

REVIEW

BIG DATA IN HEALTHCARE – A REVIEW

Magesh G^{1*}, Swarnalatha P²¹School of Information Technology and Engineering, VIT University, Vellore, INDIA²School of Computer Science and Engineering, VIT University, Vellore, INDIA

ABSTRACT

In exhibit world, information is developing step by step. Enormous information is created in a medicinal services association. Information is futile unless we won't locate any valuable data from it. Information examination turns into a test when we have such huge measure of information. In this paper, we have assessed recent year's papers. Information mining methodologies and planning for huge information is additionally surveyed. This paper will help future specialists to grow new methodologies and calculations to explain few difficulties in the healthcare services industry.

INTRODUCTION

Finding talent is one of the most challenging tasks today, but is not usually prevailing because of the problem, not being technical. We are living in a world where everyone looks for profits. According to most of the industries, most of the students graduating from India are not fit for work and thus they go through about a year of industrial training, shouldn't all colleges be responsible for that? I think all colleges should take care of that field training, instead of relying on industries. Most students can learn programming on their own and if not, should be specially taken care of, but for the rest, logic and algorithms should be the focus of their study, applying and integrating it in field should be the primary objective.

The next in line is the expensiveness of the server hardware, for the time being, we cannot reduce the costs immediately, but we can make a smart investment, and may be reduce the costs in the long run. I believe we should work on real time compression techniques, thus improving its efficiency and file size. Taking this into account, an intermediate system should be integrated with the original system between the server and desired application, here map reduce, dedicated for file compression and decompression. During the first stages of implementation we might not notice a huge difference, but in the long run a significant reduction in server load may be noticed.

There are very few studies regarding the wearing and laundering of lab coats in hospitals and medical practice. This study highlights the role of lab coats acting as vector for transmitting health care infections to the patients and the common areas where contamination occurs.

Literature Review

Analyzing patients sentiments using hadoop and machine learning

Prescient demonstrating is basic to changing substantial clinical informational indexes, or "enormous clinical information," into significant learning for different medicinal services applications. Machine learning is a noteworthy prescient demonstrating approach; however, two hindrances make its utilization in social insurance testing. Initial, a machine learning instrument client must pick a calculation and dole out at least one model parameters called hyper-parameters before demonstrate preparing. The calculation and hyper-parameter esteem utilized regularly affect demonstrate precision by more than 40 %, yet their determination requires many works concentrated manual cycles that can be troublesome notwithstanding for PC researchers. Second, numerous clinical qualities are more than once recorded after some time, requiring worldly accumulation before prescient demonstrating can be performed. Much work serious manual emphasis are required to distinguish a decent combine of total period and administrator for each clinical characteristic. The two obstructions result in time and human asset bottlenecks and block medicinal services overseers and specialists from soliciting an arrangement from imagining a scenario in which questions while testing chances to utilize prescient models to enhance results and lessen costs. This paper portrays their plan of and vision for PredicT-ML (forecast instrument utilizing machine taking in), a product framework that means to defeat these boundaries and mechanize machine learning model working with enormous clinical information [Fig-2]. The paper introduces the point by point plan of PredicT-ML. Foresee ML will open the utilization of enormous clinical information to a huge number of social insurance managers and analysts and increment the capacity to progress clinical research and enhance human services. [1]

The possibility that the simply phenomenological information that they can separate by breaking down a lot of information can be valuable in medicinal services appears to negate the craving of VPH analysts to construct itemized unthinking models for singular patients. Yet, practically speaking no model is ever altogether phenomenological or totally unthinking. They propose in this position paper that enormous information investigation can be effectively joined with VPH advancements to create hearty and compelling

KEY WORDS

Big Data, Healthcare,
Hadoop, Internet,
Information Technology
(IT)

