A Review on Recent Work in Medical Text Mining

Rajeevgandhi. V, B.Ashmitha

Department of Biomedical Engineering

Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India

Abstract: This paper discusses the software of text mining in clinical field at home and overseas based totally on overview and evaluation of contemporary literature data. Foreign researchers have unique textual content mining tools, and they use it in the search engine records and digital scientific record. In addition, it is additionally used to predict side consequences between drugs. In India, text mining based totally on medical literature records occupies a large part. On the one hand, they can monitor the self disclosure of fitness information. On the different hand, they discover whether on line information can help humans get out of the disease. With the development of data technology, text mining will come to be extra and more broadly utilized in the clinical subject in the future.

Keywords: Medical Field, Text Mining, Data Source

1. Introduction

With the fast improvement of scientific and health data development process, the kind and measurement of scientific information is developing at an extraordinary pace, so that they cannot be easily collected, managed and integrated by contemporary mainstream software program tools within a lifelike time, consequently the health center cannot take advantages of the advanced system to assist make extra positive selections which could assist these who go through from disease. There are four frequent sources of content-rich scientific data. Firstly, the records pharmaceutical companies and lifestyles science. The data generated by using drug discovery is quite dense, even for small and mid-sized enterprises, the statistics reaches the scale of TB. In the discipline of lifestyles sciences, with the gradual enlarge in computing ability and gene sequence checking out capabilities, John Wiley argues that there will be more human beings who have genetic maps, and for a single genomic sequence file, the size is about 750MB. Secondly, medical medicine. Lumping the data from clinical and laboratory together, the medical institutions are faced with a fast-paced increase in clinical data. For example, a CT photo consists of about 150MB of data and a popular pathological image is close to 5GB. Thirdly, information about costs and medical insurance plan generated during the clinical cure of patients is also any other supply of scientific data. Fourthly, healthcare social media. With the fast development of mobile devices and mobile Internet and the recognition of social media and physiological transportable devices, the statistics stored in each health communities is immeasurable.

Text mining, which is also called text-mining technique, is the discovery by using laptop of new and until now unknown information, such as interesting patterns and relationships in textual data, by routinely extracting information from one of a kind written texts. Text mining is certain up with information mining, an important research region focused on the extraction of beneficial facts from facts records. However, text mining has tested to be more challenging than

information mining, as the statistics source consists of severa unstructured texts alternatively than structured information in well-defined databases. The goal of this article is to summarize the software of text-mining approach in clinical domain. This article has two separate perspectives, the states of foreign and domestic fields respectively. In phrases of the overseas research, they no longer only focus their attentions on how to mine data from the website and social media, but additionally on developing expert software to assist they discover specific disease in journals or expert databases. Therefore, this article will make a quick introduce in this two sections. With regard to the domestic research, the main tries are the medicine regularity of a specific sickness thru normal Indian remedy (TCM), in the equal time, some western medicine's researchers to concisely understand the modern-day states of improvement in medical textual content mining and their drawbacks as well as the future directions.

2. Research Review

2.1. Foreign Research

Foreign researchers may be plenty extra in advance in the improvement of scientific textual content mining than domestic community. Their works are extra extensive, the information are more wide-ranged, from journals and databases to website and social media to explorer and digital clinical record. And their targets are an awful lot more special from every different with their personal novel viewpoint.

2.1.1. Text Mining in Journals and Databases

Luis Tari, Saadat Anwar, Shanshan Liang & James Cai1 and many others (2010) discover drug-drug interactions (DDI) via an progressive method which integrates text-mining tool [1]. Using this approach, they can not solely observe the explicit interactions however additionally the possible interactions that should be derived thru the use of a logical illustration of the area knowledge and automatic reasoning. Hodong Lee, Gwan-Su Yi & Jong (2008) center of attention their interest in particular on ubiquitin-protein ligase called E3 [2]. They enhance E3Miner which is a web-based text-mining tool that extracts and organizes comprehensive and entire information about E3s from the abstracts of journal articles and the relevant databases, providing without difficulty reachable to users with a excellent grasp of E3s and their associated information. The tool analyzes text sentences to discover protein names for E3s and additionally retrieves E3 statistics about protein functions, other E3-interacting companions and E3-related human diseases from the protein databases, which should supply us more useful information.

