Semantic-Level New Information Identification in Electronic Health Records Using Text-Mining Techniques

Ya-Han Hu
Department of Information Management, National Central University, Taiwan
yhhu@mgt.ncu.edu.tw

Hsiao-Ting Tseng
Department of Information Management, National Central University, Taiwan
httseng@mgt.ncu.edu.tw

Chun-Feng Huang
Department of Information Management, National Chung Cheng University, Taiwan
weilin0933602899@gmail.com

Abstract

Electronic health records (EHRs) are widely used in healthcare systems to store and transmit patients' health records. They have many advantages, such as saving space, increasing efficiency, and facilitating communication. However, they also have a major drawback: information redundancy. Healthcare professionals often use copy and paste to write clinical notes, which leads to excessive similarity and low diversity in EHRs. This impairs the readability and quality of EHRs and hinders decision making. To address this problem, this study proposes a text-mining approach to identify new information at semantic-level in EHRs. Unlike previous studies that focused on word-level identification, we use concept occurrence and concept similarity score methods to annotate new information at semantic-level and evaluate them with gold standards. The experimental evaluation demonstrates that the method proposed in this study achieves an F1-score ranging from 78.57 to 80.31 under various parameter combinations. The proposed method enables healthcare professionals to read EHRs more efficiently and make more informed decisions.

Keywords: Data mining, New information, Semantic similarity, UMLS, Electronic health records

1. Introduction

The emergence of smart healthcare and the COVID-19 pandemic has underscored the significance of electronic health records (EHRs). EHRs play a crucial role in medical informatics as they chronologically document a diverse range of clinical information about patients. This includes structured data like vital signs, medication details, and physical examination results, as well as unstructured data such as progress notes, discharge summaries, and diagnostic test reports (Menachemi & Collum, 2011; Shoolin et al., 2013). By providing comprehensive medical histories, EHRs offer numerous benefits to clinical practice. They assist physicians in decision-making, particularly in emergency situations where patient overcrowding can compromise healthcare quality (Batley et al., 2011). Moreover, EHR systems enhance emergency efficiency and healthcare quality by enabling rapid communication and real-time responses to infectious diseases (Chaudhry et al., 2006).

Previous study has highlighted that clinical practitioners spend over 37% of their time reviewing EHRs, detracting from direct patient interaction and examinations (Hingle, 2016). Prolonged viewing of EHRs indirectly diminishes patient satisfaction with overall medical care. Overhage & McCallie (2020) also noted that internists using EHRs spend an additional 48 minutes per day on document indexing compared to traditional paper records. Therefore, marking out new information in medical records to facilitate quick reading for doctors can not only enhance patient satisfaction with medical care but also maximize the value of clinical decision support systems.

Unstructured data in EHRs primarily consist of textual information written by medical staff. These texts may exhibit varying structures depending on the writing style and habits of individuals and may also contain grammatical or spelling errors, short sentences, informal abbreviations, or dialects. These factors pose challenges when applying natural language processing to EHRs (Yadav et al., 2017). Additionally, redundancy is an increasingly significant issue. Most modern EHR systems incorporate copy-and-paste functionality to reduce the time doctors spend entering information for each patient visit. However, this practice also contributes to increased redundancy in EHR content, particularly for patients undergoing complex treatments or long-term hospitalization. Consequently, their EHR texts become difficult to read and excessively verbose (Hirschtick, 2006; Markel, 2010). Such challenges

URI: https://hdl.handle.net/10125/106786
978-0-9981331-7-1

further exacerbate the cognitive burden faced by clinical physicians.

In the fast-paced clinical environment, physicians face numerous challenges when reviewing comprehensive patient data. Redundant information within the clinical text acts as noise, obscuring clinically relevant and new information in the EHRs. Furthermore, these redundancies often consist of outdated or erroneous information, with the issue becoming more severe due to the use of copy and paste functions, making it difficult for physicians to effectively understand and utilize the data for clinical decision support (Zhang et al., 2014). Wrenn et al. (2010b) and Zhang et al. (2011) employed automated methods to quantify data redundancy in medical records. They found that nearly 76% of the content in EHRs for inpatients and outpatients consisted of redundant and repetitive information, emphasizing the significance and urgency of this problem in clinical texts.

Text summarization methods are typically employed to address this problem. While these methods can simplify the length of medical records and improve the efficiency of reading for medical staff, redundant information still persists, leading to an ongoing cycle of reading difficulties. If summaries are directly generated from this medical record content, the redundancy remains. Therefore, a more effective method would be to initially identify new information, followed by the summarization process (Pivovarov & Elhadad, 2015; Wrenn et al., 2010a).