Received: 11 May 2017
Accepted: 10 July 2017
Published: 21 Sept 2017

*Corresponding Author

Email: magesh.g11@gmail.com

in silico prescription arrangements. So as to do this, huge information innovations must be additionally created to adapt to some particular pre requisites that rise up out of this application. Such necessities are: working with delicate information; investigation of mind boggling and heterogeneous information spaces, including non-printed data; circulated information administration under security and execution limitations; specific examination to incorporate bioinformatics and frameworks science data with clinical perceptions at tissue, organ and living beings scales; and concentrated investigation to characterize the "physiological envelope" amid the everyday life of every patient. These area particular prerequisites propose a requirement for focused financing, in which huge information advancements for in silico medication turns into the exploration need. [2]

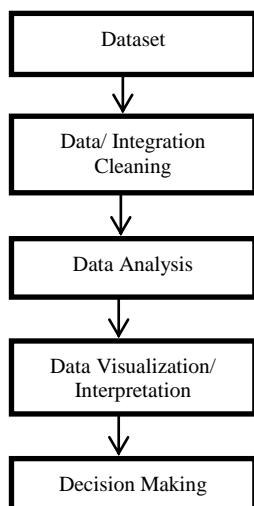


Fig. 1: Decision-making

With the advancement of savvy gadgets and distributed computing, increasingly general wellbeing information can be gathered from different sources and can be examined in an extraordinary way. The immense social and scholarly effect of such improvements caused an overall buzz for huge information. In this audit article, they compressed the most recent uses of Big Data in wellbeing sciences, incorporating the suggestion frameworks in social insurance, web based pestilence observation, sensor-based wellbeing conditions and sustenance security checking, Genome-Wide Association Studies (GWAS) and articulation Quantitative Trait Loci (eQTL), surmising air quality utilizing huge information and metabolomics and bionomics for nutritionists. They additionally checked on the most recent advancements of enormous information gathering, stockpiling, exchanging, and the cutting edge logical strategies, for example, Hadoop circulated document framework, MapReduce, proposal framework, profound learning and system Analysis. Finally, they talked about the future points of view of wellbeing sciences in the time of Big Data. [3]

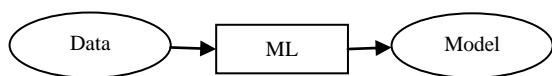


Fig. 2: Machine learning

This target of this paper to depict the guarantee and capability of enormous information examination in human services. The paper portrays the beginning field of enormous information examination in human services, talks about the advantages, diagrams a structural system and strategy, depicts illustrations revealed in the writing, quickly examines the difficulties, and offers conclusions. Results: The paper gives a wide diagram of huge information investigation for medicinal services analysts and experts. Conclusion is that enormous information examination in human services is developing into a promising field for giving understanding from huge informational indexes and enhancing results while diminishing expenses. Its potential is extraordinary; however there remain difficulties to overcome. As large information investigation turns out to be more standard, issues, for example, ensuring protection, defending security, setting up principles and administration, and constantly enhancing the devices and advancements will gather consideration. Enormous information examination and applications in medicinal services are at an incipient phase of improvement, however quick advances in stages and instruments can quicken their developing procedure. [4]

This work shows a novel, Cloud-based framework for information gathering and capacity to catch smoking conduct with e-cigarette. A client's smoking information created by the day by day utilization of e-cigarettes

is transferred to the cloud through versatile web and a Bluetooth association between an advanced mobile phone and the e-cigarette. All individual character can be scrambled and an investigation personality number will be doled out to each subject for information security insurance. The remote stage in the cloud can give productive investigative execution on an enormous volume of information with high speed of information creation. Information mining on smoking conduct will better comprehend the methods for utilizing the e-cigarette. This information framework will likewise be conceivably utilized as a part of other epidemiological investigations in general wellbeing. In this investigation, they display their work in the improvement of an information catching stage in the cloud for e-cigarettes associated with cell phones. This stage is intended for future examination on the utilization of e-cigarette for smoking lessening or restraint. Information on smoking conduct and the viability of e-cigarette utilize is caught by a cell phone APP, transmitted to the cloud, and a synopsis provide details regarding smoking status will be naturally created for get to through the cell phone. They trust that this intelligent stage on e-cigarette utilize will build the achievement rate of smoking restraint. [5]