2.1.2. Text mining in Website and Social Media

Courtney, Diane, Armin & Karan (2010) argue that the internet and social media (WSM) is just like a pool of data, which affords us necessary disease surveillance resource [3].

In the regular method, a affected person with influenza-like sickness (ILI) will be checked by means of a physician, so it takes time to behavior a diagnostic check or take blood samples which should supply a sure diagnosis. Thus it is hard to become aware of the fashion of flu seeing that many cases of influenza stay puzzled about what disease they have. On the contrary, text-mining lookup can use mining statistics to discover the flu trend. Meanwhile, they have proved that there exists a widespread correlation between the detected expand in ILI and the sufferers reporting of Centers for Disease Control and Prevention (CDC) throughout a 24-week period from 5 October 2008 to 21 March 2009. Moreover, graph-based algorithms are now not only on hand for ranking communities that discuss and disseminate records of influenza however additionally can facilitate bio-event detection by looking for anomalies in WSM. According to

the reality that Twitter customers frequently document a combination of symptoms, as a substitute than a suspected or last diagnosis, the usage of naive and everyday language to describe their diseases, Gesualdo, Stilo & Agricola (2013) observed that Twitter has the attainable to be a rich-content and inexpensive source of information for syndromic surveillance [4].

To show their assumptions, they developed a minimally educated algorithm that exploits the abundance of health-related net pages to perceive all jargon expressions related to a unique technical term. Then the definition of an influenza case had been translated into a Boolean query, every symptom being described by a technical time period and all related jargon expressions, as recognized by way of the algorithm above. After monitored all tweets that suggested a aggregate of signs and symptoms and put them into the mannequin above, they determined a excessive correlation coefficient between the trend of influenza-positive tweets and real-life ILI developments registered by US usual surveillance systems. Nikfarjam A, Sarker A&O'Connor K etc. (2015) attempt to enhance a desktop learning-based approach to extract mentions of adverse drug reactions (ADRs) from the notably informal and descriptive text in social media [5].

They introduce a laptop learning-based concept extraction system, namely ADR Mine, to mannequin words' semantic similarities by cluster approach. Using deep getting to know technique, they generate pretrained word illustration vectors with unlabeled user posts besides supervision. It turns out that ADR Mine outperforms countless general systems in ADR extraction, accomplishing an F-measure of 0.82. It supposes that phrase cluster features remarkably enhance extraction performance and weaken the demand for large, annotated coaching data sets. Apart from disorder surveillance and extraction of ADRs, it turns out that the web and social media (WSM) is additionally full of abuse-related data which may want to be used to display prescription medication abuse by means of Abeed Sarker, Karen O'Connor & Rachel Ginn etc. (2016) [6].

They take three often used abused medicinal drugs (Adderall, Oxycodone, and quetiapine) for trying out by using collecting Twitter consumer posts after having manually annotated 6400 tweets citing these three medications. Through training, they diagram an computerized supervised classification method to determine whether or not the content material includes indicators of remedy abuse or not. To testify to the utility of automatic classification approach, they additionally analyze abuse pattern over time the usage of this classification data, reaching 82% accuracy overall.

2.1.3. Text mining in Search Engine Query

Ginsberg, Mohebbi, Matthew & Patel etc. (2009) explored another issue of the text-mining region with the aid of detecting influenza epidemics with search engine query information [7]. They now not only wish to discover seasonal influenza epidemics, which motive tens of millions of respiratory ailments and deaths global each year but greater importantly the new influenza virus against which no previous immunity exists. They believed that early detection and surveillance of health-seeking behavior in the form of queries to on line search engines may additionally provide remedy when followed by a speedy response. Consequently, they supposed a method of analyzing large numbers of Google search queries to track influenza-like illness within a specific populace based totally on the theory that the relative frequency of certain queries is relatively correlated with the proportion of patients who have influenza-like symptoms. The supposed approach ought to be in a position to discover influenza epidemics in areas of United States with a giant populace of net search users with the aid of their search queries. Philip, Yiling, David & Forrest (2008) commenced from the intuitive assumption

