Clinical Documents Summarization using Text Visualization Technique

Jonah Kenei
School of Computing and Informatics University of Nairobi
Nairobi, Kenya
Email: Jonah.kenei [AT] gmail.com

Elisha T. O. Opiyo
School of Computing and Informatics
University of Nairobi
Nairobi, Kenya

Juliet Moso
Dedan Kimathi University of Technology
School Of Computer Science & IT,
Nyeri, Kenya

Robert Oboko
School of Computing and Informatics University of Nairobi
Nairobi, Kenya

***Abstract*---- The increased dependency on information systems in all spheres of our lives has led to the exponential growth of electronic data. Therefore, more and more digital information is becoming available and readily accessible for decision making. However, the constantly growing data is difficult to use due to availability of too much information. In computing, this has popularly become known as information overload problem. In recent years, electronic health records have gain wide acceptance in many healthcare delivery settings which is changing the clinical health records documentation in medical practice. Physicians therefore routinely collect and document care of their patients using electronic health records and thus generating digital health records of patients during one or more clinical encounters which can be leveraged to improve healthcare delivery and clinical research. However, the abundant availability of detailed patient information has led to the problem of information overload which makes it difficult to look for information within the health record. It's therefore difficult for users to find and utilize information to improve patient care. In clinical practice, many decisions require a doctor to read and summarize one or many clinical documents which are stored longitudinally, containing various medical assessments or observations per patient. But, there is always a limited time to read long and multiple documents and make decisions during care episodes. Users are therefore faced with an overwhelming amount of information about their patients which requires much time and effort to browse and review and make decisions under time pressure. Despite these deficiencies, visualization of textual documents has proposed as efficient technique for summarizing clinical documents so that it can be easily understood and consumed.**

In this paper we review various text visualization techniques, their applications in summarizing medical documents. Further, we identify research gaps and give an overview of our work in progress to address the aforementioned problem and the gaps. Lastly, this work follows a generalized conclusion on health data visualization and references.

***Keywords*--- Healthcare, Summarization, Visualization, Text, Electronic health records, Physician**

I. Introduction

To diagnose a patient's current illness, physician always reviews the patient's medical history and current health information. The medical history is a written account of patient past medical history which includes symptoms, diagnosis and treatment. This is usually documented as chronological sequence of clinical events. It is generated during one or more encounters with patients and it is often longitudinal. The main purpose of patient medical record is to document the course of patient's health care and to provide a medium of communication among health care professionals for current and future patient care. Medical histories vary in their depth and focus. Therefore, it is typical for a doctor to spend some time reviewing the patient record in order to get a basic understanding of the patient health profile. During care episodes, there can be one or multiple health records for a single patient and physician is required to read carefully and identify important concepts. Therefore, a patient medical history is the most versatile diagnostic and therapeutic tool. However, interpreting patient's medication history from long textual documents can be challenging. Traditionally, records were documented using paper records which according to [1] were hard to read, easy to misplace and sometimes incomplete. Electronic health records on the other hand provide immediate, reliable and readable access to the patient's health history [1]. However, like in paper records, physicians still have to read patient's medical history which is time consuming and error-prone process. According to Nygren et al [2], paper medical records major limitation is the time and effort required to find data items or to gain an overview and, despite the use of electronic health records, the same challenges experienced in using paper-based patient records still persists in Electronic health records[2],[3]. Several research works [4],[5], [6] indicates that with the use of electronic health records, the time a clinician spends interacting with patients on average has decreased while time spent interacting with the EHR has increased therefore affecting the patient time encounter. So it's still difficult to interpret the collected electronic health data and use it to make timely clinical decisions. This information needs to be organized in a way that facilitates efficient retrieval and presentation to the physicians. Despite the abundance of patient information in electronic form, using it to guide patient care still remains a challenge and yet important decisions have to be made, based on the available health history of a patient. With increased use of electronic health records, most of this information is in digital form which can be summarized using text summarization algorithms to help physicians get a glance of the patient's medical history.

In an attempt to solve this problem, many researchers have proposed a text processing technique known as text summarization. Research in automatic text summarization has increased with the rapid growth of the Web and on-line information services, which have made available vast amounts of textual data. Text summarization is the problem of creating a concise summary from a longer text document conveying the same information with the original document. Automatic text summarization is a growing research field which aims at addressing the unprecedented growth of textual documents in many domains. It helps users in discovering and consuming relevant information from longer textual documents without wasting time reading long text documents [7]. Automatic text summarization is an old [8], [[9] and active research field [10] with the aim of producing a concise and fluent summaries while preserving key information content and overall meaning [11]. The idea behind text summarization is the ability to generate a condensed representation of a longer textual document for easier user consumption [12]. The health care domain is a data-intensive industry and in the recent past, there has been unprecented growth of digital health records due to the wide acceptance and use of electronic health records in healthcare delivery. However, large volume of clinical information also renders physician access to patient health information a time-consuming and inefficient process. Electronic health records make access to health records within and across institutions easily and in a timely fashion. Therefore, data that was not previously readily available to physicians are now accessible in a central place.

Developing an automated summarizer that summarizes all relevant information to the attending physician will be very useful. A summary of the patient record is critical for the clinician to see at a glance the patient's medical history and to provide accurate & appropriate ongoing treatment. As an extension of text summarization techniques, information visualization has emerged as an active research field with the aim of providing interactive visual representations of data and information to physicians. This will then help in deepening understanding and exploration of the clinical information space, in order to support optimal use of data and information and also overcome the problem of information overload. Visualization exploits the perception capabilities of human visual system [13] which allows them to quickly understand visual information. Many research studies in other disciplines such as ubiquitous learning [14] show that text visualization has the potential of improving learning. Other research studies such as by Camey [14] shows that people learn better when words and pictures are used together. Mayer [15] proposed that text visualization is useful when textual information is converted into a visual representation while still using words to make it more meaningful.

The benefits of electronic health records can only be realized if data is presented in a format which can be easily understood. Therefore, the effective visualization of patient health data for timely decision making and research is critical. In this paper, we study the various text visualization techniques and its application in summarizing medical documents. Further, this paper identifies research gaps and presents a text visualization design which is work in progress to address the aforementioned problem and research gaps. Our long-term goal is to improve the usability of clinical data sets by developing effective and efficient clinical document visualization model to aid clinicians in clinical documents review. We will then evaluate the effectiveness of this approach using sample documents. Our approach combines textual and graphical summarization approaches to support clinical documents review activities. Lastly, this work follows a generalized conclusion on health data visualization and references.

II. BACKGROUND AND CONTEXT

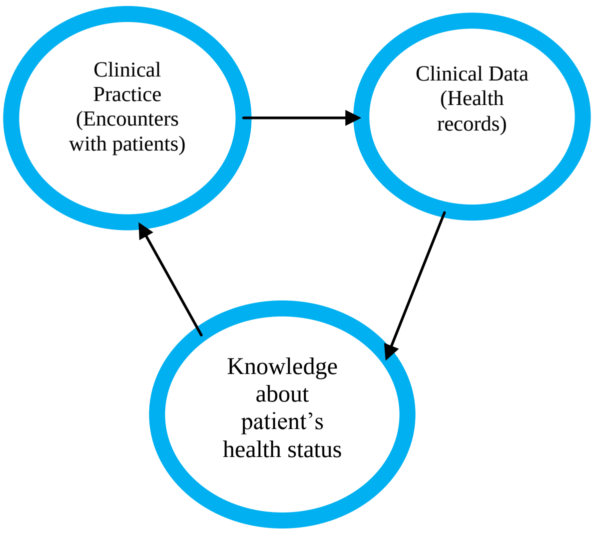


Figure 1: Learning Healthcare System framework

In the last few decades, two major concepts have emerged in healthcare research; data driven healthcare analytics and learning healthcare system. Therefore many research works has been working on applying various techniques such as machine learning, data mining and data visualization to real world healthcare datasets in order to build up a data driven healthcare analytics framework [16]. The concept of Learning Healthcare Systems (LHS) objective is to capture and generate knowledge from the routinely collected data obtained during healthcare delivery. Figure 1 above shows the Learning Healthcare System framework showing clinical practise production and use of health data.

The health care domain is a data-intensive industry and in the recent past, there has been unprecedented growth of digital health records due to the wide acceptance and use of electronic health records in healthcare delivery. Most of these are narrative information obtained from discharge summaries/reports, physicians case notes, pathologists as well as radiologists reports.

In the healthcare domain, physicians often need access to information that answer questions related to health history of a patient. This information generally comes from patient health record. With electronic health records gaining more and more acceptance in healthcare delivery, the task of recording consultations and clinical events is done using electronic health records. This often results in a number of textual clinical documents which are used at the point of care during patients' sub-sequent visits. This has provided physicians with unprecedented digital information about patients' health details. The information contained in the health record is very useful and it allows physicians to determine the patient's health history and provide informed care.

Another emerging trend is using the historical clinical information in clinical research such as population-level studies which can then be used to come up with policies that can improve the quality of care. Therefore the patient health history forms an important part of the management of a patient. Clinical documentation plays a critical role and its importance increases when documentation is communicating clinical and medical related information and data for evaluation, diagnosis and treatment related services (medical and diagnosis reports). The terms medical record, health record, and patient chart are synonymous and describes systematic documentation of a single patient's medical history and care across time within one particular health care provider's jurisdiction [17]. The medical record includes a variety of types of clinical notes entered over time by health care professionals, recording observations and administration of drugs and therapies, orders for the administration of drugs and therapies, test results, x-rays, reports, etc.[17]. Therefore, the increasing use of electronic health records in healthcare delivery, has led to the unprecedented amount of detailed patient data being generated and stored in electronic health records. This has in turn led to the problem of information overload which is increasingly becoming a problem for clinicians [18] [19], [20] and causing challenges for making diagnoses and therapeutic decisions [21]. Therefore, the use of electronic health records has led to a paradox: electronic health records have detailed information of patients' health records, which are scantly used to improve healthcare; and lack of efficient information presentation tools, which make decision making process difficult by the medical practitioners [21]. Electronic Health Records (EHRs) provides a structured approach for the collection, storage, retrieval and analysis of patient health data [22] and also supports sharing of health information among the providers. The sharing of health records also contributes to information overload [23].The increasing volume and variety of electronically available patient information poses significant challenge to physicians in analyzing and interpreting health records, for effective decision making [24]. This therefore demands novel tools to enable efficient review of patients' health records.

