A Medical Decision Support System Using Text Mining to Compare Electronic Medical Records

Pei-ju Lee[1(✉)], Yen-Hsien Lee[2], Yihuang Kang[3], and Ching-Ping Chao[1]

[1] National Chung Cheng University, No. 168 University Rd., Minhsiung 621, Chiayi County, Taiwan
pjlee@mis.ccu.edu.tw
[2] National Chiayi University, No. 300 Syuefu Rd., Chiayi City 600, Taiwan
[3] National Sun Yat-sen University, No. 70 Lianhai Rd., Gushan District, Kaohsiung City 804, Taiwan

Abstract. The electronic medical records (EMRs) contain information about the patient such as their date of birth and blood type as well as other medical information such as prescription history and previous syndromes. Physicians usually have limited time to identify critical information on medical records and to provide a summary before they make a decision. However, the content of EMRs usually be complicated, repeated, and contain many consistency problems; these issues are not only cost a lot of time for physicians to filter information out from the medical records but also increase the probability of wrong medical decisions. Therefore, this study proposed a new EMR interface to identify the new medical information such as new syndromes or the turning point in the medical records. The Metathesaurus database which contains medical information such as medical terms or classification codes in the Unified Medical Language System will be used. This study uses MetaMap tools to compare medical terms within EMRs using MetaMap and also compares the vocabulary using the bigram technique to highlight the similarities in the EMR.

Keywords: Electronic medical records · Decision support system · MetaMap

1 Introduction

The electronic medical record (EMR), or the digital medical record, contains the interactive information between patients and medical staffs (Jardim 2013), as well as patients information such as the blood type and the physical condition. This standardized information helps hospitals collect patients' data for further analysis and allows healthcare professionals to make reference decisions for the patients (Senteioet al. 2018). The EMR is stored electronically in the cloud database of the Ministry of Health and Welfare in Taiwan, so that the Ministry of Health and Welfare can grasp the basis of people's medication and the hospital's declaration of health insurance points, and can fully store patients' information for the follow-up research and analytical uses of the government.

F. F.-H. Nah and K. Siau (Eds.): HCII 2019, LNCS 11589, pp. 199–208, 2019.
https://doi.org/10.1007/978-3-030-22338-0_16

With the implementation of EMRs, it has brought convenience and speed of the whole process of medical records. Most EMR systems enable physicians to use the functions of copy and paste or automatically import previous clinical data. However, the efficiency also brings a part of problems: when physicians use the copy and paste functions, it may cause the repeatability of some contents in EMRs, prolongs the length of whole EMR or the prescriptions prescribed by the physicians do not match the current situation of patients. These situations make it very difficult for medical staff to read medical records and take a longer time to find useful information (Zhang et al. 2017).

When physicians read EMRs, often suffer from the difficulty of data connection due to access a large amount of medical information (Prados-Suárez et al. 2012). Most of the literature is to redefine the data structure, reorganize file information, or provide a new medical system (Jerding and Stasko 1998; McAlearney et al. 2010; Prados-Suárez et al. 2012). Some studies does not directly dealing with amount problems but performing knowledge mobilization (i.e. calling professionals prepared for the action) and ubiquitous computing (i.e. people can process and capture information whenever and wherever they want) (Judd and Steenkiste 2003; Kang et al. 2006; Prados-Suárez et al. 2012), exchange of information using standardized medical information systems (Cayir and Basoglu 2008; Lähteenmäki et al. 2009; Nagy et al. 2010b; Prados-Suárez et al. 2012), or based on the user's background needs to improve data retrieval, so that the data retrieval model can better meet the real needs of customers (Prados-Suárez et al. 2012). Zhang et al. (2017) conducted a medical information semantic similarity study based on the n-gram method to find redundant EMRs.

The above research did not enable physicians to quickly identify new information on medical records and summaries in no time. Instead, they simply identified information and did not perform other function such as medical abstracts, labeling information, etc. Therefore, the medical data retrieval method of this study is to unitize the words in multiple EMRs of a single patient and use the n-gram and MetaMap comparison method to find the parts of the EMRs that are similar to the selected comparison cases. This study also proposes an interface, use it in a teaching hospital in central Taiwan as the research data, compare the medical records in an n-gram manner, and output new condition of the patient and a summary of the patient's medical records. In addition, the color-coding is used to highlight the differences between these EMRs, so that the physicians can pay attention to the information on the new condition in time. Furthermore, the medical related terms in the medical record are labeled and the medical record summary of the physician's EMR is provided using MetaMap. In turn, a decision support system was developed to identify the problem from the original data, documents, and personal knowledge, to find a solution to the problem so that the user can make better decisions (Hwang et al. 2018). The information can be shown on different devices, such as computers, cell phones, etc., in a responsive webpage, so that physicians can obtain new medical record information ubiquitously and make decisions more quickly.

