

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

2014

WenZher: Comprehensive Vertical search for Healthcare Domain

Liqiang Nie

Tao Li

Mohammad Akbari

Jialie SHEN

Singapore Management University, jlshen@smu.edu.sg

Tat-Seng CHUA

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Databases and Information Systems Commons](#)

Citation

Nie, Liqiang; Li, Tao; Akbari, Mohammad; SHEN, Jialie; and CHUA, Tat-Seng. WenZher: Comprehensive Vertical search for Healthcare Domain. (2014). *SIGIR '14: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval: July 6-11, 2014, Gold Coast, Australia*. 1245-1246.

Available at: https://ink.library.smu.edu.sg/sis_research/2511

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylids@smu.edu.sg.

WenZher: Comprehensive Vertical Search for Healthcare Domain

Liqiang Nie[†], Tao Li[§], Mohammad Akbari[†], Jialie Shen[‡], and Tat-Seng Chua[†]
[†] National University of Singapore; [‡] Singapore Management University; [§] Zhejiang University
{nieliqiang, coylee917, jialie}@gmail.com; {akbari, chuats}@nus.edu.sg

ABSTRACT

Online health seeking has transformed the way of health knowledge exchange and reusability. The existing general and vertical health search engines, however, just routinely return lists of matched documents or question answer (QA) pairs, which may overwhelm the seekers or not sufficiently meet the seekers' expectations. Instead, our multilingual system is able to return one multi-faceted answer that is well-structured and precisely extracted from multiple heterogeneous healthcare sources. Further, should the seekers not be satisfied with the returned search results, our system can automatically route the unsolved questions to the professionals with relevant expertise.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval-Search Process

Keywords

Healthcare, Vertical Search, Structured Knowledge, Social Platform

1. INTRODUCTION

As health consumers take more control over managing their health, 72% of U.S. internet users have looked online for health information to improve their health, wellness and safety. This survey is conducted by the Pew Research Center in January 2013¹. It also reports that 77% of online health seekers begin their exploration at commercial search engines, such as Google, Bing and Yahoo. Another 13% explore special healthcare-oriented sites, such as WebMD and HealthTap. The remainders start search at social network or other non-mainstream retrieval forums. The seekers, however, are frequently overwhelmed by the returned vast quantity of unstructured information, i.e., lists of documents

¹<http://pewinternet.org/Reports/2013/Health-online.aspx>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.

ACM 978-1-4503-2257-7/14/07.

<http://dx.doi.org/10.1145/2600428.2611176>.

Table 1: Data types involved in our system.

Data Types	Sources
Health Provider Released Data	WebMD, Yahoo! Health, MedicineNet, etc.
Expert Generated Data	HealthTap, HaoDF, ChunYuHealth, etc.
Patient Generated Data	Patientslikeme, Manbingyou, etc.

or lists of QA pairs. Besides, the lists of answers are usually relevant redundancy rather than diverse richness, they hence may not meet the seekers' demands.

To overcome the information over-loading, unstructured and incomplete problems, this demonstration presents a novel system, named WenZher. It automatically organizes all the associated healthcare knowledge into a single view for a given question. The comprehensive answer is an analytical result of heterogeneous and multilingual data sources. These sources can be broadly categorized into health provider released data, expert generated data and patient generated data, as summarized in Table 1. WenZher is also a social platform and recommender system. It can propagate the questions among askers' connections and route the questions to the knowledgeable persons with relevant expertise.

2. SYSTEM ARCHITECTURE

The architecture of WenZher is illustrated in Figure 1, where the three knowledge bases play essential roles. To construct these bases, WenZher firstly collected approximately one thousand and half healthcare topics and their associated structured data from WebMD, Yahoo! Health and other forums. These topics cover a wide range, including pregnancy, cancer and endocrine. They are well-organized and freely-released by professional organizations. Regarding these topics as queries, WenZher crawled more than half million high-quality QA pairs from popular healthcare QA communities, such as HealthTap, HaoDF and Chunyu-isheng. The answers in these communities are generated by professionals that offer valuable references for the new coming similar questions. In addition, WenZher gathered more than 200 thousand patient generated e-records from Patientslikeme and Manbingyou, and clustered them in terms of healthcare topics. The statistical results of patient generated data reveal the insights of patients from the same group with similar conditions, and provide the health seekers more knowledge about their health concerns.

To generate comprehensive results, WenZher performs several high-order analytics. Here we briefly introduce them.