Much of the information that is generated in relation to care is stored in electronic health record systems (EHRs) and it is normally in structured or unstructured forms. However, a significant portion (~50%) of the digital health data is in unstructured form [25] stored as clinical notes written on a daily basis documenting care of individual patients and is usually [26] the most important part of the patient health record. For example unstructured data such as clinical notes contain a lot of valuable information about patients, such as diagnosis, symptoms and treatment [27].

Despite the documentation standards such as SOAP which have been generally accepted to write out clinical notes in a patient's chart, the documentation still remains in unstructured form, making it a challenge to extract and understand clinical concepts both within and across patients [26]. It's therefore difficult to extract information from large unstructured free text stored in EHRs. Clinicians often use natural language to document interactions, findings and analyses in patient cases [28]. As a result, finding information from these volumes of unstructured clinical texts requires the use of NLP techniques to automate the extraction process.

Other types of information include images, radiology reports, x-rays, MRI, CT-scan reports, etc. Text documents require reading the source document, in order to get the underlying information. In many cases textual documents are too long to read [29] and therefore text summarization has the advantage since it reduces the reading time of a document [29]. As mentioned earlier, physicians usually refer to patient medical history during care episodes in order to get a basic understanding of the patient health status. However, in many cases health histories can be long especially for patients with multiple health cases. Also, since medical histories are textual documents, it's challenging to read and interpret their contents within a short time period. This problem can be described as information overload problem which is seriously hampering physician decision making.

Effective presentation of clinical data for decision making is a major problem especially when a physician is confronted with large volumes of patient data. Therefore, there is need to tackle this problem using information reduction techniques. More, so systems that automatically generate an overview, or summary, of the information in these health records for both free text and structured information are required. Such systems would enable clinicians to spend more time treating the patients, and less time reading information about the patients.

With the rapidly growing use of electronic health records, the possibility of using text summarization in medical practice to summarize text documents has drawn much attention. It can help overcome the information overload problem by reducing the amount of text that must be read [30] and in obtaining the gist of a given topic of Interest [31] and thus helping physicians and researchers save time and effort required to seek information [32]. There are many research works which have attempted to use text summarization in clinical documents such as [33] and [34].
Using automatic text summarization can help serve clinicians and researchers looking for key information in patients' medical histories. It presents a general overview of the source text, by giving the most important concepts thus allowing users to identify and process relevant information. This can help clinicians in reviewing medical histories of patients [35].
Another emerging approach to summarize and understand textual documents is text visualization which is a sub field of data/information visualization. Just like text summarization, it provides a user with a high level summary which is effective in exploring and analyzing large text documents. According to Nan Cao and Weiwei Cui [36], a good visualization design is able to convey a large amount of information with minimal cognitive effort [35]. Visual analytics is emerging discipline that has shown significant promise in addressing many of the information overload challenges [37].
There are currently many research works, on various visualization techniques based on different types of information that needs to be visualized. Depending on the type of information being visualized, visualization systems can be classified into two categories:

1) Meta-data-based visualization and
2) Content-based text visualization.

Meta-data-based text visualization focuses on visualizing the metadata of text documents. Whereas meta-data-based visualization aids in text analysis, it is difficult to discover deeper knowledge buried in the text. In such cases, visualization fails to perform for those documents having little metadata. This limitation led to the development of content-based text visualization [38].
According to Jason Chuang [39], text visualization suffers from three major problems:

1. High Dimensionality – Textual documents usually have a large number of dimensions or parameters of the data. Visualizing all the dimensions is challenging.
2. Context & Semantics – It's difficult to provide relevant context to aid understanding when textual documents are visualized.
3. Modeling Abstraction – It's difficult to model and abstract textual documents to aid analysis.

III. LITERATURE REVIEW

In the recent decades, text visualization has attracted a lot of research interests due to a number of possible benefits attributed to it; which includes overcoming information overload experienced due to a lot of information in textual documents. Several research studies such as [40] have provided evidence that visualization of clinical information helps in synthesis of patient information from electronic health records clinical documents. Radhakrishnan et al [41] demonstrated that visualizing health data can help in reducing the cognitive load of information, allowing physicians to easily review and interpret large amounts of clinical data.
The use of visual graphical techniques to represent, understand and analyze textual narrative structures is not new. There has been extensive research works on this subject such as [42], [43], [44], and [45]. In the clinical domain there has been a large amount of prior research in visualization of health data, especially the visualization of time-oriented clinical data.
As stated earlier the problem of information overload is increasingly becoming a bottle neck to physicians and researchers intending to use clinical documents for patient care and research respectively. Thus many researchers are attempting to solve this problem from text visualization perspective using different techniques. We divide prior work into three categories namely:

A. History of information visualization
B. Overview of text visualization and
C. Visualization of clinical documents.

A. History of information visualization

William Playfair is believed to be the pioneer of basic graphical visualization of data [46]. He stated that readers' best understand and retain information by graphical representations of data [47]. William Playfair conceived the idea of visualizing data using different types of statistical graphs like line graphs, pie charts, and bar graphs [47]. Since then, data visualization has been researched and studied extensively [49] with the aim of improving its effectiveness and performance. Modern statistical graphs currently in use for different purposes are based on the pioneering works of William Playfair [47]. Most of the Playfair works exploits the potential of using visual graphs to communicate information about quantitative data. These graphs are being used to date in different domains including health as used in [48]. These graphs are popularly used today to represent and give summary information of structured numerical data. In 1858, Florence Nightingale came up with a polar-area diagram which became known as coxcomb chart to demonstrate the relationship between sanitary conditions and soldiers' deaths during the Crimean war [48]. She also pioneered the use of charts to present the statistical data that she collected about causes of war deaths and issues of sanitation and health [49]. Other works include "The Visual Display of Quantitative Information" by Edward Tufte who argued that visualization was more effective way of displaying data [50]. Later on William Cleveland [51] improved and extended data visualization techniques to include statistics. In the late nineties, information visualization emerged in the world of academia as a distinct research field with the collection and publication of previous research works in a book titled "Readings in Information Visualization: Using Vision to Think" by Card et al [52]. Since, then there has been various research works with emphasis on applying information visualization in different domains including healthcare. Card et al [52] defines information visualization as "The use of computer-supported, interactive, visual representations of abstract data to amplify cognition". The objective of visualizing information is to aid in understanding information with visual representations.

B. Overview of text visualization techniques

Text visualization which is a subfield of information visualization [53] is a text processing research field with the main objective of presenting text content of one or many documents in a visual form and thus creating a visual summary [54].

The Linnaeus University Sweden research group called "ISOVIS Group" developed "Text Visualization Browser" which provides the latest and most comprehensive text visualization research works [55]. It gives four different approaches of classifying existing text visualization techniques, i.e., categorizing by task which can either be by analysis or visualization task, by data to be visualized, by application domain, and by the style of visualization design [55]. Another classification of text visualization techniques developed can be made on the basis of goals which these visualization technique aims to fulfill. These include (1) visualizing document similarity, (2) revealing the content of textual information in a document, and (3) displaying sentiments and emotions in a text document [36].

I. Visualizing Document Similarity

Many text visualization techniques were designed to display similarities of content in a textual document. It is an old text visualization technique in which documents are visualized as points on a low-dimensional (2-Dimension or 3-Dimension) visualization plane. Other similar research works have been extensively studied and categorized into two categories namely projection-oriented and semantic-oriented [56].

i. Projection Oriented Techniques

In Projection oriented technique, the dimensions in the text document is first reduced using dimension reduction technique before being visualized. Projection Oriented technique, characterizes a document as a bag of words and it is represented as an N-dimensional feature vector. To calculate this vector, a set of most informative words W (|W| = N) that best differentiates each document (i.e., best captures the characteristics of different documents "Term Frequency Inverse Document Frequency (TF-IDF)" forms an approach to differentiate a particular document from others. As a renowned numerical statistical method designed to help extractive word features for document classification [57], the process computes a TF-IDF score for each word in the document collection. The word with a higher score is considered to be more informative, i.e., more useful than other words for classifying different documents. Generally, the projection oriented techniques can be further classified into (1) linear projections (i.e., linear dimension reduction) and (2) non-linear projections.

Both techniques could be formulated in consistent form in which pair-wise distances between data items are maximized and guided by weights that indicate the importance of separating pairs of points in the results [58].

ii. Semantic Oriented Techniques

In semantic oriented techniques, document similarity is characterized using latent topic information i.e. latent topics retrieved from textual documents. In this technique each document is viewed being made up of different topics and each document is considered to have a set of topics that are assigned to it using Latent Dirichlet Allocation (LDA). This technique is based on the theoretical foundation of topic modeling techniques.

It is not intended for visualization purposes and it is mostly used for analysis purposes [36]. There have been some efforts to improve these techniques by producing improved visual representations. These procedures which were initiated by Probabilistic Latent Semantic Visualization (PLSV) [59] implants the latent topics and documents within the generic Euclidean space simultaneously because of the fact that document similarity is directly encoded through the distances among the topics being shared.

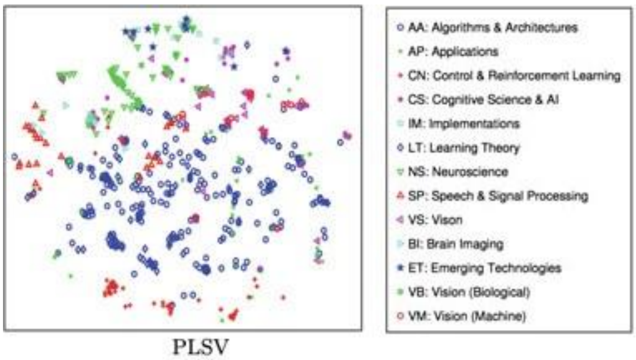


Figure 2. Visualizing document similarity based on semantic oriented techniques PLSV

II. Revealing Text Content

The most crucial step in text visualization is to represent the words in a text document visually. Many different approaches of text visualization attempts to show documents' contents using different aspects and different levels of details. This includes, summarization of a single document, viewing the words and topics, detection of events, and creation of storylines [36].

i. Summarizing a Single Document

The summarization of a document through the prevalent visualization approaches involves two prime phases: (1) substances such as words and figures and (2) characteristics for instance average sentence length and quantity of verbs. While demonstrating what a document contains, Collins et al. [60] used DocBurst that creates a tree like structure by breaking the entire document into its components. These are demonstrated in SunBurst visualization [61] (Figure 3). Rusu et al. [62] has demonstrated a node-link diagram consisting of the document constituents. This diagram is based on the extraction of semantic graph from the document. Strobelt et al. [63] introduced a system transforming the entire document into a card wherein the constituents of a document are presented in a summarized manner through the extracted keyword and figures (Figure 3). Stoffel et al. [64] offered an approach wherein a thumbnail for the document is produced based on the deformation of keywords. This technique produces a focus and context (together) representation of document at the page level.