2 Background

The types of decision support system consists of databases, computational model libraries, and knowledge outcome libraries; and the efficiency of the mining process and the information processed by the operations can be improved through visualization (Reyes et al. 2014; Tejeda et al. 2013; Zhang et al. 2017; Demergasso et al. 2018).

In the hospitals, because the medical information is very complicated, it may cause many medical problems, such as misadmission and misdiagnosis if we want to rely on the medical professionals to record the patients' medication, meals, etc. Therefore, the well-designed medical decision support system can then be used to remind physicians of medications and other information, or to provide physicians advice when reading reports of X-ray, CT, etc., and to alert them when the prescribed dosage is abnormal. Alerts and suggestions are provided to enable healthcare professionals to reduce the risk of medical malpractice (El-Sappagh and El-Masri 2014).

There are many different applications in the medical decision support system. For example, El-Sappagh and El-Masri (2014) collects medical information and transformed it into electronic format in various hospitals, and pre-processes electronic medical information with data mining technology. The format is unified, and then the knowledge is generated into a knowledge base by means of association rules, classification, grouping, etc. Finally, the knowledge is encoded by the prediction model, so that the knowledge engine can provide information to the medical professionals through the generated knowledge and speed up the decision of the physician. Parshutin and Kirshners (2013) pre-processed the data query form containing 28 attributes and 840 records using the classification methods C4.5 and CART as the basic classifier to find a better medical decision support system and to provide extra help to the professionals. Khan and Shamsi (2018) proposed a neural network-based clinical decision-making system that helps medical staff diagnose diseases and identify disease categories by performing natural language processing and multi-label classification techniques on patients' EMRs. deWit et al. (2015) used pharmacists' help to optimize clinical rules, to prune rules for less relevant rules in the medical decision support system and reduce false alarms in the medical decision support system. Almasri (2017) proposes an automated clinical decision-making system based on a neural network to assess multi-factor health problems, using a multi-layer perceptron to predict the risk of thromboembolic disease to provide decision-making for healthcare professionals. Nazari et al. (2018) proposed a medical decision support system based on fuzzy analysis, which was studied in four steps: (1) selection of criteria and sub-criteria, (2) weighted sub-criteria, (3) assess the patient's condition, (4) finally assessed the risk of heart disease in patients with heart disease by fuzzy analysis of risk factors. Wulff et al. (2018) proposed that when the medical decision support system generates knowledge-intensive problems, it will cause problems such as high cost and operational difficulty in the medical decision support system. Therefore, in order to solve these problems, they propose to use openEHR, which is a database queried with the Archetype Query Language syntax; when the rules in the database are touched, a warning is sent to alert the health care provider. Shanmathi and Jagannath (2018) proposed a remote health monitoring clinical decision-making system that can use

multiple signs of life (such as blood pressure, heart rate, etc.), using multi-label classification, clinical terminology, and context-aware technology, then the information can be passed to the telemedicine monitoring center for the physician to make clinical decisions when the patient is in a critical situation.

Medical information retrieval refers to the use of a retrieval model to find medical information that meets the needs of medical professionals from a large number of medical databases in hospitals. Rui et al. (2015) mentioned that when searching for medical information, he often encounters search difficulties such as searching for fever and rash when the EMR contains "the patient has fever and rash" but because of the symptoms are not very obvious so many other symptoms are found when searching for these keywords and causing ambiguity in the search. Therefore, the authors use Consumer Health Vocabulary to convert words that are not medical words into medical words and use WordNet to find possible similar words to form a search vocabulary to avoid blurring medical information search. Shamna et al. (2018) proposed an unsupervised image retrieval framework based on spatial matching patterns of visual words, which can effectively calculate the spatial similarity of visual words. Ma et al. (2017) proposed to combine semantic and visual similarity and use context-related similarity for medical image retrieval to improve image retrieval. Zhao et al. (2018) proposed a semantic search of graphics based on file relevance and document popularity. Choi et al. (2014) proposed using MetaMap to analyze medical records to find concepts and use concept identifiers to make medical information search concepts more correct. Rui et al. (2017) proposed six steps to find new information in the medical records: (1) data pre-processing, (2) delete stop words and use TF-IDF to find high frequency but not important word deletion, (3) normalize words, (4) apply rules to classify words, (5) establish bigram semantic similarity model, and (6) generate semantic similarities by formulas modified from Pedersen (2010).