Several researches have been conducted on identifying new information at the semantic level (Chen et al., 2019; Hu et al., 2015; Liang et al., 2019; Moradi & Ghadiri, 2017; Zhang et al., 2012; Zhang et al., 2017; Wang & Fang, 2016). However, they often utilized default parameters when using the medical terminology tool such as MetaMap to map medical concepts from the records. This approach does not guarantee the best results, particularly for patients with comorbidities.

This study utilizes hospital records of stroke patients from the case hospital as the primary data source. In the mapping stage of medical concepts, we employ diverse parameter combinations within the MetaMap, a well-known tool designed for mapping biomedical text to concepts within the Unified Medical Language System (UMLS), aiding in the extraction and organization of complex medical information for analysis. Furthermore, we explore various medical record retrospective days and similarity algorithms to effectively identify new information within the cases. Through a series of comprehensive experiments, we aim to identify the optimal parameter combinations that can be applied to enhance clinical decision support systems.

The organization of this paper is as follows: Section 2 reviews past research on text summarization and the retrieval of new information from clinical texts. Section 3 provides a detailed description of the system architecture, the text preprocessing techniques employed, and the experimental design. Section 4 presents the complete experimental results, while Section 5 concludes our study.

2. Literature review

2.1. Text summarization

Text summarization methods can be broadly categorized into two main types: extractive and abstractive. In the context of medical texts, most of the existing literature focuses on extractive methods for summarization. Extractive summarization involves generating a summary by selecting phrases or sentences directly from the original document. One advantage of using extractive methods in the clinical medical field is the ability to preserve the original content written by physicians (Liang et al., 2019).

In previous studies, researchers have employed various approaches for extractive summarization. Some have utilized regular expressions to identify different types of information such as test results, adjectival parameters (e.g., tumor texture, location, quantity), therapy or negation words. They then applied self-defined algorithms for natural language processing (Chen et al., 2019). Others have constructed sentence sequence models using linear-chain conditional random fields (Linear-chain CRF) and subsequently used a simple CNN-rand technique for summarization (Liang et al., 2019). Another approach involved employing MetaMap to map out Semantic Type information in the text and then grouping the results into itemsets. Through scoring, important and relevant information was selected for summarization (Moradi & Ghadiri, 2017).

2.2. Identification of new information in medical texts

EHRs consist of continuous text content, and although previous studies have addressed information redundancy through summarization, relying solely on summarization can still lead to information overload. Furthermore, some methods separate the summary from the original medical record, which may hinder physicians' ability to focus on potentially important key information within the original record.

Researchers have explored alternative approaches, such as improving the user interface of medical decision support systems or searching for new information within the continuous medical record data.

The identification and aggregation of similar information within EHRs can be categorized into three different levels of language processing: word-level, semantic-level, and sentence-level methods. Zhang et al. (2012) employed a Bi-gram model to analyze textual data and identify differences for identifying new information. Other approaches involve recognizing noun entities in medical texts to extract useful information (Wang & Fang, 2016). Some researchers have used semantic similarity measures to determine whether concepts with similar meanings should be considered as new information (Hu et al., 2015; Zhang et al., 2017). These methods have shown promising results particularly at the semantic level.

However, past research has focused on understanding the content of medical texts but has overlooked the issue of significant repetitive information present in actual clinical texts. For clinicians, filtering redundant information from clinical texts is more clinically meaningful than developing automated tools for medical text comprehension. This study aims to address this research gap, exploring how to filter out redundant information from clinical texts and assist in identifying valuable new information. The goal is to enhance the efficiency with which clinicians review electronic medical records.

3. Research method

The research process comprises the following steps: (1) Data collection and preprocessing; (2) Utilizing the medical terminology tool, MetaMap, to map medical terms; (3) Computing similarity scores among different concepts; (4) Annotating new information by the proposed methods; and (5) Evaluating the annotation performance.

This study delves into the chronological notes of stroke patients from a case hospital in Taiwan. These medical records are preprocessed and subsequently fed into MetaMap, which extracts relevant medical concepts from the text, encompassing Concept Unique Identifiers (CUI), Semantic Type, Negation, and other attributes. Additionally, similarity scores discern whether these concepts convey new or pre-existing information.

During the annotation phase, both concept dichotomy and similarity scores assist in pinpointing new information in the medical records. Ultimately, the system's output is compared to the gold standard (new information co-confirmed by the two physicians at the case hospital) to gauge the performance of the system.

3.1. Data collection and preprocessing

The dataset consisted of records from 10 inpatients, each with a duration of 9 to 18 days. Clinical physicians recommended focusing solely on two types of data: admission note and progress note (as indicated in Table 1). For each inpatient, there is one admission note upon admission and daily progress notes recorded throughout their hospital stay. Other types of data, such as emergency medical records or discharge summaries, were excluded from the analysis due to their limited content and the challenges associated with comparing new information within them.