Figure 3. The outcome of the search query for words that begin with pl using DocuBurst visualization. The results which match the query are marked with gold.

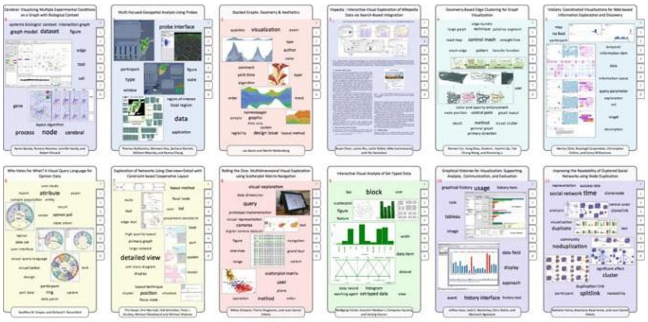


Figure 4. Transformation of the entire document into cards

ii. Showing Content at the Word Level

The prevailing visualization approaches within this classification were established for providing the solution to three main problems: (1) how to represent the words aesthetically in a visual form to clearly depict the content of the text; (2) how to summarize and represent the semantic relationships such as "A is B" and "A of B" amongst the textual words and (3) the way for revealing word-level patterns for example recurrences and co-occurrences. TagCloud [65] has emerged to be a spontaneous and universally practiced methods for visualizing words. This technique generates a summary for the input content from the bag of words and inputs them in the cloud where the words, with font size representing their own significance are packed together without any overlap.

iii. Visualizing Topics

An increasing number of visualization techniques were developed to:

i) Summarize and explore static topic information,
ii) Illustration of the topic dynamics with time,
iii) Assist in the comparison of topics
iv) Demonstrate the occurrences of events and storylines.

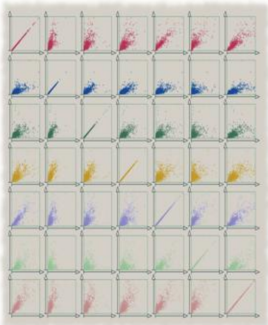
Capturing the topic dynamics is another research direction that has attracted great attention in the area of text visualization. In particular, Theme River [66] is a premature techniques developed to show the frequency of the keywords that changes with time. It visualizes a set of keywords as stripes (i.e., themes) with their wideness altering with time thereby demonstrating the alterations in the frequencies of the keywords.

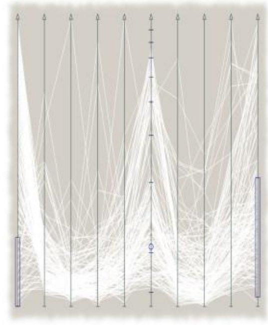
C. Text visualization of clinical documents

In the healthcare domain, a wide variety of data visualizations techniques have been successfully used in visualization of medical images from x-rays machines, CT scans and MRIs images(medical images produced by Magnetic Resonance Imaging (MRI) and Computed Tomography (CT)) which can be visualized in three dimensional view by applying computer graphics. The visualizations of these images provide accurate three dimensional volume visualizations that highlight problem areas in different parts of interest [67]. Other visualization works include, surgical planning [68], telesurgery, pharmaceutical drug discovery, chem-informatics and genomic expression analysis [69].

In the recent years there have been some research works on visualizing textual health documents which mainly targets patients' medical histories available in electronic form. Most of these have concentrated on visualization of temporal information in Electronic health records (EHRs) [70], [71], [72], and [73]. Timelines is the most widely used technique used in visualizing temporal events of a patient's health history. A timeline is defined as a visual representation of chronological events on a line. There are many examples reported in the literature on the use of TimeLine technique to visualize electronic health records as described in [74, 75]. The use of timelines for visualizing health records was proposed by Powsner and Tufte [76] who designed a one page graphical summary and mapping medical findings and treatments over time. Timelines main idea involves developing a graphical summary using a table of individual scatterplots of relevant medical variables such as test results and treatment data. Other research works in the literature using timelines to visualize health data include Ozturk et al [77] who developed a visualization tool based on timeline which visualizes patient's medication history and also converts a patient prescription data into simple timeline visualization. Other recent examples of temporal data visualizations include [78], [79] and [80]. The classic example is in [81] proposed a visualization tool call HARVEST". HARVEST is a longitudinal patient record summarization system which visualizes patient health information using a timeline. HARVEST works by extracting medical content from a patient's longitudinal documentation and aggregates information from multiple care settings. Finally, it then visualizes content through a timeline of a patient's problem documentation and clinical encounters [81].

Most healthcare data are multivariate in nature and researchers have also developed techniques for visualizing multivariate data. The two common techniques use scatter plot matrix [82] and Parallel Coordinates [83]. The figure (Figure 5) below shows Scatter Plot Matrix and Parallel Coordinates.





Scatter Plot Matrix **Parallel Coordinates**

Figure 5. Scatter Plot Matrix and Parallel Coordinates. (Adapted from [84])

Scatter Plot Matrix and parallel coordinates have found practical applications in visualizing electronic health data e.g. comparing diagnoses and showing correlations among different diagnoses [84]. Borland et al [85] used scatter plot and parallel coordinates to visualize multivariate clinical data. They developed a visualization tool for visual exploration and analysis of longitudinal clinical data. Their tool was mainly used for visualizing clinical data for large cohorts of patients. The tool was useful in identifying and selecting temporal patterns of interest, and also displaying various relationships and patterns of clinical data. The table below summarizes Borland et al [85] visualization work.

Table 2: Borland et al [85] five views of the data visualization

Visualization technique	Description
Scatter plot and Parallel coordinates	Provide multivariate visualizations of multiple descriptive features extracted from the longitudinal data
A hexagonal binning	Provides an aggregated visualization of the temporal distribution of data points
An icicle plot variant	Visualizations of diagnoses
Parallel-sets	Visualizations of Demographic data

Another practical application is the work of Dogu and Nembhard [86] who studied the use of scatter plot matrix, star glyph plot and Andrew's curves in visualizing and monitoring multivariate clinical hypertension data. They found out that; incorporating glyphs together with multivariate control charts helped users visualized data easily and also allowing users to monitor the univariate and multivariate process information simultaneously and hence improving the decision making ability and supporting the use of clinical information more effectively[86].

Visualization and representation of multivariate data with respect to geographical location has also been found to be useful in facilitating data driven knowledge discovery. Karol et al [87] designed Community Health Map to explore and visualize health care data with respect to geographical region. Users can therefore be able visualize health care data for any given region of their interest. Other similar works include Cartographic Treemaps [88], [89].

Another variant of TimeLine is Lifelines which uses horizontal bars to represent the temporal location and duration of data [90] and it is used to present the history of a patient's medical record. Plaisant et al [90] designed and implemented a visualization tool called LifeLines, and tested it using real clinical data. The main idea behind this approach is to use familiar metaphor of timelines, and taking the advantage of human beings ability to visually analyze information abundant displays, and facilitating access to the details in the record.

Lifelines evolved to become Lifelines2 [91], a visualization tool using categorical point event data across multiple records. Unlike Lifelines which only visualized a single patient record at a time, Lifelines2 is able to visualize records of multiple patients [91]. Another interesting visualization system is Midgaard [91] which visualizes patient data by mapping it to a human body template. It was designed mainly for ICU data [101]. Midgaard is also an extension of Lifelines with a semantic zoom chart technique that shows more details as the chart is enlarged [92]. Another emerging powerful visualization tool is Eventflow [93] developed by the University of Maryland Human Computer Interaction Lab (HCIL). It is similar to Lifeline2 [93] and it helps in visualizing and reviewing the data from individual records and their event sequences [92]. It summarizes and displays time-oriented as well as interval data [94]. Lifelines2 only handles data with a single timestamp. In many cases there are situations where data have multiple time intervals (i.e. with a start and end time). EventFlow therefore addresses the shortcoming of Lifelines2 by supporting the visualization of data with time intervals. In addition it provides advanced query capabilities [94]. EventFlow has been successful and there are many practical applications from the literature in visualizing health data using EventFlow such as [95], [96], and [97].

There are other several visualizations techniques reported in the literature such as clinical pathways [95] and [96]. Zhang, Yiye et al [98] proposed a practice-based clinical pathway development process and a data-driven methodology for extracting common clinical pathways from electronic health record (EHR) data that is patient-centered and consistent with clinical workflow, and facilitates evidence-based care. Bettencourt-Silva et al [99] proposed a data driven clinical pathways generated from routinely collected hospital data. It was found out that, the clinical path way can help in revealing information about patients and diseases. It also helps in summarizing, visualizing, and querying of patient centric clinical information, as well as the computation of quality indicators and dimensions [99].

Others approaches include use of body-centric data layouts and flowcharts. Body-centric data layouts such as by Zhang, et al [100] use body-centric data layouts, which takes advantage of clinical data which usually have some relation to human body anatomy. A template of a virtual human body is used which provides an index to the corresponding part of the human body [100]. Figure 5 illustrates an example of body centric layout showing a human body and medical problems in different parts of the body. Midgaard [101] uses this concept to map collected patient records to an image of the human body. The Five Ws for Information Visualization [102] has also been used to visualize patient records using a radial display integrated on a human body map. It helps physicians by giving them the ability to mark the location of symptoms in the body and present the corresponding diseases in a sunburst tree around the body map. The display is used to capture the past and present health conditions of patients and it can give attending physician a summarized overview of a patient's health status [102].