In this paper creators portray the principal ever endeavor of constant information investigation for social insurance datasets utilizing as a part of memory databases and they benchmark and analyze two such in-memory database frameworks to examine responsiveness and capacity to deal with multifaceted nature of run of the mill wellbeing information investigation undertakings, they share their work in advance outcomes and blueprint key issues that should be tended to for approaching advances in this imperative enormous information vertical. This paper investigates persuading situations for investigating exceptionally complex human services informational collections continuously by performing specially appointed inquiries. They look into two economically accessible In-memory database advances Mem SQL and Volt DB for responsive medicinal services information investigation. Both of these databases are New SQL social databases. For correlation they utilized Medicare Claims Synthetic information from Center for Medicare and Medicaid Services (CMS). The primary commitments of this work are speedier information investigation and empowering ongoing expectations. [6]

An expanding number of the elderly populace wish to carry on with an autonomous way of life, as opposed to depend on meddling consideration programs. A major information arrangement is displayed utilizing wearable sensors fit for completing nonstop observing of the elderly, cautioning the significant guardians when vital and sending relevant data to a major information framework for examination. A test for such an answer is the improvement of setting mindfulness through the multidimensional, dynamic and nonlinear sensor readings that have a powerless relationship with detectable human practices and wellbeing conditions. To address this test, a wearable sensor framework with an astute information forwarder is talked about in this paper. Ping Jiang et al. displayed an arrangement of utilizing wearable sensors fit for ceaseless observing of the elderly and sending the information to the huge information frameworks. A major information framework oversees high volume, speed and assortment of data from various sources, to address this test a wearable sensor framework with keen information forwarder was presented receiving Hidden Markov display for human conduct acknowledgment for nourishing important information to the framework. The sensor readings and conditions of a client are sent to Big Data investigation for enhancing and customizing the nature of care. [7]

This paper concentrates on enormous information issues, points of view and arrangements as for human services industry. Diverse enormous information arrangements and a structure for taking care of heterogeneous clinical information has been examined which if actualized successfully could demonstrate supportive in basic leadership for understanding consideration and general medicinal services results. The social insurance industry is good to go for change with a specific end goal to enhance quiet care and to make it savvy by utilizing on most recent advances. Digitization has prompted huge measure of advanced information particularly in human services industry. This paper proposes a technique for taking care of heterogeneous human services information utilizing the correct innovation and design which can possibly change social insurance results and the nature of patient care at ideal cost. This paper likewise shows different logical instruments that can be utilized to use profits by the immense arrangement of social insurance information. Legitimate choice of devices to do examination on medicinal services information can give promising outcomes. The distinctive reception challenges highlighted in the paper are a portion of the open issues and can be considered as a territory of research for future work. [8]

In this paper, creators propose a Genetic Algorithm based scheduler for such Big Data Cloud where decoupled computational and information administrations are offered as administrations. The approach depends on developmental techniques concentrated on information conditions, computational assets and powerful use of data transfer capacity in this way accomplishing higher throughputs. They portray a booking model in view of GA and the model is assessed and contrasted and before works, for example, coordinate making and different heuristics systems through the reproduced information. They build up a planning system, where computational assets and information stockpiles are decoupled with the information recreated over capacity storehouses which are geological scattered. Here, the issue is centered on gathering the occupations in view of the information necessities, and the goal is to limit the aggregate make traverse considering both computational assets and correspondence transmission

capacity viably. The proposed family/gather booking model tends to the information concentrated issues to limit the turnaround time of the employments where the registering and information assets are decoupled. [9]