The medical record is dynamic information that continuously monitors the health of patients through clinical records and medical plans. Electronic Health Record (EHR) or EMR stores the medical activities received by patients are recorded in time order, recording the patient's treatment time, diagnosis and treatment, and treatment location (Jardim 2013). In various medical information retrieval methods, for file comparison, traditional information retrieval mainly uses two elements for document ranking: text relevance and text content views. The text-related index indicates whether the search results match the user's needs; and the number of text content views can show which words are more important for users (Zhao et al. 2018). However, there is not much research in the medical information comparison.

3 Methodology

The structure of the research is shown in Fig. 1. It is expected to use the admission record and the progress note of a regional teaching hospital in Taiwan. Visual Studio will be used as the development tool and Windows 7 environment. This research also uses the Python3's natural language processing tool - Natural Language Toolkit (NLTK) to perform medical data processing on the medical data set (SpellChecker kit), stemming (PorterStemmer kit), sentence break, punctuation, etc. Then, the binary

comparison is performed and the MetaMap medical dictionary is searched to obtain new information and a summary of the medical record.

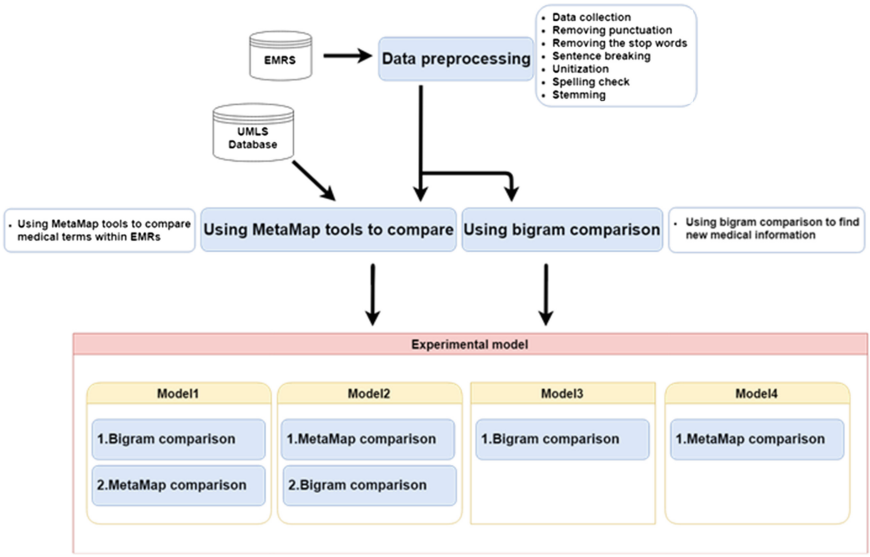

Fig. 1. Research structure.

When the EMR is processed through data integration, sentence segmentation, unitization, spelling check, stemming, removing punctuation, removing the stop word, etc., the data set without noise can be obtained. The purpose of this study is to use the hospital's EMRs to find new medical information such as new syndromes or the turning point in the medical records. The UMLS (Unified Medical Language System) Metathesaurus database will be used and which contains medical information such as medical terms, classification codes, etc. This study uses MetaMap tools to compare medical terms within EMRs, and comparing the vocabulary in the Metathesaurus database with the EMRs using the bigram comparison method to find the same vocabulary and highlight these words.

This study will use medical data from UMLS which is a system developed by the National Library of Medicine (NLM). The medical information in the system is numerous and credible. UMLS utilizes the Metathesaurus database to retrieve medical information such as Concept, Concept Unique Identifier (CUI), Semantic Type, Definition, etc. as shown in Fig. 2. And using MetaMap, the researcher can get the medical terminology of the hospital's EMR, and use the obtained medical terminology to carry out a short medical summary, so that physicians can understand the patient's condition quickly.