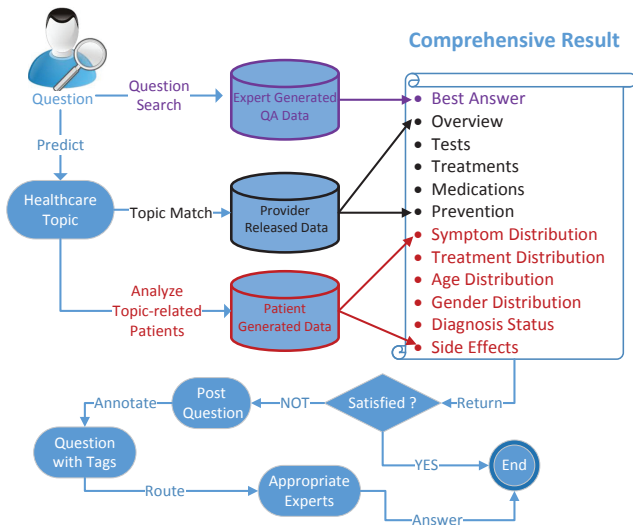


Figure 1: Illustrative Architecture of our system.

Best Answer Selection WenZher supports multilingual search, so it first identifies the type of question language. It then locates the similar language-aware question by utilizing the syntactic tree structure approach [2]. The expert crafted answers of the first positioned question are returned.

Healthcare Topic Prediction. To better capture the semantics of healthcare data and reduce the feature dimension, WenZher exploits the LDA-based topic-level features for data representation instead of traditional low-level n-gram features. Feeding the features into various classifiers, decision tree C4.5 achieves the best performance, slightly less than 90%, compared to Naïve Bayes, KNN and SVM.

Annotation with Terminologies Inspired by [1], WenZher codes each new question with medical terminologies that benefits cross-system operability and inter-user reusability. It first locally extracts the medical concepts from the question itself and maps them to terminologies in the external authenticated vocabularies, SNOMED-CT. Following that, it globally learns the terminologies from neighbors.

Question Routing WenZher models the matching relations between given questions and doctors from the expertise and attitude, respectively. It then adaptively fuses these two factors via regression models.

3. DEMONSTRATION

WenZher is a social platform. It supports unidirectional connections among patients and experts, which naturally forms the tightly linked communities in terms of similar healthcare concerns, habits and practices. The latest posted questions and answers can be promptly updated and fed within each community, so users can learn key knowledge and offer advice via their personalized pages.

Before adding a new question, WenZher encourages users to perform vertical search over archived question-answer pairs, which is a policy to constrain duplicate questions. Figure 2 illustrates a comprehensive answer page for a given question. This page contains the best answer selected from multiple QA communities, organized knowledge from health provider released authoritative sources, as well as statistics from patient crowdsourcing forums.

WenZher allows users to post new questions, as shown in Figure 3. Here media documents can be uploaded to intuitively complement the question descriptions. In addi-

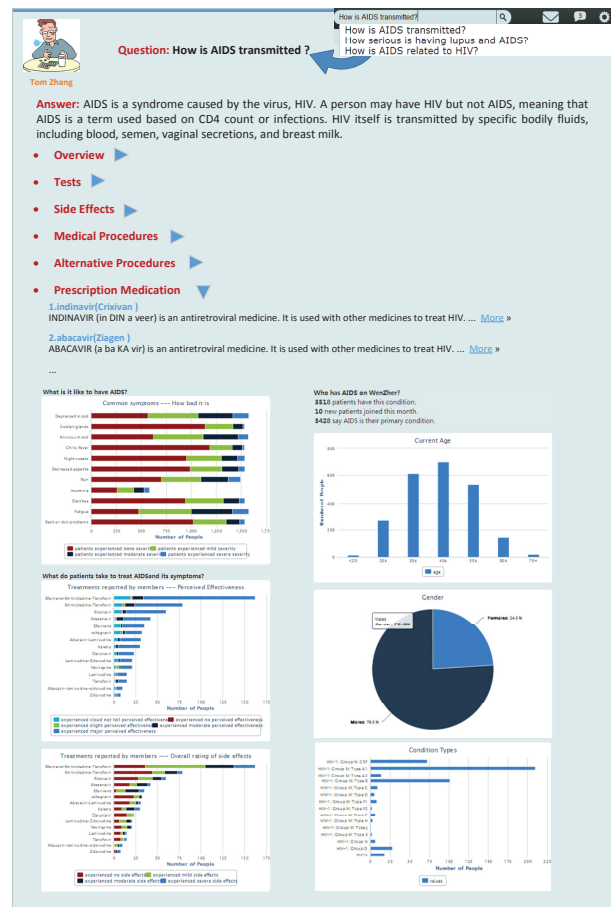


Figure 2: Comprehensive vertical search results.

tion, those automatically recommended tags are manually selectable via easily dragging.

4. ACKNOWLEDGEMENTS

This work was supported by NUS-Tsinghua Extreme Search project under the grant number: R-252-300-001-490.

5. REFERENCES

- [1] L. Nie, Y.-L. Zhao, X. Wang, J. Shen, and T.-S. Chua. Learning to recommend descriptive tags for questions in social forums. *TOIS*, 2013.
- [2] K. Wang, Z. Ming, and T.-S. Chua. A syntactic tree matching approach to finding similar questions in community-based qa services. In *SIGIR*, 2009.

Figure 3: Question post interface of WenZher.