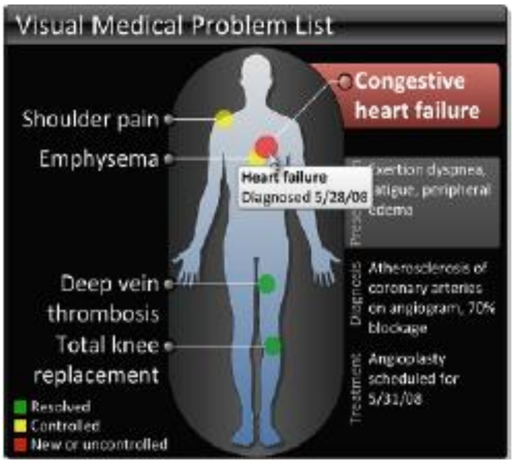


Figure 5: A map using a human body as a visual index for medical problems. (Adapted from [103]

LifeFlow [104] is another recent visualization tool that provides an overview of event sequences to support users' exploration [104]. In Electronic Health Records (EHRs) event sequences include doctor visits, lab results, medication orders, and transfer among hospital departments which are series of time stamped events [104].

Borland et al [105] developed a multivariate visualization system using Radial Coordinates. They applied real data sets from the National Health Service (NHS) in the United Kingdom (UK) to demonstrate lung cancer prevalence and lung cancer rates. One unique characteristic of this approach is its ability to visualize continuous, discrete, and categorical data types. Lee et al [106] proposed a service-oriented architecture that supports visualization of patient population data sets as a service [106]. Kumar et al [107] developed a visualization chart called Patient Journey Visualizer (PJV), a visualization tool used for visualizing patients' medical histories over time called "Patient Journeys". It uses Parallel Coordinates, Sankey charts, and Sunburst charts. Perer and David [108] designed a visualization tool called "CareFlow" to help doctors and patients communicate about possible care plans and their outcomes. CareFlow allows physicians and patients to understand which treatments have historically been most effective. They used Sankey charts. Krause et al [109] designed a tool called, "Patient-viz," where patient data is visualized and used to tell the story of a patient by using a visual representation of data using charts over time. Perer and Sun [110] designed a matrix base approach called "MatrixFlow" which tracks symptoms evolution during disease progression. It uses matrix-based icons to provide visual links and comparisons between multiple events. It works by taking clinical event sequences of patients as input, constructs time-evolving networks and visualizes them as a temporal flow of matrices. Goovaerts [111] studied the visualization of spatial and temporal data to address the problem where spatial and temporal data were being displayed in separate views. Previous tools were not taking advantage of the human visual processing engine to extract knowledge from the spatial interconnectedness of information over time and geography [111]. He reviewed different approaches for the space-time visualization of health data with the aim of finding solutions for the 3D interactive visualization of health outcomes in a combined time and geography space and contextualization of disease maps through the incorporation of familiar markers, such as highways, rivers or topographic details [111]. Rind et al [112] designed and evaluated a visual exploration tool known as "VisuExplore" which visualizes time oriented data to support health care and medical review of patients with chronic diseases. One distinguishing feature of this tool was the ability to allow users interact with the system hence giving users techniques for effective exploration of patient information. The tool was designed and evaluated in close collaboration with the end users (physicians).

Table 2. Summary of visualization systems for medical documents [113, 114, 115, 116, 117, 118, 119, 120, 121,122,123,124,125,126]

Visualization type	**Reference**
Timeline	A visualization technique that is a problem-centric which is used to visualize temporal patient health records [113].
Lifelines	A visualization technique that uses a timeline visualization technique to represent patient health history [114]. Provides timeline of a patient's temporal events.
Lifelines2	An improvement of Lifelines, which gives the user the ability to analyze records of multiple patients [115]. It is used for analyzing large number of patient records trends.
CLEF	A data visualization framework called Clinical e-Science Framework (CLEF) used for browsing health histories, while integrating visual navigation tools and the ability to generate textual summaries automatically [116][117] and semantic network of health record events
KNAVE-I	A visualization framework called Knowledge-based Navigation of Abstractions for Visualization and Explanation (KNAVE) used for interactive visualization and exploration of time-oriented clinical data and semantically-related concepts [118].
KNAVE-II	An extension of KNAVE-I used for visualization and interactive exploration of time-oriented data at different levels of temporal abstractions[119], [120]
Asbruview	A visualization system which represents treatment procedures as structured time-oriented plans [121].
Radial Starburst	A visualization system with the ability to represent data with over 100-dimensional space[122]
VISITORS	Visualization of Time-Oriented Records (VISITORS) is a visualization system which has evolved from KNAVE/KNAVEII which supports diverse temporal data from multiple records [123]. Temporal data from multiple records. Usability testing found the system feasible for exploring longitudinal data for quality or clinical results.
DICON (Dynamic Icon)	A visualization system which interactively explores clusters of similar patients and representing clusters as icons using a treemap[124].
Outflow	A visualization system which visualizes disease progression paths and allowing users to look at a visual display consisting of multiple events, their sequences, and outcomes[125]
EventFlow	A visualization system which visualizes treatment patterns and outcomes[126]

Several clinical data visualization techniques have been extensively investigated and some have been successfully deployed in clinical practice. The propose techniques have shown the potential of having a tremendous impact in clinical practice in terms of assisting physicians in diagnosis and treatment of diseases. Most successful text visualization methods that have been employed for clinical data visualization fall into four categories:

- Timeline-based methods for visualizing temporal events of a patient health history.
- Body centric layouts methods which maps patient data to a human body template.
- Data plot charts methods which visualize multivariate clinical data and
- Map based methods which visualizes multivariate clinical data with respect to geographical location.

Timeline, Lifelines, LifeLines2, KNAVE I, KNAVE II], LifeFlow, OutFlow, EventFlow use timeline-based tools to organize and display health records. They use dots to indicate time events and line segments to represent periods at which medical events occur along a horizontal time axis.
Hovering a mouse pointer over a line or double clicking on any line enables access to more details, while graphical attributes, such as color and line thickness, improve system display. Body centric layouts methods examples includes: The Five Ws system and Midgaard which uses body centric layouts by mapping patient data to a human body template.
Data plot charts are used for visualizing multivariate health data. Examples include Scatter Plot Matrix and Parallel Coordinates. Map based techniques visualizes multivariate and geographically distributed health datasets. Examples of map based techniques include Cartographic Treemaps, Community health map.

IV. GAPS IN THE LITERATURE

Unstructured electronic health documents such as clinical notes present enormous opportunities in healthcare delivery and also for knowledge discovery through secondary uses such as data mining [127], [128] big data analytics [129], [130] Natural language processing [131], [132] clinical text analysis [133] and data visualization [134], [135] and the emerging commercial software applications such as Tableau [136] used for visualizing real world health data.
Previous studies have shown that, summarizing patient health records and presenting a structured overview of patient medical history can have a positive impact on overall patient care [137].
When such data is analyzed using various visualization tools, privacy disclosure risks arise, making privacy preserving text visualization a fundamental requirement. However, existing visualization research works have not addressed this requirement. While some solutions to this problem exist in other areas such as data publishing [138] and data mining [139], there is little research in privacy preserving data visualization. Therefore exploratory visual analytics tools for healthcare data which take into consideration patient privacy is needed to fill this gap.

From the literature, it can be noted that the mostly well studied area is the visualization of temporal data using variants of timelines and extension of the Lifelines such as LifeLines2, EventFlow, and LifeFlow. The second mostly studied approach uses body centric layouts which maps patient data to a human body template. Also, initial, research works focused mainly on visualizing records of a single patient and the same has been extended to records of multiple patients and other complex data. The above two techniques however ignores the fact that capturing and visualizing semantics and semantic relationships in clinical records is equally important. The second major limitation is the inability of supplementing visual navigation with textual explanations, which improves the clarity of the visual summary.

From the foregoing discussions it’s irrefutable that existing health visualization techniques suffer a combination of the following issues:-

a) Inability of the current visualization models to address the problem of preserving patient privacy in electronic health records [140]. Current visualization techniques assume that there is unrestricted access to the underlying data [135]. From the literature, visualizing electronic health records have useful practical applications in healthcare care delivery and research, however visualizing this information which is sensitive in nature is a problem in real-world applications. Using visualization in healthcare delivery may not be a problem, however when it’s used for research purposes, the patient privacy problem comes into focus. Privacy-preserving data visualization is a relatively new area of research compared to the more established research areas of privacy-preserving data publishing [138] and data mining [139] and the nascent works such as [140], [141], [142] and [143] aims at addressing this problem.
b) Most of the electronic health data are unstructured health documents which just like any other text document have high dimensional nature [138]. It’s still challenging to design effective visualization systems to represent large corpora of text due to the unstructured and high-dimensional nature of text [144]. Data with high dimension is a major challenge to visualization researchers [145].
c) Lack of semantics in text visualization - Lack of semantics in text visualization [145]. Most medical applications visualize patient data without integrating additional semantic information to structure the analysis [146]
d) Several information visualization toolkits and tools have been developed to facilitate users work. However, evaluation studies for these toolkits and

tools from a user perspective have been overlooked [147].

V. WHY VISUALIZATION OF CLINICAL DOCUMENTS

In recent years, electronic health records have gain wide acceptance in many healthcare delivery settings. Physicians therefore routinely collect and document care of their patients using electronic health records and thus generating digital health records of patients during one or more clinical encounters in any care delivery setting. The digital health record is a myriad of patient information such as patient demographics, problems, symptoms, diagnoses, progress notes, treatments, medications, vital signs, past medical history, immunizations, laboratory data, radiology reports, etc. [148]. There is great potential in leveraging these large amounts of clinical data in healthcare delivery [149] and clinical research [150]. This therefore demands tools that can assist clinicians and researchers in automatically condensing records to provide succinct summaries of a patient's medical history. Research works such as [151], [152] demonstrates the potential of health data visualization in helping doctors during care episodes.

Visualization of clinical information has the potential to address information overload problem by supporting clinicians to understand EHR data visually. According to [153], several research studies have been done to address two visualization goals; 1) to provide patient data to support physician's medical decision making at point of care, and 2) to provide population data to help establish new clinical knowledge [153].