that the frequency of Internet searches may also grant facts related to infectious ailment endeavor [8]. Thus they performed an scan to testify the relationship between search queries contained influenza-related search terms in the Yahoo! search engine for influenza from March 2004 thru May 2008 and genuine influenza incidence in the United States. By the use of searches with 1-10-week lead instances as explanatory variables, linear models have been built to predict the proportion of cultures tremendous for influenza and deaths attributable to pneumonia and influenza in the United States. The end result confirmed that their models estimated an extend in cultures high-quality for influenza 1-3 weeks in increase of when they occurred, as nicely as an expand in mortality attributable to pneumonia and influenza up to 5 weeks in advance. Search-term surveillance might also furnish an extra tool for sickness surveillance. Hulth, Rydevik & Linde (2009) calculated the occurrence of a number of queries related to influenza from search logs submitted to a Swedish clinical web website for two influenza seasons [9]. These figures were consequently used to generate two models, one to estimate the number of laboratories established influenza cases and one to estimate the proportion of sufferers with influenza-like sickness suggested by means of selected General Practitioners in Sweden. Particularly, they applied an strategy designed for incredibly correlated data, partial least squares regression. The end result verified that sure web queries on influenza follow the same pattern as that obtained by way of the two other surveillance structures for influenza epidemics and that they have equal electricity for the estimation of the influenza burden in society. However, web queries are accurate, less costly and labor giant alternative for syndromic surveillance. Samaras, Garcia, Elena & Sicilia etc. (2012) located that a whole lot current lookup has paid interest to the viable of Web queries as a supply for syndromic surveillance and the potential for estimation and prediction of the development of a syndromic disease, such as the use of a linear model to become aware of influenza [10]. But they went in addition to build a new statistical approach to show the relationship between associated search engine queries with scarlet fever cases in the UK and the actual facts got from the professional agencies. After optimizing, every other choice statistical method based on gamma distributions used to be built and could reap better results in all cases, which capability the mannequin has the capability to predict the height and the unfold of the distributions that past the functionality of different models, in particular in those with a smaller correlation factor.

2.1.4. Text Mining in Electronic Medical Records

Luther, Mc Cart, Berndt & Hahm etc. (2015) determined whether statistical textual content mining (STM) can identify fall-related injuries in digital health report (HER) documents and the affect on STM models of training on documents from a single or more than one amenities [11]. STM fashions primarily based on training data from a single facility resulted in the inaccuracy of 87.5% and 87.1%, F-measure of 87.0% and 90.9%, the sensitivity of 92.1% and 94.1%, and specificity of 83.6% and 77.8% at the visit and affected person levels, respectively. Results from coaching statistics from multiple amenities were almost identical. STM has the viable to enhance identification of fall-related injuries in the VHA, offering a mannequin for wider utility in the evolving national HER system. Hammond, Ben, Laundry & Samore etc. (2015) applied text-mining methodology to learn about medical questions in large populations [12]. To check the feasibility of textual content mining, investigation of the relationship between publicity to negative childhood experiences (ACEs) and recorded diagnoses was carried out amongst all VA-treated Gulf battle veterans, utilizing all development notes recorded from 2000-2011. Text mining to realize ACE publicity in a massive populace was feasible. Analysis of the relationship between ACE rating and adult fitness prerequisites yielded patterns of association steady with prior research. Duggal, Shukla & Chandra etc. (2016) categorized the sufferers into two exclusive risk companies of readmission inside 30 days of discharge primarily based on patients' traits using 2-year clinical and administrative records [13]. It proposed an structure of this prediction mannequin and identified a number of threat factors the usage of textual content