This study also adopts bigram to construct experimental models. In this study, the EMR in the database are first segmented, and then each EMR is unitized using a binary language model that using two words as a group. When the medical records are

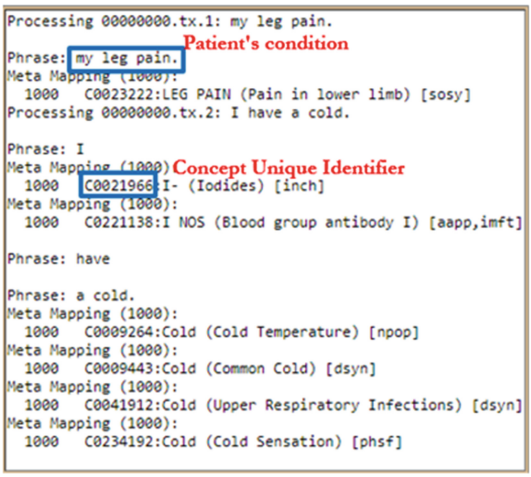

Fig. 2. Medical information retrieved from MetaMap.

compared, the medical records of the new symptoms are compared with the other medical records of the same patients. For example, as shown in Fig. 3: choose John's medical record on August 15 and records on August 14 and 13 to compare the differences, and any new symptoms on August 15 will be highlighted after the comparison. This allows physicians to quickly read the differences in medical records.

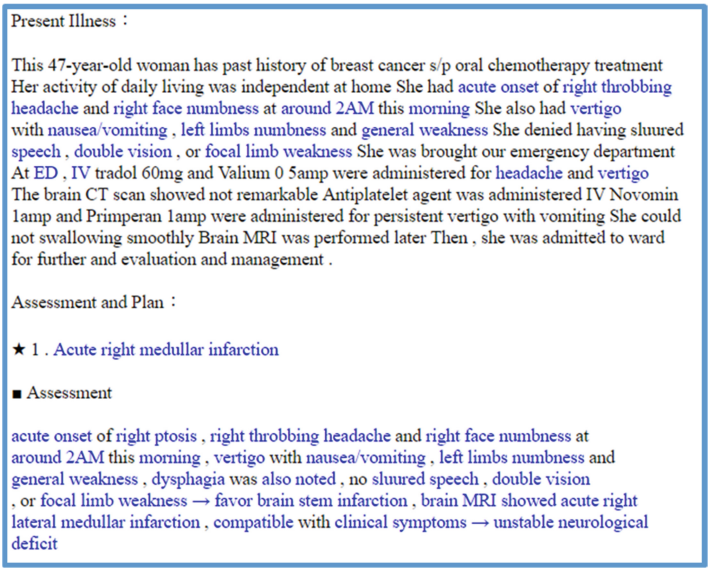

Fig. 3. Example of EMRs comparison.

According to the physicians' clinical suggestion, because the medical record system of the hospital is too complicated, it will cause inconvenient and a waste of time when reading and comparing multiple EMRs. Therefore, this study will construct a new interface based on the new information on the EMR and construct a medical decision support system that facilitates reading by healthcare professionals. At present, the web language HTML and PHP will be used for interface presentation. The initial design contains the EMR search page, the EMR comparison page, and the new information page (shown in Figs. 4, 5, and 6 respectively). The webpage will be the respond page that allows medical staff to view new medical records and medical record summaries on different devices.

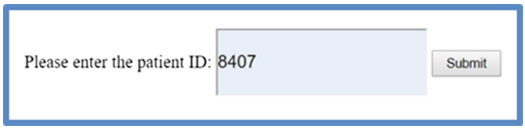

Fig. 4. Example of EMR search page.

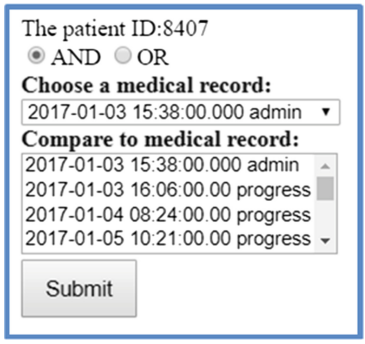

Fig. 5. Example of EMR comparison page.