Electronic health records (EHRs) which is defined as "a repository of information regarding the health status of a subject of care, in computer process able form" [154] has the advantage of providing increased access to lots of patient information. This information is both in structured and unstructured text-documents [155]. However, the abundant availability of detailed patient information has led to the problem of information overload in healthcare delivery where physicians have access to a lot of information to make decisions from [156], which can negatively affect physician ability to identify important clinical data, and may also lead to medical errors [157], delays [158], [159] and the general patient safety [148]. Physicians suffers cognitive overload due to a lot of electronic health record documentation [160]. Much of the information is available in unstructured (free) text and thus the clinical documents are often lengthy with data redundancy problem which makes it difficult for physicians to navigate and synthesis information during time-constrained patient care [161], [162], and [163]. Without this comprehensive, accessible, information, physicians may not be able to deliver quality healthcare. On the other hand, when there is too much information in a time constrained environment, the impact on physician performance is exacerbated. According to Chittaro [164] humans have cognitive and perceptual limitations and therefore the quantity of information a user can examine and handle at a given instant is very limited. Research works confirms that information visualization help to reduce cognitive information overload [165] which is experienced by users when reading long textual documents. It is generally accepted that using images along with text help people understand and remember text information more easily [166].

VI. AN OVERVIEW OF OUR APPROACH

Building upon the concepts on the existing literature and the research gaps identified in the literature, we now describe an ongoing project of implementing a privacy preserving visual analytics model. In this section, we only highlight a general overview of our proposed work. Our work introduces an approach for summarizing clinical documents and presenting it in an intuitive graphical structure for easy consumption by the intended users.

In this paper we propose an approach for using text visualization technique to summarize and visualize clinical documents to aid physicians and researchers in navigating textual documents. We use a medical chart document and Kenya health setting as a case study. Our approach can be modeled in the following five steps:

I. Anonymizing (de-identification)

The objective of this step is to remove individually identifiable Information from the medical chart with the aim of preserving patient privacy. Specifically, the following information will be removed:

i. Names
ii Dates,
iii Places

The first step in this task will be training the system to identify string segments containing names, dates or places. To achieve this, the system will be trained using n-grams (specifically five-grams) and a Bayesian classifier.

To minimize the need to manually annotate training data, several automation techniques will be employed:

i. Names - To generate the training dataset containing n-grams with names, a large, pre-existing dataset of names of Kenyans will be used. The dataset was generated from publicly-available records from the I.EB.C voter register, KRA register of PINs, HELB records and KNEC results data. Collectively, these datasets have millions of records.
ii. Dates- Dates generally have a certain layout for example 26/10/2017 or Thursday, October 26th, 2017. By running a regular expression engine against the texts, a large collection of

ham n-grams will be created for training purposes.

iii. Places - As with names and dates, there's need to generate a large dataset of ham n-grams with a place. This will be achieved first by using Google Maps Place Search API to generate a large collection of place names in Kenya. Next, the clinical texts will be searched for any n-grams containing any of the known place names. This dataset will constitute the ham dataset.

II. Detecting Keywords - In this step, the system will be trained to identify keywords from clinical texts. The first step will involve employing a Part of Speech tagger to eliminate all non-verb and non-noun tokens. Next, the Brown Corpus will be used to eliminate the most common words. Also, the system will use the WordNet synset database to eliminate less common words with a similar meaning to the most common words. Finally, we group semantically related keywords from medical charts (e.g. symptoms might be grouped together, followed by diagnosis then medication), the keywords extracted will be grouped into the following categories:

i. Symptoms
ii. Diagnosis
iii. Treatment

The categories dictate where they are displayed to the user. We developed the groupings (i.e., dimensions) on the basis of the types of information that would be of interest to physicians. To ensure accuracy of the system, there might be need to manually verify the resultant dataset.

III. Create the word net
Based on the keywords detected in step (II) above, a word net of medical terms will be created. For example, 'headache' might be classified as a hyponym of 'pain'. Primarily, this will be built on top of the existing open-source WordNet database. Where necessary, terms will be re-classified by an expert

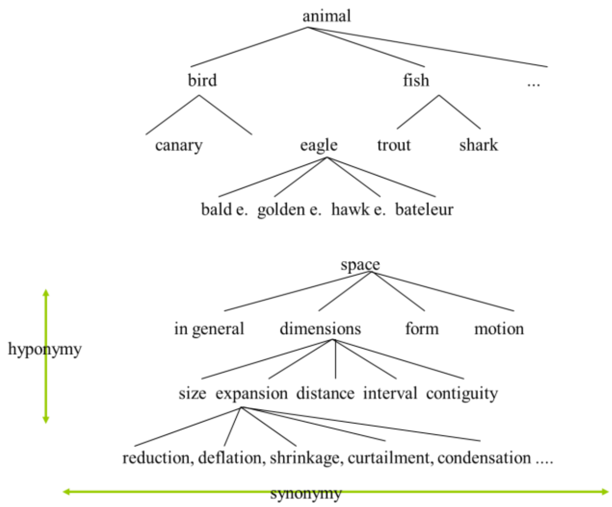


Figure 6: Sample WordNet

IV. Create semantic maps
Based on the WordNet created, for each chart, we generate a graphic semantic map showing the most important terms and how they relate to each other. A sample of the desired output is shown is Figure 7.

Figure 7: Sample Visual summary illustration

VII. CONCLUSION AND ONGOING WORK

Although, Electronic health records have the potential of improving diagnosis by making it easier to collect, store, organize and retrieve patient data in a more organize way, reviewing clinical documents in electronic health records is a tedious and time-consuming process. The unstructured nature of data available makes it hard for the physician to recognize important concepts in textual documents and the progression over time. Therefore, the need to have effective support to access, retrieve and understand electronic clinical documents is necessary. As patient data are collected over multiple visits, visualization tools will greatly help physicians to quickly navigate the general contours of patient health history, looking for important medical concepts, and then drilling for details in order to understand the health profile of a patient. From the literature, information visualization continues to be acknowledged as the most innovative approach to support healthcare delivery and clinical research in overcoming the problem of information overload due to the explosion of electronic health data.

We have surveyed various research works in the literature and identified important findings that can help guide our current research in the area of clinical documents visualization. We have also observed several research gaps in the literature and described our proposed model to address the gaps observed. For this purpose we propose to use semantic map models that are automatically created from clinical documents. Our proposed model visualizes text documents allowing users to visualize and navigate textual documents. By visualizing health documents, we are able to support physicians in reviewing medical histories of patients and also researchers interested in conducting studies on healthcare data. In our current work, we intend to evaluate the performance of our model in terms of effectiveness to de-identify personal identifying information. We will also evaluate the usability of our model with healthcare professionals to ascertain the effectiveness and usability in practical situations.

REFERENCES

[1] Graber, M.L., Byrne, C.M., & Johnston, D. (2017). The impact of electronic health records on diagnosis. Diagnosis, 4 4, 211-223

[2] Nygren, E., Wyatt, J., & Wright, P. (1998). Helping clinicians to find data and avoid delays. Lancet, 352 9138, 1462-6.

[3] Bui AAT, Hsu W, Taira RK. Medical Data Visualization: Toward Integrated Clinical Workstations. Berlin: Springer-Verlag Berlin; 2009: 139–193

[4] Gilchrist, V., McCord, G., Schrop, S. L., King, B. D., McCormick, K. F., Oprandi, A. M., … Zaharna, M. (2005). Physician Activities During Time Out of the Examination Room. Annals of Family Medicine, 3(6), 494–499. http://doi.org/10.1370/afm.391

[5] Laxmisan, A., Hakimzada, A.F., Sayan, O.R., Green, R.A., Zhang, J., & Patel, V.L. (2007). The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. International journal of medical informatics, 76 11-12, 801-11.

[6] Weigl, M., Müller, A., Zupanc, A., & Angerer, P. (2009). Participant observation of time allocation, direct patient contact and simultaneous activities in hospital physicians. BMC Health Services Research, 9, 110. http://doi.org/10.1186/1472-6963-9-110

[7] Horacio Saggion, Thierry Poibeau. Automatic Text Summarization: Past, Present and Future. T. Poibeau; H. Saggion. J. Piskorski, R. Yangarber. Multi-source, Multilingual Information Extraction and Summarization, Springer, pp.3-13, 2012, Theory and Applications of Natural Language Processing, 978-3-642-28569-1.

[8] Luhn, H. The automatic creation of literature abstracts. IBM Journal of Research and Development 2(92) (1958) 159-165

[9] Sparck-Jones, K. Automatic summarizing: factors and directions. In Mani, I.; Maybury, M. Advances in Automatic Text Summarization. The MIT Press (1999) 1-12

[10] Patil, M.P.D., Kulkarni, N.J.: Text summarization using fuzzy logic. Int. J. Innov. Res. Adv. Eng. (IJIRAE) 1(3), 42–45 (2014)Google Scholar

[11] Allahyari, M., Assefi, M., Gutierrez, J.B., Kochut, K.J., Pouriyeh, S.A., Safaei, S., & Trippe, E.D. (2017). Text Summarization Techniques: A Brief Survey. CoRR, abs/1707.02268.

[12] Mani, I. Automatic Summarization. J.Benjamins Publ. Co. Amsterdam Philadelphia (2001)

[13] Langelier, G., Poulin, P., & Sahraoui, H.A. (2005). Visualization-based analysis of quality for large-scale software systems. ASE.

[14] Carney, R. N. and Levin, J. R. Pictorial Illustrations Still Improve Students' Learning From Text. Educational Psychological Review, vol. 14, no. 1, pp. 5-26, 2002.

[15] Mayer, R.E. Multimedia Learning. Cambridge University Press, 2nd edition, 2009

[16] [10] Hu J., Perer A., Wang F. (2016) Data Driven Analytics for Personalized Healthcare. In: Weaver C., Ball M., Kim G., Kiel J. (eds) Healthcare Information Management Systems. Health Informatics. Springer, Cham

[17] CMS,(2011) Personal Health Records. Retrieved from https://www.medicare.gov/Publications/Pubs/pdf/11397.pdf

[18] Hall, A., Walton, G.: Information overload within the health care system: a literature review. Health Information & Libraries Journal 21(2) (2004) 102–108 2. Van Vleck, T.T., Stein, D.M.,

[19] Stetson, P.D., Johnson, S.B.: Assessing data relevance for automated generation of a clinical summary. In: AMIA Annual Symposium Proceedings. Volume 2007., American Medical Informatics Association (2007) 761

[20] Ammenwerth, E., Duftschmid, G., Hübner-Bloder, G., Kohler, M., Rinner, C., & Saboor, S. (2013). The EHR-ARCHE project: Satisfying clinical information needs in a Shared Electronic Health Record System based on IHE XDS and Archetypes☆. I. J. Medical Informatics.