mining techniques. Also, businesses of consistently taking place factors that inference readmission costs had been revealed by associative rule mining, It then evaluated the classification accuracy the use of 5 distinct data mining classifies and performed a fee analysis, Out of complete 9381 records, 1211 encounters have been located as readmissions. This learn about concludes that the model could be incorporated in healthcare establishments to witness its effectiveness. Cost analysis indicates large savings which are giant for any healthcare machine specially in creating counties like India. Casillas, Perez & Oronoz etc. (2016) extracted negative drug reaction events from electronic health data in Spanish [14]. They current a hybrid system utilizing a self-developed morpho-syntactic and semantic analyzer for clinical texts in Spanish. It performs named entity recognition of drugs and diseases and negative drug reaction event extraction. The tournament extraction stage operates using rule-based and desktop studying techniques. One of the contributions of the machine studying based totally machine is its capacity to deal with both intra-sentence and inter-sentence events in a enormously skewed classification environment. Moreover, the knowledge-based and the inferred model are complementary in terms of precision and recall. While the former presents high precision and low recall, the latter is the different way around. As a result, an appropriate hybrid strategy seems to be in a position to benefit from each procedures and also enhance them. This is the underlying motivation for choosing the hybrid approach. In addition, this is the first system dealing with real digital fitness files in Spanish. Ford, Carroll & Smith etc.(2016) extracted information from the textual content of electronic clinical archives to enhance case detection: a systematic review [15]. Electronic medical records (EMRs) are revolutionizing health-related research. One key issue for study quality is the accurate identification of patients with the condition of interest. Information in EMRs can be entered as structured codes or unstructured free text. The majority of lookup studies have used solely coded parts of EMRs for case-detection, which can also bias findings, omit cases, and limit learn about quality. This review examines whether or not incorporating records from textual content into case-detection algorithms can improve lookup quality. Text in EMRs is accessible, especially with open source data extraction algorithms, and substantially improves case detection when blended with codes. More harmonization of reporting within EMR research is needed, mainly standardized reporting of algorithm accuracy metrics like advantageous predictive price and sensitivity.

2.2. Domestic Research

Under the condition of information blast, it is vital and essential to investigate the extensive scale medicinal information by content mining system. In local research, our researchers give careful consideration principally to the normality of conventional Indian drug towards a specific illness, at times they may think about the diverse treatment of customary Indian medication between specific maladies. Barely any specialists pursue remote analysts' pace, how to use the data in web and web based life to assist the legislature with monitoring malady and individual to help is their core interest.

2.2.1. Content Mining Technique Towards a Particular Disease

The relationship among side effect, example and drug regularities of hypertension have been explored by He Dan, Jiang Miao, Zheng Chi and so forth (2014) through applying content mining system [16]. Through examination of the information in the pertinent writing, they touched base at the end exposed to information cutting calculation. The outcomes are that the most well-known indications of hypertension are a cerebral pain (2650), unsteadiness (1734) and the most harmed organ are liver and kidney and the most valuable herb are Gastrodiae Rhizoma and Uncaria Ramulus Cum Uncis in TCM. Guo Hongtao, Zheng Guang, Zhao Jing and so on (2011) utilized the content mining method to download crude information from the Indian BioMedical Literature Database (CBM) with the watchword H1N1 [17]. In the wake of

preprocessing, the information was dug for customary Indian remedial regularities in the treatment of flu H1N1 dependent on the discrete subordinate calculation. It is demonstrated that treatment of flu H1N1 focuses on scattering warmth and settling poisonous quality in TCM, that the broadly utilized TCM medications are Lian-Hua-Qing-Wen Capsule, Tan-Re-Qing Injection, Qing-Kai-Ling Injection, that the disorder of flu H1N1 can be characterized by lethality, warmth, and mucus when utilizing TCM design. The drug regularities of constant gastritis have been investigated by Li, Zhou Qi, Zheng Guang and so forth (2011) in view of content mining method [18]. The information are gathered from the rich-content writing in CBM, they through preprocessing, the crude, unstructured XML type information are moved into organized information into the database. To demonstrate the end, various types of diagrams are presented by Cytoscape programming, imagining the aftereffect of their examination. It is proposed that the most as often as possible utilized Indian patent drug are Sini San and Lizhong Wan.

2.2.2. Content Mining Technique Towards Disease Comparison

Guo Hongtao, Zheng Guang, Zhang Chi and so forth (2010) began from the well known saying in TCM, treating diverse with a similar treatment, since a few maladies have a similar rule in TCM sense, for example, the connection between Rheumatoid Arthritis (AR) and Diabetes Mellitus (DM) [19]. In the paper, a content mining calculation dependent on the word recurrence has been proposed to mine the information from writing in CBM with the catchphrases AR and DM. It is recommended that the Indian herb Huang Qi, Dang Gui are the basic center drug for both AR and DM, as their capacities are blood-actuating and stasis-settling. The past supposition turns out to be a strong logical establishment. Another part of hypertension, the examination of regularities of Indian herbs, Indian patent medication and western drug in treatment, is being dissected by Wang Liying, Zheng Guang and Guo Hongtao and so forth (2013) utilizing the content mining procedure [20]. Writing about hypertension from CBM have been extricated, and afterward exposed to information process. They reasoned that gastrodia elata and Uncaria rhynchophylla were most generally utilized Indian herbs, infusion of red sage root and zhenju jiangya tablet were most unavoidable Indian patent medication, captopril, and nifedipine were most well known western drug in the treatment of hypertension. In addition, they ran further with the best blend of Indian patent medication and western prescription, including captopril and zhenju Jiangya tablet, nifedipine and zhenju jiangya tablet.

