experts or can be the results of other search engines if there are no relevant documents already selected. Moreover, the Data Mining produces the rules for filtering and sorting the different documents.

In the following we give a short description of the PMIR building blocks.

1.1. Semantic Annotation

The semantic annotator parses related texts from external sources using Natural Language Processing (NLP) techniques [3] such as pre-processing algorithms for tokenization, lemmatization and part of speech. Then the platform communicates with the Bioportal [4] annotator and extracts ontology terms, concepts and semantic types for a set of medical ontologies. Using the ontology terms and its semantic types we get annotated documents which feed the data mining services and generate a topic model and the corresponding map for the annotations.

1.2. Data Mining Service

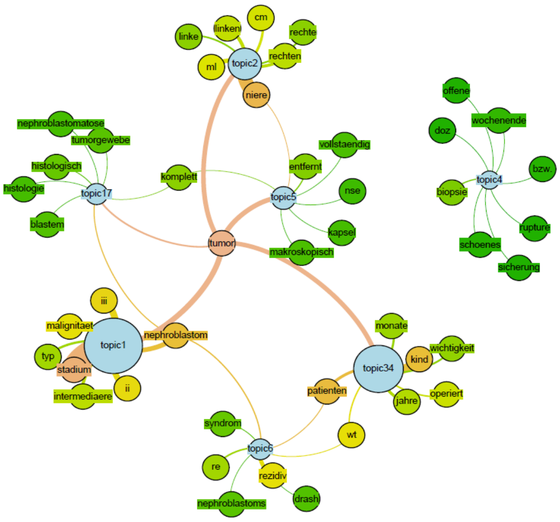

Figure 2. Small excerpt from the Topic Map (in German – labels represent topics/subtopics identified)

Data mining approaches allow the PMIR platform to scale beyond the complexity that is possible with a purely manual approach of content generation. A core problem for personalized information recommendation, based on both structured patient data and unstructured text content, is that appropriate meta-data has to be generated for the text content. While the semantic annotation of texts based on ontology terms is an interesting approach to identify already known concepts, this approach has limited capability to identify new, unknown concepts. For this reason, as an additional source of information, topic maps are used. The topic models, which are based on the LDA algorithm [5], provide an intuitive overview over important discussion topics in the texts which are extracted directly from the texts itself. The advantage of this approach is that it not only allows to generate interesting meta-data (by assigning to each text its topic membership degrees), but also allows a clinician or any other interested party to get a very quick overview over the content which is recommended to patients without having to read all available documents (an example is shown in Figure 2). This is a key issue to monitor the quality of the recommended information and keeping track of the patient's interests without an excessive demand on precious expert's time.

1.3. Recommendation Service

The recommendation service is the bridge between the knowledge from the data mining services and the patient's information and preferences. This service is built over rules based on the metadata, information from the PHR, and the recorded patient feedback.

The service also supports the scoring mechanism for the retrieved information. The content that has been identified as relevant is scored according to its dynamic relevance to the patient. In addition to the objective relevance, the system takes into account subjective information such as the content which the patient has already seen, or the type of content the patient prefers (articles for the general public or specialist information, text or image, etc). The patient-specific relevance is learned based on patient feedback. The recommendation rules on the other hand are automatically learned from data and optionally be evaluated from experts.

1.4. Personal Health Record

IndivoX, the EURECA's PHR system, is an open-source personal health platform, enabling an individual to own and manage a complete, secure, digital copy of her health and wellness information. It has been extended to include the PMIR search engine and to respond to the different recommendations provided by the recommendation service. This covers a wide range of techniques such as: *adaptive presentation* which deals with the content presentation in a manner that best suits individual users' needs, *adaptive navigation* which comprises all the ways to alter visible links to support hyperspace navigation and *intelligent alerts* generated through the recommendation services.

The extended PHR supplies feedback about the displayed information from the patient, either implicitly or explicitly, such that the relevance of the displayed information can be adopted over time. The process of adapting content to specific user needs can be thought of as two main sub processes. The first sub process involves understanding what content can be most relevant to the current user's interests, and how this content should be organized. The second sub process involves decision on how to effectively present the selected content to the user.

2. Conclusion and Discussion

The changing nature of information distribution due to the evolution of the web has important implications for health care. Given the wide use of the web in providing medical information, feeding patients with appropriate content might further enhance the patient's education and experience. The validity and the quality of the available healthcare information on the internet is an area of major concern mainly because these have not been well documented [6]. Although healthcare professionals should continue to strive to be the main source of information for patients, we should also be aware that most will continue to use the internet to gather information [7]. A recent publication from Berg et al [8] concluded that the optimal solution for patients is to be guided by healthcare providers to more optimal resources over the web. Delivering accurate sources to patient increases his knowledge and changes the way of thinking which is usually referred as patient empowerment. As a result, the patient's dependency for information from the doctor is reduced. Moreover, patients feel autonomous and more confident about the management of their disease [9].

PMIR platform focuses on making the available information timelier and more relevant with respect to dynamic influences in the individual patient's treatment. The idea is that even if two patients suffer from the same disease and they are in the same phase of the treatment, their interests on available information may differ based on various factors. For example, patients might have different medical background that is not directly related to the treatment, they might receive additional drugs due to other, independent treatments, or they might be affected by other external factors such as the weather in case of allergies. More subjective factors include for example patient preferences for more simple or more complex information. To the best of our knowledge no other system providing medical information is able to be dynamically adapted in such a diverse environment.

Acknowledgements

This work has been supported by eHealthmonitor, p-Medicine and EURECA projects and has been funded by the European Commission under contracts FP7-287509, FP7-270089 & FP7-288048.

References

[1] Berg GM, Hervey AM, Atterbury D, Cook R, Mosley M, Grundmeyer R, Acuna D. Evaluating the quality of online information about concussions. JAAPA. 2014;27(2):1-8.

[2] Adida B, Sanyal A, Zabak S, Kohane IS, Mandl KD. Indivo-X: developing a fully substitutable personally controlled health record platform. AMIA Annual Symposium Proceedings. 2010;13:6-10.

[3] Jurafsky D, Martin JH, Kehler A, Vander LK, Ward N. Speech and language processing: An introduction to natural language processing, Computational Linguistics, and Speech Recognition, 2000.

[4] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Musen MA. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research. 2009;37(2):170-173.

[5] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. The Journal of machine Learning research. 2003;3:993-1022.

[6] Agliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. BMJ. 2002;324:569–73.

[7] Scullard P, Peacock C, Davies P. Googling children's health: reliability of medical advice on the internet. Archives of disease in childhood. 2010;95(8):580-582.

[8] Berg GM, Hervey AM, Atterbury D, Cook R, Mosley M, Grundmeyer R, Acuna D. Evaluating the quality of online information about concussions. Journal of the American Academy of Physician Assistants. 2014;27(2):1-8.

[9] Wiesner M, Pfeifer D. Adapting recommender systems to the requirements of personal health record systems. In Proceedings of the 1st ACM International Health Informatics Symposium. 2010. 410-414.