[21] Thiessard, F., Mougin, F., Diallo, G., Jouhet, V., Cossin, S., Garcelon, N., Campillo-Gimenez, B., Jouini, W., Grosjean, J., Massari, P., Griffon, N., Dupuch, M., Tayalati, F., Dugas, E., Balvet, A., Grabar, N., Pereira, S., Frandji, B., Darmoni, S.J., & Cuggia, M. (2012). RAVEL: Retrieval And Visualization in ELectronic health records. Studies in health technology and informatics, 180, 194-8.

[22] Nair, V., Kaduskar, M.V., Bhaskaran, P., Bhaumik, S., & Lee, H. (2011). Preserving Narratives in Electronic Health

Records. 2011 IEEE International Conference on Bioinformatics and Biomedicine, 418-421.

[23] Ammenwerth, E., Duftschmid, G., Hübner-Bloder, G., Kohler, M., Rinner, C., & Saboor, S. (2013). The EHR-ARCHE project: Satisfying clinical information needs in a Shared Electronic Health Record System based on IHE XDS and Archetypes☆. I. J. Medical Informatics.

[24] Max Craft,Bev Dobrenz,Erik Dornbush, "An assessment of visualization tools for patient monitoring and medical decision making," 2015 Systems and Information Engineering Design Symposium, Charlottesville, VA, 2015, pp. 212-217.doi: 10.1109/SIEDS.2015.7116976

[25] Hicks 2003. Hicks J. 2003. The potential of claims data to support the measurement of health care quality. San Diego, CA: RAND; 2003.

[26] Sondhi, P., Sun, J., Tong, H., & Zhai, C. (2012). SympGraph: a framework for mining clinical notes through symptom relation graphs. KDD.

[27] Ling, Y., An, Y., & Hu, X. (2014). A matching framework for modeling symptom and medication relationships from clinical notes. 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 515-520.

[28] Nair, V., Kaduskar, M., Bhaskaran, P., Bhaumik, S., and Lee, H. Preserving narratives in electronic health records. In Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on (Nov 2011), 418–421.

[29] Torres-Moreno, J. (2014). Automatic Text Summarization. Cognitive science and knowledge management series. Wiley. https://books.google.es/books?id=aPHsBQAAQBAJ.

[30] Reeve LH, Han H, Brooks AD. The use of domain-specific concepts in biomedical text summarization. InformationProcessing & Management. 2007;43:1765-76.

[31] Mishra R, Bian J, Fiszman M, Weir CR, Jonnalagadda S, Mostafa J, et al. Text summarization in the biomedical domain: a systematic review of recent research. Journal of biomedical informatics. 2014;52:457-67.

[32] Moradi, Milad & Ghadiri, Nasser. (2016). Different approaches for identifying important concepts in probabilistic biomedical text summarization. Artificial intelligence in medicine. 84. . 10.1016/j.artmed.2017.11.004.

[33] Rimma Pivovarov, Noémie Elhadad; Automated methods for the summarization of electronic health records, Journal of the American Medical Informatics Association, Volume 22, Issue 5, 1 2015, Pages 938–947, https://doi.org/10.1093/jamia/ocv032

[34] Hans Moen, Laura-Maria Peltonen, Juho Heimonen, Antti Airola, Tapio Pahikkala, Tapio Salakoski, Sanna Salanterä, Comparison of automatic summarisation methods for clinical free text notes,Artificial Intelligence in Medicine,Volume 67,2016, Pages 25-37,ISSN 0933-3657,https://doi.org/10.1016/j.artmed.2016.01.003.

[35]] Mishra, Rashmi, Jiantao Bian, Marcelo Fiszman, Charlene R. Weir, Siddhartha Jonnalagadda, Javed Mostafa, and Guilherme Del Fiol. "Text summarization in the biomedical domain: A systematic review of recent research", Journal of Biomedical Informatics, 2014.

[36] Nan Cao, Weiwei Cui. Introduction to Text Visualization. Atlantis Briefs in Artificial Intelligence 1, Atlantis Press 2016, ISBN 978-94-6239-185-7

[37] Jesus J Caban, David Gotz; Visual analytics in healthcare – opportunities and research challenges, Journal of the American Medical Informatics Association, Volume 22, Issue 2, 1 March 2015, Pages 260–262, https://doi.org/10.1093/jamia/ocv006

[38] Shixia Liu, Michelle X. Zhou, Shimei Pan, Weihong Qian, Weijia Cai, Xiaoxiao Lian. "Interactive, topicbased visual text summarization and analysis", Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09, 2009

[39] Jason Chuang, (2011)- Text VIE Lecture Notes, Stanford University Retrieved from http://hci.stanford.edu/courses/cs448b/f11/lectures/CS448B-20111117-Text.pdf

[40] [32] Farri, O., Rahman, A., Monsen, K., Zhang, R., Pakhomov, S.V., Pieczkiewicz, D.S., Speedie, S.M., & Melton, G. (2012). Impact of a prototype visualization tool for new information in EHR clinical documents. Applied clinical informatics, 3 4, 404-18.

[41] Radhakrishnan, K., Monsen, K. A., Bae, S.-H., & Zhang, W. (2016). Visual Analytics for Pattern Discovery in Home Care: Clinical Relevance for Quality Improvement. Applied Clinical Informatics, 7(3), 711–730. http://doi.org/10.4338/ACI-2016-03-RA-0049

[42] Bruce, D; Newman, D. (1978). Interacting Plans. In Cognitive Science, Volume 2, Issue 3, July-September: 195-233

[43] Lehrnert, W. (1981). Plot Units and Narrative Summarization. In Cognitive Science, Volume 5, Issue 4: 293–331

[44] Dyer, G. (1983). The Role of Affect in Narratives. In Cognitive Science 7: 211-242,

[45] Carley, K. (1993). Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis. In Social Methodology, Vol. 23: 75-126

[46] Elmqvist, N., Javed, W., & McDonnel, B. (2010). Graphical Perception of Multiple Time Series. IEEE Transactions on Visualization and Computer Graphics, 16, 927-934.

[47] Spence I. William Playfair and the psychology of graphs. American Statistical Association JSM Proceedings; 2006:2426–2436.

[48] Powsner S, Tufte E. Graphical summary of patient status. Lancet. 1994;344:386–389.

[49] West, V. L., Borland, D., & Hammond, W. E. (2015). Innovative information visualization of electronic health record data: a systematic review. Journal of the American Medical Informatics Association : JAMIA, 22(2), 330–339. http://doi.org/10.1136/amiajnl-2014-002955

[50] Tufte, Edward R., 1942-. (2001). The visual display of quantitative information. Cheshire, Conn. :Graphics Press,

[51] William S. Cleveland. 1993. Visualizing Data. Hobart Press

[52] Card, S.K., Mackinlay, J.D., & Shneiderman, B. (1999). Readings in information visualization - using vision to think.

[53] K. Kucher and A. Kerren, "Text visualization techniques: Taxonomy, visual survey, and community insights," 2015 IEEE Pacific Visualization Symposium (PacificVis), Hangzhou, 2015, pp. 117-121.doi: 10.1109/PACIFICVIS.2015.7156366

[54] Neto, Joel & Freitas, Alex & Kaestner, Celso. (2002). Automatic Text Summarization Using a Machine Learning Approach. 2507. 205-215. 10.1007/3-540-36127-8_20.

[55] Kucher, K., Kerren, A.: Text visualization browser: a visual survey of text visualization techniques. Poster Abstracts of IEEE VIS (2014)

[56] Cao, Nan & Cui, Weiwei. (2016). Visualizing Document Similarity. 49-56. 10.2991/978-94-6239-186-4_4.

[57] Neto, J.L., Santos, A.D., Kaestner, C.A., Freitas, A.A.: Document clustering and text summarization. In: Proceedings

of the International Conference Practical Applications of Knowledge Discovery and Data Mining, pp. 41–55. The Practical Application Company (2000)

[58] Koren, Y., Carmel, L.: Visualization of labeled data using linear transformations. In: Proceedings of IEEE Symposium on Information Visualization, pp. 121–128 (2003)

[59] Iwata, T., Yamada, T., Ueda, N.: Probabilistic latent semantic visualization: topic model for visualizing documents. In: Proceedings of SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 363–371. ACM (2008)

[60] Collins, C.,Carpendale, S., Penn, G.:Docuburst: visualizing document content using language structure. Comput. Graph. Forum **28**(3), 1039–1046 (2009)

[61] Stasko, J., Zhang, E.: Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: IEEE Symposium on Information Visualization, 2000. InfoVis 2000, pp. 57–65. IEEE (2000)

[62] Rusu, D., Fortuna, B., Mladeni´c, D., Grobelnik,M., Sipos, R.: Document visualization based on semantic graphs. In: 13th InternationalConference InformationVisualisation, pp. 292–297. IEEE (2009)

[63] Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D., Deussen, O., et al.: Document cards: a top trumps visualization for documents. IEEE Trans. Vis. Comput. Graph. 15(6), 1145–1152 (2009)

[64] Stoffel, A., Strobelt, H., Deussen, O., Keim, D.A.: Document thumbnails with variable text scaling. Comput. Graph. Forum **31**(3pt3), 1165–1173 (2012)

[65] Kaser, O., Lemire, D.: Tag-cloud drawing: algorithms for cloud visualization. arXiv preprint cs/0703109 (2007)

[66] Havre, S., Hetzler, B., Nowell, L.: Themeriver: visualizing theme changes over time. In: IEEE Symposium on Information Visualization, 2000. InfoVis 2000, pp. 115–123. IEEE (2000)

[67] B. Shneiderman, C. Plaisant and B. W. Hesse, "Improving Healthcare with Interactive Visualization," in Computer, vol. 46, no. 5, pp. 58-66, May 2013.doi: 10.1109/MC.2013.38

[68] L. T. De Paolis, M. Pulimeno and G. Aloisio, "Visualization and interaction systems for surgical planning," Proceedings of the ITI 2010, 32nd International Conference on Information Technology Interfaces, Cavtat/Dubrovnik, 2010, pp. 269-274.