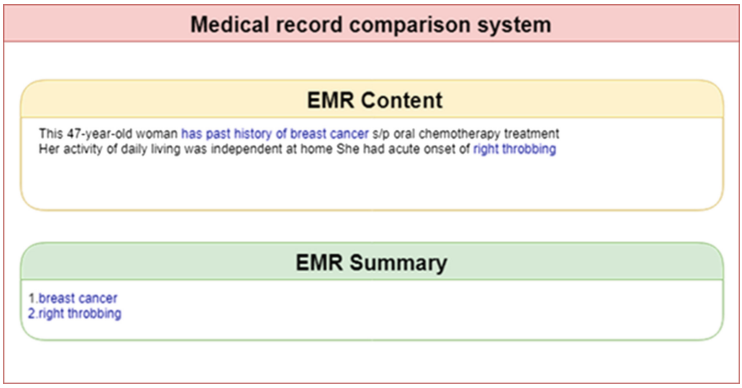

Fig. 6. Example of the new information page.

4 Evaluation

This study is expected to collect neurological EMR, and carry out the MetaMap and bigram model to find out new medical information in different EMRs and compare with the medical information found in traditional EMR interface in our experimental hospital. Before the experiment, the physician expert was asked to mark the medical information on all experiment datasets as the Gold Standard for these medical records. This study will recruit 12 medical professionals and divided them into two groups. The first group will use the new EMR interface to find new symptoms, and then use the old interface to find out new symptoms; the second group first uses the old interface, and then uses the new interface. Each group of participants will receive two different electronic medical records for the new interface and the old interface for the experiment to avoid the given knowledge of previous EMR will affect the speed and correctness of the answer. In addition, two different EMRs will be randomly selected from the experiment EMR database to avoid the difficulty varied between EMRs and will affect the outcome of the answer. Finally, evaluate the medical narrative written by the participants after reading the EMR within five minutes, and evaluate the correct rate of each participant of the experimental model according to the Gold Standard.

The Mean Absolute Error (MAE) is used to evaluate the performance. The MAE value can clearly indicate the magnitude of the error between the predicted value and the actual value. The predicted value of this study is the number of new symptoms found by recruiting medical professional using the new interface and the old interface. The actual value is the number of new symptoms (also Gold Standard) found for the neurological clinician. The size of the MAE value can be used to indicate the error value of the experiment; the larger the MAE value, the lower the accuracy of the experiment, and the smaller the MAE value, the higher the accuracy of the experiment.

5 Conclusion

This research is expected to collect EMRs, and carry out the MetaMap and bigram comparison using Python and presented using the web page as an interface. This study is expected to use patients' EMRs in a teaching hospital in the central region of Taiwan as research data, compare the medical records at a different time to locate the differences among these records in an n-gram manner or in MetaMap terminologies. The outputs of the system are the patient's new syndromes and the patient's medical record summary. And the information can be presented on different devices such as computers, mobile phones, etc., using a responsive webpage so that physicians can obtain new medical record information more quickly and make proper decision efficiently; in addition, they can catch the different types of diseases or complications and other medical information more easily.

References

Almasri, N.: Clinical decision support system for venous thromboembolism risk classification. Appl. Comput. Inform. (2017). https://doi.org/10.1016/j.aci.2017.09.003

Cayir, S., Basoglu, A.N.: Information technology interoperability awareness: a taxonomy model based on information requirements and business needs. In: Proceedings of PICMET: Portland International Center for Management of Engineering and Technology, pp. 846–855 (2008). https://doi.org/10.1109/PICMET.2008.4599692

deWit, H.A.J.M., et al.: Evaluation of clinical rules in a standalone pharmacy based clinical decision support system for hospitalized and nursing home patients. Int. J. Med. Inform. **84** (6), 396–405 (2015). https://doi.org/10.1016/j.ijmedinf.2015.02.004

Demergasso, C., et al.: Decision support system for bioleaching processes. Hydrometallurgy **181** (September), 113–122 (2018). https://doi.org/10.1016/j.hydromet.2018.08.009

El-Sappagh, S.H., El-Masri, S.: A distributed clinical decision support system architecture. J. King Saud Univ. Comput. Inf. Sci. **26**(1), 69–78 (2014). https://doi.org/10.1016/j.jksuci.2013.03.005