Table 1. Type and data fields in EHRs.

Type	Data fields
Admission Note	Chief complaint, Present illness, Past illness, Family history, Vital signs
Progress Note	Subject, Object, Assessment, Plan, Doctor sign

Hospital admission records primarily document a patient's status at the time of admission, including their medical history, physical examination results, reason for hospitalization, and initial care guidelines. In contrast, progress notes follow the Subjective, Objective, Assessment, and Plan (SOAP) format and adhere to standardized practices for medical record-keeping. However, EHRs, which are semi-structured data pulled from the hospital's EHR database, can vary in content across different sections. These records may also contain extraneous elements, such as symbols from human body examinations and Chinese punctuation marks, particularly within the "Objective Description" section. To mitigate these issues, several data preprocessing steps are carried out. These steps encompass data extraction from the database, removal of irrelevant symbols, restoration of line breaks, spell-checking, and abbreviation expansion. Only after these steps are completed is the text forwarded to MetaMap for further analysis.

To assess the system's performance, the Gold Standard is established based on the new information in all medical records, which has been collaboratively verified by two physicians from the case hospital. They were each instructed to independently highlight words they recognized as new information. When discrepancies appeared between their annotations, the final Gold Standard was determined by overlapping the ranges provided by both physicians.

3.2. Using the medical terminology tools

All preprocessed texts from EHRs are processed using MetaMap. Following the methodology of Aronson and Lang (2010), this study employs MetaMap for simple text preprocessing steps. Upon inputting the text, a sequence of syntactic analysis is performed, as described below:

1. Sentence boundary detection: Sentences within the text are delineated using line breaks or periods, and tokenization is employed to facilitate the identification and expansion of acronyms and abbreviations.
2. Part-of-speech tagging: Every word segmented in the previous step is annotated with its corresponding part of speech.
3. Expert dictionary lookup: Words listed in the expert dictionary are matched against the segmented words.
4. Phrase generation: Words located in the expert dictionary are grouped into phrases using a minimal commitment parser.

Upon completion of the syntactic analysis stage, variant generation is carried out for the generated phrases. These variants are identified and ranked based on their similarity to the original phrases, producing candidate results. Subsequently, these candidate results are mapped with the medical terms or medical concepts recorded in the UMLS comprehensive index. The mapped results are then outputted as fielded indexing, as depicted in Figure 1. The outputted results are formatted with separators to facilitate the extraction of medical information embedded within the text in subsequent stages.

Figure 1. Results after text mapping with MetaMap.

Table 2 illustrates the various types of medical information and their respective uses. These medical information types are categorized into five categories. Furthermore, each output result is formatted and stored as a .csv file, ensuring convenient accessibility and readability.

Table 2. Extracted medical information content.

Medical information	Description
CUI	Each medical term that is mapped to has a unique concept identifier, which corresponds to a concept. Through various similarity calculation methods, the similarity between a specific pair of CUIs can be calculated.
Semantic type	Each concept has its own semantic type.
Position	The position of the word mapped to this concept in the text.
Negation	Whether this concept has a negative meaning, such as: deny, without, non...
Trigger	Show the words or phrases in the text that trigger MetaMap to map this concept

3.3. Calculating the similarity scores between different concepts

By leveraging similarity scores, it becomes possible to assess the conceptual resemblance between two medical terms at the semantic level. The UMLS metathesaurus serves as a valuable resource, encompassing a vast range of medical terms and their associated concepts and semantic types. These terms and concepts are organized within a knowledge ontology system, structured across four levels. To facilitate the calculation of similarity scores, this study employs various methods, namely path-based, ontology-based, and note-based. These methods are integrated into the UMLS-Similarity module, implemented using the Perl programming language. This module enables the computation of similarity scores for any pair of CUIs. The subsequent sections will elaborate on each method individually.

3.3.1. Path-based. This fundamental method computes similarity based on the position of concepts within the ontology framework and all conceivable paths between two concepts. By traversing the hierarchical structure, it measures the similarity between concepts. While its strength lies in its straightforward calculation, a drawback is that to determine the shortest path, one must first traverse all possible routes between the two concepts, leading to a high computational cost. The two methods employed in this study are:

- Path: Calculate the shortest path between two CUIs, C_1 and C_2.

$$score_{path} = ShortestPath(C_1, C_2) \quad (1)$$

- Lch: In addition to the shortest path, the depth of the ontology system is also taken into account. Let D represents the total depth of the taxonomy, the $score_{lch}$ is caculated as follows.