[69] Shortliffe, E. H. and Cimino, J. J. (Editors), Biomedical Informatics: Computer Applications in Healthcare and Biomedicine: 4th Edition, Spr

[70] A. Younas, M. S. A. Malik, Khalil-ur-Rehman and R. Shahid, "A detailed study on temporal data visualization techniques in electronic health records," 2016 Sixth International Conference on Innovative Computing Technology (INTECH), Dublin, 2016, pp. 638-643. doi: 10.1109/INTECH.2016.7845074

[71] Khalil-ur-Rehman, Malik, M.S., Shahid, R., & Younas, A. (2016). A detailed study on temporal data visualization techniques in electronic health records. 2016 Sixth International Conference on Innovative Computing Technology (INTECH), 638-643.

[72] Aigner W., Miksch S., Schumann H., Tominski C. (2011) Visualization Aspects. In: Visualization of Time-Oriented Data. Human-Computer Interaction Series. Springer, London

[73] Mcpeek Hinz, Eugenia & Borland, David & Shah, Hina & West, Vivian & Hammond, Ed. (2014). Temporal Visualization of Diabetes Mellitus via Hemoglobin A1c Levels.

[74] Stab C., Nazemi K., Fellner D.W. (2010) SemaTime - Timeline Visualization of Time-Dependent Relations and Semantics. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2010. Lecture Notes in Computer Science, vol 6455. Springer, Berlin, Heidelberg

[75] A. A. T. Bui, D. R. Aberle and H. Kangarloo, "TimeLine: Visualizing Integrated Patient Records," in IEEE Transactions on Information Technology in Biomedicine, vol. 11, no. 4, pp. 462-473, July 2007.doi: 10.1109/TITB.2006.884365

[76] Powsner, S.M., Tufte, E.R., Graphical summary of patient status. The Lancet, 344 (Aug 6, 1994) 386-389

[77] Ozturk S, Kayaalp M, McDonald CJ. AMIA Annual Symposium Proceedings. Vol. 2014. American Medical Informatics Association; 2014. Visualization of Patient Prescription History Data in Emergency Care; p. 963.

[78] Lee, J., Kong, H.-K., Lin, S., & Karahalios, K. (2016). Plexlines: Tracking Socio-communicative Behaviors Using Timeline Visualizations. AMIA Annual Symposium Proceedings, 2016, 1890–1899.

[79] Dima AL, Dediu D (2017) Computation of adherence to medication and visualization of medication histories in R with AdhereR: Towards transparent and reproducible use of electronic healthcare data. PLoS ONE 12(4): e0174426. https://doi.org/10.1371/journal.pone.0174426

[80] Cousins, S., Kahn, M., The visual display of temporal information, Artificial Intelligence in Medicine 3 (1991) 341-357

[81] Hirsch, J. S., Tanenbaum, J. S., Lipsky Gorman, S., Liu, C., Schmitz, E., Hashorva, D., … Elhadad, N. (2015). HARVEST, a longitudinal patient record summarizer. Journal of the American Medical Informatics Association : JAMIA, 22(2), 263–274. http://doi.org/10.1136/amiajnl-2014-002945

[82] Cleveland, W. S. (1993). Visualizing Data. Hobart Press, Summit, New Jersey.

[83] Inselberg, A. (1998). A Survey of Parallel Coordinates. In Hege, [57] H.-C. & Polthier, K., editors, Mathematical Visualization. Springer-Verlag, Berlin.

[84] Tominski, C. (2008). Visual Methods for Analyzing Human Health Data.

[85] Borland, D., Hammond, E., & West, V.L. (2016). Multivariate Visualization of Longitudinal Clinical Data.

[86] Dogu, Eralp & Nembhard, Harriet. (2011). Multivariate Data Visualization Tools to Explore Hypertension.

[87] Karol, S., Lee, G., Noh, A.S., Rosenfeld, P., Sopan, A., & Shneiderman, B. (2012). Community Health Map: A geospatial and multivariate data visualization tool for public health datasets. Government Information Quarterly, 29, 223-234.

[88] Laramee, B., Tong, C., McNabb, L., Laramee, B., Lyons, J., Walters, A., . . . Berridge, D. (2017). Time-Oriented Cartographic Treemaps for the Visualization of Public Healthcare Data. Proceedings of the Computer Graphics and Visual Computing (CGVC) Conference 2017. doi:10.2312/cgvc.20171273

[89] Berridge, D., Tong, C., Roberts, R., Laramee, R., Berridge, D., & Thayer, D. (2017). Cartographic treemaps for the visualisation of public health care data.

[90] Plaisant Catherine, Mushlin Richard, Snyder Aaron, Li Jia, Heller Dan, Shneiderman Ben. Life-Lines: Using Visualization to Enhance Navigation and Analysis of Patient Records. In: Chute CG, editor. Proceedings of the 1998

American Medical Informatic Association Annual Fall Symposium; Bethesda, MD: AMIA; Nov 9–11, 1998. pp. 76–80.

[91] Lifelines2: Discovering Temporal Categorical Patterns across Multiple Records. http://www.cs.umd.edu/hcil /lifelines2/. Accessed September 5, 2013.

[92] AIGNER, W., KAISER, K., & MIKSCH, S. (2008). Visualization Methods to Support Guideline-Based Care Management. *Studies in Health Technology and Informatics*, *139*, 140–159.

[93] Event Flow www.cs.umd.edu/hcil/eventflow. Accessed March 11, 2018.

[94] Lifelines2: Discovering Temporal Categorical Patterns across Multiple Records. http://www.cs.umd.edu/hcil/lifelines2/. Accessed March 11, 2018

[95] Meyer, T., Monroe, M., Plaisant, C., Lan, R., Wongsuphasawat, K., Coster, T., Gold, S., Millstein, J., Shneiderman, B., Visualizing Patterns of Drug Prescriptions with EventFlow: A Pilot Study of Asthma Medications in the Military Health System, Proc. Of Workshop on Visual Analytics in HealthCare, VAHC2013 - Copyright retained by the authors (2013)

[96] Carter, E., Burd, R., Monroe, M., Plaisant, C., and Shneiderman, B. Using eventflow to analyze task performance during trauma resuscitation. Proceedings of the Workshop on Interactive Systems in Healthcare (WISH 2013) (2013).

[97] Monroe, M., Meyer, T. E., Plaisant, C., Lan, R., Wongsuphasawat, K., Coster, T. S., Gold, S., Millstein, J., and Shneiderman, B. Visualizing patterns of drug prescriptions with eventflow: A pilot study of asthma medications in the military health system. Proceedings of Workshop on Visual Analytics in Healthcare (VAHC 2013) (2013).

[98] Yiye Zhang, Rema Padman, Nirav Patel, Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data, In Journal of Biomedical Informatics, Volume 58, 2015, Pages 186-197, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2015.09.009

[99] Bettencourt-Silva, J.H., Clark, J., Cooper, C.S., Iglesia, B.D., Mills, R., & Rayward-Smith, V.J. (2015). Building Data-Driven Pathways From Routinely Collected Hospital Data: A Case Study on Prostate Cancer. JMIR medical informatics.

[100] Zhang Z, Ahmed F, Mittal A, Ramakrishnan I, Zhao R, Viccellio A, Mueller K (2011) AnamneVis: a framework for the visualization of patient history and medical diagnostics chains. In: Proceedings of the IEEE visual analytics in health care (VAHC) Workshop

[101] Bade, Ragnar, Stefan Schlechtweg, and Silvia Miksch.Connecting time-oriented data and information to a coherent interactive visualization. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 105-112. ACM, 2004.

[102] Ahmed, F., Mueller, K., Ramakrishnan, I.V., Viccellio, A., Wang, B., Zhang, Z., & Zhao, R. (2013). The Five Ws for Information Visualization with Application to Healthcare Informatics. IEEE Transactions on Visualization and Computer Graphics, 19, 1895-1910.

[103] Bui, A.A., & Hsu, W. (2010). Medical Data Visualization: Toward Integrated Clinical Workstations.

[104] Gómez, J.A., Plaisant, C., Shneiderman, B., Taieb-Maimon, M., Wongsuphasawat, K., & Wang, T.D. (2011). LifeFlow: visualizing an overview of event sequences. CHI.

[105] Borland, David & West, Vivian & Hammond, Ed. (2014). Multivariate Visualization of System-Wide National Health Service Data Using Radial Coordinates.

[106] Lee, Jaehoon & Hulse, Nathan & Oniki, Thomas & Huff, Stanley. (2015). Service Oriented Development of Information Visualization of the Electronic Health Records for Population Data Set.

[107] Kumar Kaushal, Kulendra & Kaushik, Shruti & Choudhury, Abhinav & Viswanathan, Krish & Chellappa, Balaji & Natarajan, Sayee & Pickett, Larry & Dutt, Varun. (2017). Patient Journey Visualizer: A Tool for Visualizing Patient Journeys.

[108] Perer, Adam, and David Gotz. "Visualizations to Support Patient-Clinician Communication of Care Plans." In Proceedings of ACM CHI Workshop on Patient-Clinician Communication (2013). 2013

[109] Krause, Josua, Narges Razavian, Enrico Bertini, and David Sontag. "Visual Exploration of Temporal Data in Electronic Medical Records." In AMIA. 2015

[110] Perer, A., & Sun, J. (2012). MatrixFlow: Temporal Network Visual Analytics to Track Symptom Evolution during Disease Progression. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2012, 716-25.

[111] Goovaerts, P. (2010). Three-dimensional Visualization, Interactive Analysis and Contextual Mapping of Space-time Cancer Data.

[112] Rind A, Aigner W, Miksch S, Wiltner S, Pohl M, Turic T, Drexler F (2011) Visual exploration of time-oriented patient data for chronic diseases: design study and evaluation. In: Holzinger A, Simonic K-M (eds) USAB 2011: information quality in e-health. LNCS, vol 7058. Springer, Heidelberg, pp 301–320

[113] Bui, A., Aberle, D.R., Kangarloo, H., 2007. TimeLine: Visualizing Integrated Patient Records. IEEE Transactions on Information Technology in Biomedicine 11(4):462-473

[114] Catherine Plaisant , Brett Milash , Anne Rose , Seth Widoff , Ben Shneiderman, LifeLines: visualizing personal histories, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, p.221-227, April 13-18, 1996, Vancouver, British Columbia, Canada [doi>10.1145/238386.238493

[115] Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S., Shneiderman, B., 2008. Aligning temporal data by sentinel events: discovering patterns in electronic health records, CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp. ACM—466

[116] Hallett, C., 2008. Multi-modal presentation of medical histories, IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces, pp. ACM--89.