Hwang, B.G., Shan, M., Looi, K.Y.: Knowledge-based decision support system for prefabricated prefinished volumetric construction. Autom. Constr. **94**(June), 168–178 (2018). https://doi.org/10.1016/j.autcon.2018.06.016

Jardim, S.V.B.: The electronic health record and its contribution to healthcare information systems interoperability. Procedia Technol. **9**, 940–948 (2013). https://doi.org/10.1016/j.protcy.2013.12.105

Jerding, D.F., Stasko, J.T.: The information mural: a technique for displaying and navigating large information spaces. IEEE Trans. Visual Comput. Graphics **4**(3), 257–271 (1998). https://doi.org/10.1109/2945.722299

Jiang, J.J.: Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy (Rocling X) (1997)

Judd, G., Steenkiste, P.: Providing contextual information to ubiquitous computing applications. In: 1st IEEE International Conference on Pervasive Computing and Communications (PerCom 2003), pp. 133–142 (2003). https://doi.org/10.1007/s00703-015-0406-0

Kang, D.O., Lee, H.J., Ko, E.J., Kang, K., Lee, J.: A wearable context aware system for ubiquitous healthcare. In: Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 5192–5195 (2006). https://doi.org/10.1109/IEMBS.2006.259538

Khan, S., Shamsi, J.A.: Health quest: a generalized clinical decision support system with multi-label classification. J. King Saud Univ. Comput. Inf. Sci. (2018). https://doi.org/10.1016/j.jksuci.2018.11.003

Lähteenmäki, J., Leppänen, J., Kaijanranta, H.: Interoperability of personal health records. In: Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009, pp. 1726–1729 (2009). https://doi.org/10.1109/IEMBS.2009.5333559

Ma, L., Liu, X., Gao, Y., Zhao, Y., Zhao, X., Zhou, C.: A new method of content based medical image retrieval and its applications to CT imaging sign retrieval. J. Biomed. Inform. **66**, 148–158 (2017). https://doi.org/10.1016/j.jbi.2017.01.002

McAlearney, A.S., Robbins, J., Hirsch, A., Jorina, M., Harrop, J.P.: Perceived efficiency impacts following electronic health record implementation: an exploratory study of an urban community health center network. Int. J. Med. Inform. **79**(12), 807–816 (2010). https://doi.org/10.1016/j.ijmedinf.2010.09.002

Nazari, S., Fallah, M., Kazemipoor, H., Salehipour, A.: A fuzzy inference-fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases. Expert Syst. Appl. **95**, 261–271 (2018). https://doi.org/10.1016/j.eswa.2017.11.001

Parshutin, S., Kirshners, A.: Research on clinical decision support systems development for atrophic gastritis screening. Expert Syst. Appl. **40**(15), 6041–6046 (2013). https://doi.org/10.1016/j.eswa.2013.05.011

Pedersen, T.: Information content measures of semantic similarity perform better without sense-tagged text, pp. 329–332, June 2010

Senteio, C., Veinot, T., Adler-Milstein, J., Richardson, C.: Physicians' perceptions of the impact of the EHR on the collection and retrieval of psychosocial information in outpatient diabetes care. Int. J. Med. Inform. **113**(February), 9–16 (2018). https://doi.org/10.1016/j.ijmedinf.2018.02.003

Shamna, P., Govindan, V.K., Abdul Nazeer, K.A.: Content-based medical image retrieval by spatial matching of visual words. J. King Saud Univ. Comput. Inf. Sci. (2018). https://doi.org/10.1016/j.jksuci.2018.10.002

Shanmathi, N., Jagannath, M.: Computerised decision support system for remote health monitoring: a systematic review. IRBM **39**, 359–367 (2018). https://doi.org/10.1016/j.irbm.2018.09.007

Zhang, R., Pakhomov, S.V.S., Arsoniadis, E.G., Lee, J.T., Wang, Y., Melton, G.B.: Detecting clinically relevant new information in clinical notes across specialties and settings. BMC Med. Inform. Decis. Mak. (2017). https://doi.org/10.1186/s12911-017-0464-y

Zhao, Q., Kang, Y., Li, J., Wang, D.: Exploiting the semantic graph for the representation and retrieval of medical documents. Comput. Biol. Med. **101**(May), 39–50 (2018). https://doi.org/10.1016/j.compbiomed.2018.08.009