Predictive big data analytics

This paper talks about the part of Big Data in the present medicinal services segments. The change of utilizing complex advancements by medicinal services suppliers to pick up bits of knowledge from clinical datasets and settle on educated choices had changed by Big Data Analytics. With the assistance of Hadoop, the objective of successful human services administration can be accomplished by giving compelling information driven administrations to individuals by anticipating their requirements. This paper gives an outline of putting away and recovery strategies, Big Data apparatuses and methods utilized as a part of social insurance mists, part of Big Data Analytics in medicinal services and talks about the advantages, viewpoints in early fields of prescient investigation, confronts challenges and gives arrangements. The method for association of information after extraction from various layers and coordination of it is additionally a testing undertaking. Overseeing and keeping up the human services information is smidgen basic since social insurance information is kept up in medicinal services records, they are colossal in nature and need much space, further, on the off chance that they require related information it needs to experience with examination to get significant information. The part of technocrats in social insurance field ought to keep up adjust amongst business and innovation, which clear a way for Big Data in medicinal services data frameworks. Applying Big Data in medicinal services space made an upheaval in keeping up and managing intricate, unstructured heterogeneous information. [10]

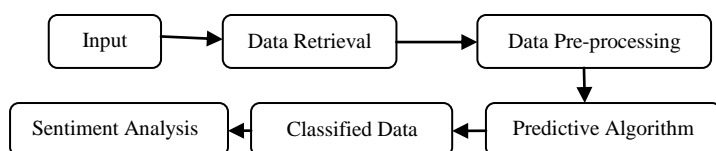


Fig. 3: Sentiment analysis

This paper portrays the outline, usage of fluffy master framework for conclusion of diabetes. The parts of fluffy master framework are fuzzification interface, Fuzzy evaluation technique and De-fuzzification interface. Fuzzification interface changes over the fresh esteems into fluffy esteems. Fluffy evaluation strategy utilizes fluffy administrators, enrollment work, connection fluffy rationale and likelihood to oversee uncertainty in rules. De-fuzzification interface changes over the subsequent fluffy set into fresh esteems. To show the viability of the proposed calculation MATLAB Fuzzy Logic tool kit is utilized for execution evaluation. The outcome shows that the fluffy evaluation system is exceptionally compelling in enhancing the exactness for diabetes application. [11]

Master frameworks are late result of manmade brainpower. It is an arrangement of projects that control encoded information to take care of issues in a particular area. Diabetes is a constant ailment that requires persistent medicinal care in tolerant self-administration training to avoid intense difficulties. Gestational diabetes mellitus (GDM) is a kind of Diabetes display in around 3-4% of all pregnancies. This paper proposes a technique to distinguish the GDM in pregnant ladies. Analysts contributed their work in diagnosing the diabetes mellitus. Their proposed work concentrates on displaying a specialist framework to analyze GDM utilizing Feed forward Neural Network design. [12]

This paper manages utilizing fluffy rationale to limit vulnerability impacts in reconnaissance. It thinks about the origination of a productive fluffy master framework that had two qualities: bland and strong to vulnerabilities. Dissecting separation between factors ideal and genuine esteems is the primary thought of the examination. Fluffy induction framework chooses, at that point, about noteworthy factors state: ordinary or irregular. A correlation between three proposed fluffy master frameworks is displayed to highlight the impact of participation number and sort. Next to, being bland this framework could likewise be connected in three fields: modern observation, camera reconnaissance and medicinal observation. To uncover brings about these fields, MATLAB is utilized to understand this approach and to reproduce frameworks reactions which uncovered intrigued conclusions. [13]

Choice emotionally supportive networks give quiet particular suggestions to mind suppliers amid clinical experiences. The MobiGuide extend additionally addresses patients by settling on them clients of the choice emotionally supportive network through their Smartphone interface, by customizing direction as per the patients' inclinations in a way touchy to their own specific circumstance, and by including patients in a common basic leadership together with their care suppliers. Furthermore, the MobiGuide framework gives choice help in non-clinically-controlled situations, for example, the patient's working environment and home, and can convey some of its proposals even without a care-supplier's intercession; consequently, it is a Ubiquitous Guidance System. The clinical learning that constitutes the rationale of the universal direction

framework depends on confirm based clinical rules. These properties make the MobiGuide framework possibly exceedingly available and quick. Also, patients' wellbeing is expanded. While keeping most checked patients out of the facility, the framework prompts an expansion in wellbeing quality and a lessening in medicinal services costs. [14]