[117] Roberts, A., Gaizauskas, R., Hepple, M., Davis, N., Demetriou, G., Guo, Y., … Wheeldin, B. (2007). The CLEF Corpus: Semantic Annotation of Clinical Text. *AMIA Annual Symposium Proceedings*, *2007*, 625–629.

[118] Cheng, C., & Shahar, Y. (1999). Intelligent Visualization and Exploration of Time-Oriented Clinical Data. Topics in health information management, 20 2, 15-31.

[119] Shahar Y, Goren-Bar D, Galperin M, Boaz D, and Tahan G. KNAVE-II: A distributed architecture for interactive

visualization and intelligent exploration of time-oriented clinical data The 7th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2003), pp. 103-110, Protaras, Cyprus, Oct. 2003.

[120] Goren-Bar, D., Shahar, Y., Galperin-Aizenberg, M., Boaz, D., Tahan, G., 2004. KNAVE II: the definition and implementation of an intelligent tool for visualization and exploration of time-oriented clinical data, AVI '04: Proceedings of the working conference on Advanced visual interfaces, pp. ACM--174.

[121] Kosara, R., Miksch, S., 2001. Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. Artif Intell Med 22, 111-131.

[122] Joshi, R., & Szolovits, P. (2012). Prognostic Physiology: Modeling Patient Severity in Intensive Care Units Using Radial Domain Folding. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2012, 1276-83. Coster, T.S., Gold, S., Lan,

[123] Klimov, D., & Shahar, Y. (2005). A Framework for Intelligent Visualization of Multiple Time-Oriented Medical Records. AMIA Annual Symposium Proceedings, 2005, 405–409.

[124] Cao, N., Gotz, D., Qu, H., & Sun, J. (2011). DICON: Interactive Visual Analysis of Multidimensional Clusters. IEEE Transactions on Visualization and Computer Graphics, 17, 2581-2590.

[125] Gotz, D., & Wongsuphasawat, K. (2011). Outflow: Visualizing Patient Flow by Symptoms and Outcome.

[126] R., Monroe, M., Meyer, T., Millstein, J., Plaisant, C., Shneiderman, B., & Wongsuphasawat, K. (2013). Visualizing Patterns of Drug Prescriptions with EventFlow: A Pilot Study of Asthma Medications in the Military Health System.

[127] ȚĂRANU, I. (2016). Data mining in healthcare: decision making and precision.

[128] Huang, Z., Juarez, J.M., & Li, X. (2017). Data Mining for Biomedicine and Healthcare. Journal of healthcare engineering.

[129] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems.

[130] Reddy, C.K., & Sun, J. (2013). Big data analytics for healthcare. KDD.

[131] Chute, C.G., Masanz, J.J., Ogren, P.V., Savova, G.K., Sohn, S., Schuler, K.K., & Zheng, J. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association : JAMIA, 17 5, 507-13.

[132] Sumithra Velupillai, D. Mowery, Brett R. South, Maria Kvist, and Hercules Dalianis. 2015. Recent advances in clinical natural language processing in support of semantic analysis. Yearbook of Medical Informatics, 10:183--193.

[133] Chapman, W.W. (2010). Closing the gap between NLP research and clinical practice. Methods of information in medicine, 49 4, 317-9.

[134] Horn, W., Popow, C., & Unterasinger, L. (2001). Support for fast comprehension of ICU data: visualization using metaphor graphics. Methods of information in medicine, 40 5, 421-4

[135] Harris, D. R., & Henderson, D. W. (2016). i2b2t2: Unlocking visualization for clinical research. AMIA Summits on Translational Science Proceedings, 2016, 98-104

[136] Chang, H., & Ko, I. (2017). Interactive Visualization of Healthcare Data Using Tableau. Healthcare informatics research.

[137] Burns, L., Brunker, C., Cannon, W., Christensen, K., Clayton, P.D., Dorr, D.A., Jones, S.S., Larsen, A., Narus, S.P., Radican, K., Thornton, S.N., & Wilcox, A.B. (2005). Use and Impact of a Computer-Generated Patient Summary Worksheet for Primary Care. AMIA ... Annual Symposium proceedings. AMIA Symposium, 824-8.

[138] Rashid, A.H., & Yasin, N.B. (2015). Privacy preserving data publishing: Review.

[139] Dhawale, S., Ingale, P., Kedar, S., Kadam, P., & Wani, S. (2013). Privacy Preserving Data Mining.

[140] Dasgupta, A., Maguire, E., Abdul-Rahman, A., and Chen, M. 2014. Opportunities and challenges for privacy-preserving visualization of electronic health record data. In IEEE VIS 2014 Workshop on Visualization of Electronic Health Records.

[141] Dasgupta, A., Kosara, R.: Adaptive privacy-preserving visualization using parallel coordinates. IEEE Trans. Vis. Comput. Graph. 17(12), 2241–2248 (2011)

[142] Dasgupta, A., Chen, M. and Kosara, R. (2013), Measuring Privacy and Utility in Privacy-Preserving Visualization. Computer Graphics Forum, 32: 35–47. doi:10.1111/cgf.12142

[143] Dasgupta, A., & Kosara, R. (2011). Privacy-preserving data visualization using parallel coordinates. Visualization and Data Analysis.

[144] Halpern Y, Horng S, Nathanson L, Shapiro N, Sontag D (2012) A comparison of dimensionality reduction techniques for unstructured clinical text. In: Proc. ICML Workshop on Clinical Data Analysis.

[145] Tomoharu Iwata , Takeshi Yamada , Naonori Ueda, Probabilistic latent semantic visualization: topic model for visualizing documents, Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, August 24-27, 2008, Las Vegas, Nevada, USA [doi>10.1145/1401890.1401937]

[146] Zilner, S., Hauer, T., Rogulin, D., Tsymbal, A., Huber, M. and Solomonides, T. (2008) Semantic visualization of patient information. The 21st IEEE International Symposium on Computer-Based Medical Systems. pp. 296-301. ISSN 1063-7125 Available from: http://eprints.uwe.ac.uk/13413

[147] Lauesen, S., & Pantazos, K. (2012). Constructing Visualizations with InfoVis Tools - An Evaluation from a user Perspective. GRAPP/IVAPP.

[148] Latha, N., Murthy, B. & Sunitha, U. (2012). Electronic Health Record. International Journal of Engineering Research & Technology. Volume 1. International Journal

[149] Burton, L. C., Anderson, G. F., & Kues, I. W. (2004). Using Electronic Health Records to Help Coordinate Care. The Milbank Quarterly, 82(3), 457–481. http://doi.org/10.1111/j.0887-378X.2004.00318.x

[150] Ross, M. K., Wei, W., & Ohno-Machado, L. (2014). "Big Data" and the Electronic Health Record. Yearbook of Medical Informatics, 9(1), 97–104. http://doi.org/10.15265/IY-2014-0003

[151] Borland, D., & Gotz, D. (2016). Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization. IEEE Computer Graphics and Applications, 36, 90-96.

[152] Vivian L West, David Borland, W Ed Hammond; Innovative information visualization of electronic health record data: a systematic review, Journal of the American Medical

Informatics Association, Volume 22, Issue 2, 1 March 2015, Pages 330–339, https://doi.org/10.1136/amiajnl-2014-002955

[153] Lee J, Holmen JR, Catmull SL, Pollock SE, Haug PJ, Huff SM:Research oriented clinical data visualization based on the analytic health repository: A case study at Intermountain Healthcare, VIS 2013, Atlanta, GA

[154] ISO/DTR 20514. Health informatics - electronic health record: definition, scope and context. Schloeffel P, ed. Fourth Draft, March 2004.

[155] Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inf 2008;77(5):291-304.

[156] Farri, Oladimeji Feyisetan. (2012). Understanding clinician information demands and synthesis of clinical documents in electronic health record systems. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/131431.

[157] W Pickering, Brian & Gajic, Ognjen & Ahmed, Adil & Herasevich, Vitaly & T Keegan, Mark. (2013). Data Utilization for Medical Decision Making at the Time of Patient Admission to ICU. Critical care medicine. 41. 10.1097/CCM.0b013e318287f0c0.

[158] Murphy, D.R., Thomas, E.J., Meyer, A.N.D., Singh, H. Development and validation of electronic health record-based triggers to detect delays in follow-up of abnormal lung imaging findings. Radiology. 2015;277:81–87.

[159] McDonald CJ, Callaghan FM, Weissman A, Goodwin RM, Mundkur M, and Kuhn T. 2014. “Use of Internist’s Free Time by Ambulatory Care Electronic Medical Record Systems.” JAMA Internal Medicine, September. doi:10.1001/jamainternmed. 2014.

[160] Belden, J.L., Clarke, M.A., Canfield, S., Koopman, R.J., Kim, M.S., Moore, J.L., & Steege, L.M. (2015). Physician Information Needs and Electronic Health Records (EHRs): Time to Reengineer the Clinic Note. Journal of the American Board of Family Medicine : JABFM, 28 3, 316-23.

[161] Wrenn JO, Stein DM, Bakken S, Stetson PD. Quantifying clinical narrative redundancy in an electronic health record. J. Am. Med. Inform. Assoc. 2010;17(1):49.

[162] Beasley JW, Wetterneck TB, Temte J, Lapin JA, Smith P, Rivera-Rodriguez AJ, Karsh BT. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med 2011;24(6):745-751.

[163] Zhang R, Pakhomov S, McInnes BT, Melton GB. Evaluating Measures of Redundancy in Clinical Texts. AMIA Annual Symposium Proceedings: American Medical Informatics Association; 2011.

[164] Chittaro, L. (2001). Visualization and its Application to Medicine, Artificial intelligence in medicine, Vol 22(2), pp. 81-88

[165] [164] Eppler, M.J., and Mengis, J., "The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, Mis, and Related Disciplines", The Information Society, 50(5), 2010, pp. 325-344.

[166] [165] Glenberg, A.M., and Langston, W.E., "Comprehension of Illustrated Text: Pictures Help to Build Mental Models", Journal of Memory and Language, 31(1), 1992, pp. 129-151.