2.2.3. Content Mining in Website and Social Media

Ku, Chiu and Zhang and so forth (2014) investigated self-revealing wellbeing data to help general wellbeing reconnaissance and social insurance utilizing information from two noteworthy HIV/AIDS gatherings (Yahoo! Learning and Taiwan AIDS Foundation gathering) in Taiwan [21]. The structure they constructed incorporates a steady web crawler program to secure information and the Stanford Word Segmenter to parse sentence dependent on streamlined Indian. Moreover, a few capabilities (sack of-words, word/POS, pack of-words+word/POS, chose sack of-words+word/POS, and DSF) were produced from the gathered postings by various content portrayal approaches, combined with SVM and NB calculations to prepare the classifier for distinguishing uncommon messages. Their outcome has demonstrated the adequacy of applying content mining systems on HIV/AIDS-related self-unveiling wellbeing data on web discussions and the promising pertinence to social insurance practice, for example, other self-revealing wellbeing data issues (e.g. suicide, weight reduction). Lu Yan and Yong Tan (2014) make another point of view, they explore whether social help traded in an online medicinal services network benefits patients' psychological well-being [22]. Social help trade implies that patients can login in an online network where they can find the solution and compose their riddles about specific illness. So as to gauge the social help, right off the bat they grouped social help into three kind, enlightening help, passionate help and fellowship. Furthermore, they

proposed a non-homogeneous Partially Observed Markov Decision Process (POMDP) model to look at conceivable wellbeing results for the individuals. What they found was that patients profit by learning for moms and that their support in the online network causes them to enhance their wellbeing and to all the more likely take part in their malady self-administration process. It was additionally demonstrated that distinctive types of social help trade have diverse impact on patients' wellbeing conditions, among which enlightening help is the most regular sort and enthusiastic help assumes the most essential job in helping patients move to a more beneficial state. At long last, the proposed POMDP model can anticipate patients' wellbeing states with moderate exactness and can renew absent or inaccessible explanation of patients' wellbeing conditions.

3. Conclusion

In the period of enormous information, how to rapidly get to the data in therapeutic assets, and how to utilize pertinent data to upgrade the medicinal experience of individuals is a sensible issue that should be unraveled later on. Remote specialists might be substantially more ahead in the improvement of restorative content mining than local network. Their principle work can be isolated into four separate parts. Right off the bat, they create explicit devices or projects for their exceptional interest, extending from the data of aubiquitin-protein ligase to sedate medication communications. Furthermore, the web and online networking (WSM) has been viewed as another rich asset of medicinal information, numerous endeavors have been taken to distinguish malady patterns, discover content containing unfavorable medication responses and screen doctor prescribed prescription maltreatment. Thirdly, the utilization of content mining instruments in web search tool inquiry information is another region for infection location and reconnaissance. At long last, the use of content mining apparatuses in electronic medicinal records is another technique to foresee the likelihood of some ailment. Additionally scientists can discover the association between medications. Likewise it was utilized in anticipating symptoms among medications and enhancing a case identification.

In residential network, writing based looks into record generally everything being equal, for example, the writing from database CBM. The distinction between scientists is that some utilization edited compositions of explicit articles, however others think full-content is more far reaching. Their subject focuses on the consistency of customary Indian prescription regardless of for single illness or malady correlation, including hypertension, H1N1, influenza, interminable gastritis, diabetes mellitus and rheumatoid joint pain. Scarcely any analysts save exertion on the data pool in web and internet based life as of late. On one hand, they can screen self-revealing wellbeing data for medicinal services organizations, then again, they investigate whether the online data could push individual to help from malady themselves.

Taking everything into account, the specialists center around the fundamental strategies and procedures of content mining in therapeutic field and its application in prescription and related ailment. Anyway the examination about content mining in malady misdiagnosis, clinical obsessive investigation, and mental ailment is generally few. Scientists can utilize content mining in these perspectives. Also with the calculations of content mining are continually progressed. Scientists can discover more esteem data in medicinal field. Later on content mining will turn out to be increasingly broadly connected in the restorative field.