Patient-centric clinical decision support system

Clinical practice rules (CPGs) expect to enhance the nature of care, decrease unjustified practice varieties and lessen social insurance costs. With the end goal for them to be viable, clinical rules should be coordinated with the care stream and give persistent particular exhortation when and where required. Henceforth, their formalization as PC interpretable rules (CIGs) makes it conceivable to create CIG-based choice emotionally supportive networks (DSSs), which have a superior shot of affecting clinician conduct than story rules. This paper audits the writing on CIG-related systems since the initiation of CIGs, while centering and drawing subjects for characterizing CIG explore from CIG-related productions in the Journal of Biomedical Informatics (JBI). The subjects traverse the whole life-cycle of CIG advancement and include: learning obtaining and determination for enhanced CIG configuration, including (1) CIG demonstrating dialects and (2) CIG procurement and particular techniques, (3) mix of CIGs with electronic wellbeing records (EHRs) and authoritative work process, (4) CIG approval and check, (5) CIG execution motors and strong devices, (6) exemption taking care of in CIGs, (7) CIG support, including investigating clinician's consistence to CIG suggestions and CIG forming and development, lastly (8) CIG sharing. I analyze the transient patterns in CIG-related research and examine extra subjects that were not distinguished in JBI papers, including existing topics, for example, defeating execution boundaries, demonstrating clinical objectives, and worldly articulations, and in addition advanced topics, for example, tolerant driven CIGs and circulated CIGs. [15]

Enthusiasm for distinguishing individuals at danger of creating Type 2 Diabetes Mellitus (T2DM) has increased extraordinary significance particularly at its asymptomatic stage, when early intercessions have a demonstrated helpful impact on clinically important results. The current indicative criteria are centered on recognizing bunches with fundamentally expanded commonness of small scale vascular difficulties. This proposes the current indicative strategies are feeling the loss of the chance to recognize IGT and IFG (pre-diabetes) and early side effects of T2DM, which prompts a late ID and treatment of patients and the ensuing improvement of inconveniences, which could be stayed away from with a prior mediation. This paper presents the MOSAIC venture which means to enhance the present principles for diabetes conclusion and administration. [16]

Diabetes is an incessant sickness that requires a man with diabetes to make a huge number of day by day self-administration choices and perform complex care exercises. Diabetes self-administration training and support (DSME/S) gives the establishment to help individuals with diabetes to explore these choices and exercises and has been appeared to enhance wellbeing outcomes.1-7 Diabetes self-administration instruction (DSME) is the way toward encouraging the information, expertise, and capacity essential for diabetes self-mind. Diabetes self-administration bolster (DSMS) alludes to the help that is required for executing and maintaining adapting abilities and practices expected to self-oversee on a continuous premise (see advance definitions in Table 1). Albeit diverse individuals from the social insurance group and group can add to this procedure, it is vital for medicinal services suppliers and their training settings to have the assets and an efficient referral procedure to guarantee that patients with sort 2 diabetes get both DSME and DSMS in a steady way. The underlying DSME is regularly given by a wellbeing proficient, while continuous help can be given by work force inside training and an assortment of group based assets. DSME/S programs are intended to address the patient's wellbeing convictions, social needs, current learning, physical restrictions, enthusiastic concerns, family bolster, budgetary status, medicinal history, wellbeing education, numeracy, and different elements that impact every individual's capacity to address the difficulties of self-administration. [17]

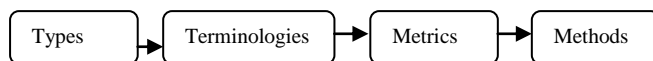


Fig. 4: Clinical decision support system.

Meeting the perplexing needs of patients with perpetual disease or weakness is the single most noteworthy test confronting sorted out therapeutic practice. Common care is not doing the occupation; many reviews and reviews have uncovered that sizable extents of incessantly sick patients are not accepting viable treatment, have poor malady control, and are miserable with their care (1). Consequences of randomized trials likewise demonstrate that compelling malady administration projects can accomplish significantly preferred results over normal care, the control mediation. These trials, alongside the thoughts and endeavors for development talked about in this issue, demonstrate that they can enhance care and results. As the articles recommend, these upgrades won't come effectively. In the event that they are to enhance administer to most patients with interminable disease, the proof firmly recommends that they

reshape their walking look after this reason. Essential care hone was to a great extent intended to give prepared get to and care to patients with intense, shifted issues, with an accentuation on triage and patient stream; short arrangements; analysis and treatment of manifestations and signs; dependence on research center examinations and solutions; brief, instructive patient training; and patient started development. Patients and families battling with interminable ailment have diverse necessities, and these requirements are probably not going to be met by an intense care association and culture. They require arranged, customary collaborations with their parental figures, with an emphasis on capacity and counteractive action of intensifications and complexities. This connection incorporates precise appraisals, thoughtfulness regarding treatment rules, and behaviorally advanced help for the patient's part as self-chief. These associations must be connected through time by clinically significant data frameworks and proceeding with follow-up started by the medicinal practice. [18]

Numerous wellbeing frameworks in Canada are starting to recognize the need to enhance nurture people with endless sickness. Their present framework is basically intended for intense care; dire indications regularly trump the need or longing to bring the patient's incessant infection under ideal administration. As the Institute of Medicine's (IOM) turning point report "Intersection the Quality Chasm" recommended, investing more energy or accomplishing business as usual won't work and just genuine framework change may endeavor to conquer the apparently inconceivable inadequacies in the system.¹ because of this suggestion to take action, the Chronic Care Model (CCM) was produced as a guide to enhance the nature of interminable ailment mind. The CCM does not recommend a handy solution, but rather a multi-dimensional, efficient way to deal with a mind boggling issue. [19]

Decision making in healthcare

Basic leadership in doctor's facilities has developed from being assessment construct to being situated in light of sound logical proof. This basic leadership is perceived as proof based practice. Ceaseless production of new confirmation consolidated with the requests of consistently hone makes it troublesome for wellbeing experts to stay up with the latest. Clinical pathways are archive based devices that give a connection between the best accessible proof and clinical practice. They give suggestions, procedures and time allotments for the administration of particular medicinal conditions or intercessions. Clinical pathways have been actualized worldwide however the confirmation about their effect from single trials is conflicting. This survey expected to abridge the confirmation and evaluate the impact of clinical pathways on proficient practice (e.g. nature of documentation), understanding results (e.g. mortality, complexities), and length of healing center stay and doctor's facility costs. Twenty-seven examinations including 11,398 members were incorporated for investigation. The fundamental outcomes were lessening in-healing facility intricacies and enhanced documentation related with clinical pathways. Entanglements evaluated included injury diseases, draining and pneumonia. Most investigations revealed a diminished length of stay and decrease in healing facility costs when clinical pathways were executed. Significant variety in ponder outline and settings counteracted factual pooling of results for length of stay and healing center expenses. By and large poor revealing kept the recognizable proof of qualities regular to fruitful clinical pathways. The creators reasoned that clinical pathways are related with diminished in-healing facility intricacies. [20]

CONCLUSION

Enormous information in health insurance is the rising examination field; this paper is attempting to introduce a diagram of huge information investigation, its applications in different fields. This paper likewise draws out the difficulties and the exploration openings in huge information.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

None

FINANCIAL DISCLOSURE

None

REFERENCES

- [1] Gang Luo. PredicT ML: [2016] a tool for automating machine learning model building with big clinical data. Luo Health Inf Sci Syst 4:5.
- [2] Marco Viceconti, Peter Hunter, Rod Hose. [2015] Big data, big knowledge: big data for personalized healthcare. IEEE Journal of Biomedical and Health Informatics .
- [3] Tao Huang, Liang Lan, Xuexian Fang, Peng An, Junxia Min, Fudi Wang. [2015] Promises and Challenges of Big Data Computing in Health Sciences. Big Data Research.
- [4] Wullianallur Raghupathi, Viju Raghupathi. [2014] Big data analytics in healthcare: promise and potential. Health Information Science and Systems.
- [5] Kelvin KF Tsoi, Yong-Hong Kuo, Helen M. Meng. [2015] A Data Capturing Platform in the Cloud for Behavioral Analysis Among Smokers. IEEE International Congress on Big Data.
- [6] Muaz Mian, Ankur Teredesai, David Hazel, Sreenivasulu Pokuri, Krishna Uppala. [2014] Work in Progress - In-