References

[1]   Tari, Luis, et al. "Discovering drug–drug interactions: a text-mining and reasoning approach based on properties of drug metabolism." Bioinformatics 26.18 (2010): i547-i553.

[2]   Lee, Hodong, Gwan-Su Yi, and Jong C. Park. "E3Miner: a text mining tool for ubiquitin-protein ligases." Nucleic acids research 36.suppl_2 (2008): W416-W422.

[3]   Corley, Courtney D., et al. "Text and structural data mining of influenza mentions in web and social media." International journal of environmental research and public health 7.2 (2010): 596-615.

[4]   Gesualdo, Francesco, et al. "Influenza-like illness surveillance on Twitter through automated learning of naïve language." PLoS One 8.12 (2013): e82489.

[5]   Nikfarjam, Azadeh, et al. "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features." Journal of the American Medical Informatics Association 22.3 (2015): 671-681.

[6]   Sarker, Abeed, et al. "Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter." Drug safety 39.3 (2016): 231-240.

[7]   Ginsberg, Jeremy, et al. "Detecting influenza epidemics using search engine query data." Nature 457.7232 (2009): 1012.

[8]   Polgreen, Philip M., et al. "Using internet searches for influenza surveillance." Clinical infectious diseases 47.11 (2008): 1443-1448.

[9]   Hulth, Anette, Gustaf Rydevik, and Annika Linde. "Web queries as a source for syndromic surveillance." PloS one 4.2 (2009): e4378.

[10]   Samaras, Loukas, Elena García-Barriocanal, and Miguel-Angel Sicilia. "Syndromic surveillance models using Web data: The case of scarlet fever in the UK." Informatics for Health and Social Care 37.2 (2012): 106-124.

[11]   Luther, Stephen L., et al. "Improving identification of fall-related injuries in ambulatory care using statistical text mining." American journal of public health 105.6 (2015): 1168-1173.

[12]   Hammond, Kenric W., et al. "The Feasibility of Using Large- Scale Text Mining to Detect Adverse Childhood Experiences in a VA- Treated Population." Journal of traumatic stress 28.6 (2015): 505-514.

[13]   Duggal, Reena, et al. "Predictive risk modelling for early hospital readmission of patients with diabetes in India." International Journal of Diabetes in Developing Countries 36.4 (2016): 519-528.

[14]   Casillas, Arantza, et al. "Learning to extract adverse drug reaction events from electronic health records in Spanish." Expert Systems with Applications 61 (2016): 235-245.

[15]   Ford, Elizabeth, et al. "Extracting information from the text of electronic medical records to improve case detection: a systematic review." Journal of the American Medical Informatics Association 23.5 (2016): 1007-1015.

[16]   Yanhong, Li, Song Anmeng, and Wang Jingling. "A Survey of Current Work in Medical Text Mining---Data Source Perspective." International Journal of Biomedical Science and Engineering 5.3 (2017): 29.

[17]   Yanhong, Li, Song Anmeng, and Wang Jingling. "A Survey of Current Work in Medical Text Mining---Data Source Perspective." International Journal of Biomedical Science and Engineering 5.3 (2017): 29.

[18]   Li, Li, et al. "Based on Text Mining Techniques to Explore Medication Regularity of Chinese Patent Medicine and West Medicine Application for Chronic Gastritis [J]." Chinese Journal of Experimental Traditional Medical Formulae 24 (2011): 066.

[19]   Yanhong, Li, Song Anmeng, and Wang Jingling. "A Survey of Current Work in Medical Text Mining---Data Source Perspective." International Journal of Biomedical Science and Engineering 5.3 (2017): 29.

[20]   Yanhong, Li, Song Anmeng, and Wang Jingling. "A Survey of Current Work in Medical Text Mining---Data Source Perspective." International Journal of Biomedical Science and Engineering 5.3 (2017): 29.

[21] Ku, Yungchang, et al. "Text mining self- disclosing health information for public health service." Journal of the Association for Information Science and Technology 65.5 (2014): 928-947.

[22] Yan, Lu, and Yong Tan. "Feeling blue? go online: an empirical study of social support among patients." Information Systems Research 25.4 (2014): 690-709.