- Memory Analysis for Healthcare Big Data. IEEE International Congress on Big Data.
- [7] Ping Jiang, Jonathan Winkley, Can Zhao, Robert Munnoch, Geyong Min, Laurence T Yang. [2014] An Intelligent Information Forwarder for Healthcare Big Data Systems with Distributed Wearable Sensors. IEEE SYSTEMS JOURNAL.
- [8] Prabha Susy Mathew, Anitha S Pillai. [2015] Big Data Solutions in Healthcare: Problems and Perspectives. ICIIACS.
- [9] Raghavendra Kune, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige, Rajkumar Buyya. [2014] Genetic Algorithm based Data-aware Group Scheduling for Big Data Clouds. IEEE/ACM International Symposium on Big Data Computing.
- [10] Rishika Reddy A, Suresh Kumar P. [2016] Predictive Big Data Analytics in Healthcare. Second International Conference on Computational Intelligence & Communication Technology.
- [11] Aronson A R. [2001] Effective mapping of biomedical text to the UMLS Metathesaurus: the Meta Map program. Proceedings / AMIA Annual Symposium. AMIA Symposium, pages 17–21.
- [12] Ayers SL, Kronenfeld JJ. [2007] Chronic illness and health-seeking information on the Internet. Health (London), 11(3):327–347.
- [13] Baeza-Yates R, Calder´on-Benavides L, Gonz´alez-Caro C. [2006] The intention behind web queries. In Proceedings of the 13th International Conference on String Processing and Information Retrieval, SPIRE'06, Berlin, Heidelberg. Springer-Verlag. 98–109.
- [14] Bhavnani SK, Jacob RT, Nardine J, Peck FA. [2003] Exploring the distribution of online healthcare information. In CHI '03 Extended Abstracts on Human Factors in Computing Systems, CHI EA '03, New York, NY, USA, ACM. 816–817.
- [15] Boot CR, Meijman FJ. [2010] Classifying health questions asked by the public using the icpc-2 classification and ataxonomy of generic clinical questions: an empirical exploration of the feasibility. Health communication. 25(2):175–181.
- [16] Boser BE, Guyon IM, Vapnik VN. [1992] A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory. ACM. 144–152
- [17] Broder A. [2002] A taxonomy of web search. SIGIR Forum, 36(2):3–10.
- [18] Cartright M A, White R W, Horvitz E. [2011] Intentions and attention in exploratory health search. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, New York, NY, USA. ACM. 65–74.
- [19] Chee B, Berlin R, Schatz B. [2009] Measuring population health using personal health messages. In AMIA Annual Symposium Proceedings. American Medical Informatics Association.92.
- [20] Chen L, Zhang D, Levene M. [2013] Question retrieval with user intent. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13, New York, NY, USA. ACM.973-976.
- [21] Chen L, Zhang D, Mark L. [2012] Understanding user intent in community question answering. In Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion, New York, NY, USA. ACM.823-828.
- [22] Chih-Chung Chang and Chih-Jen Lin. LIBSVM – A Library for Support
 [23] Vector Machines, April 2013.
- [24] Cho J H, Liao V Q, Jiang Y, Schatz B R. [2013] Aggregating personal health messages for scalable comparative effectiveness research. In Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, BCB'13, New York, NY, USA. ACM.907:907–907:916.
- [25] Cohen J. [1960] A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 20(1):37.
- [26] De Choudhury M, Morris M R, White R W. [2014] Seeking and sharing health information online: Comparing search engines and social media. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, ACM.1365–1376.