$$score_{lch} = -log \frac{ShortestPath(C_1,C_2)}{2 * D} \quad (2)$$

3.3.2. Ontology-based. By viewing the entire UMLS semantic network as a knowledge ontology, we determine the similarity ratio between two CUIs through possible super-concepts. This approach is more efficient than calculating all potential paths. Let *P* denotes the union of ancestor concepts for any two CUIs and *Q* denotes the intersection of these ancestor concepts. The two ontology-based methods utilized in this study are:

- Sanchez (Sanchez et al., 2012):

$$score_{sanchez} = \frac{-log(\frac{P-Q}{P})}{log(2)} \quad (3)$$

- Batet (Batet et al., 2011):

$$score_{batet} = \frac{P-Q}{P} \quad (4)$$

3.3.3. Note-based. The computational method employed in this study is named "Vector," derived from the approach proposed by Patwardhan & Pedersen (2006). To calculate the semantic similarity between two target concepts, the following steps are taken: 1) Initially, identify words related to the target concepts in WordNet. Within WordNet, each word may have one or more definitions, termed as "senses."; 2) Construct context vectors. These senses, or explanations, serve as the context for the respective words. The context can be determined by referencing the hierarchical relationships found in WordNet. The vector representing the relationship between the two target concepts is then computed; 3) Finally, calculate cosine similarity between the two target concepts. The $\rightarrow_{V_1}$ and $\rightarrow_{V_2}$ in formula (5) are respectively the target concept C_1 and the target concept C_2 mentioned in formula (1), each with their explanations in WordNet. The calculated vector value, the formula is:

- Vector:

$$score_{vector} = \frac{\overrightarrow{V_1} \cdot \overrightarrow{V_2}}{|V_1||V_2|} \quad (5)$$

3.4. Annotating new information

This study proposes two new information identification methods, Method 1 and Method 2, based on semantic levels, with the aim of identifying new information and annotating it in the text, producing the final annotation results.

Method 1 uses concept dichotomy to identify new information. The goal of Method 1 is to find out which combination of MetaMap mapping parameters can achieve the highest F1-score. Based on different parameter settings listed in Table 3, concept strings are generated after mapping the text. These concept strings differ depending on the patient and their medical record days. Method 1 compares the difference between the concept string of day *N* and the previous *M* days, where *N* is the medical record to be checked for new information, and *M* is the number of days to be traced back for comparison. If a concept appears in any concept string within the trace-back days, it is regarded as old information and removed from the concept string. After the comparison, the remaining concepts are labeled as new information according to their original position in the text. The labeling results are assessed for annotation performance with Gold Standard.

Table 3. MetaMap parameter setting in Method 1.

Parameter	**#1**	**#2**	**#3**
Composite Phrases	4	4	2
Ignore Word Order	Off	On	On
Prefer Multiple Concept	Off	On	On
Allow Concept Gap	Off	Off	On
Word Sense Disambiguation	On	On	On
User Define Acronyms	On	On	On
Negated Setting	On	On	On

Method 2 identifies new information based on conceptual similarity scores. Method 2 aims to explore which similarity calculation method can achieve the best F1-scorem. In the annotation stage, different thresholds will be selected as the criteria for judging new and old information, corresponding to the scores in the score matrix, judging the similarity between different days and different CUIs. If the similarity score is greater than the set threshold, it is regarded as having high similarity, equivalent to old information that has appeared before, and no annotation is performed; otherwise, if it is lower than the set threshold, it is regarded as new information that has not appeared before, and annotation is performed. The results of the annotation will be evaluated with the Gold Standard for annotation effectiveness.

The similarity calculation methods used in this paper are shown in Table 4. Using the similarity values obtained from these computational methods, we can determine the degree of similarity between any two concepts, and select an optimal similarity threshold to decide whether to treat them as new or old information.

Table 4. Concept similarity measurement methods.

Type	Measurement	Description
Path-based	Path, Lch	Using the SNOMED-CT dictionary, only calculate the PARENT, CHILD relationships.

Ontology-based	Sanchez, Batet	Using the SNOMED-CT dictionary, the data is calculated separately for different record.
Note-based	Vector	Combining the structure, content and target concept of WordNet with the medical record text, calculate the similarity using the cosine theorem.

3.5. Performance evaluation

When performing the assessment, this study will only focus on the medical information in the text, so the non-medical terminology in the Gold Standard will be excluded. Let True Positive (TP) refers to the count of phrases/CUIs in the medical records that are concurrently identified as new information by both the Gold Standard and the system, False Positive (FP) refers to the count of phrases/CUIs in the medical records that the system identifies as new information but are not considered as such by the Gold Standard, False Negative (FN) refers to the count of phrases (or CUI) in the medical records that the Gold Standard considers as new information, but the system does not identify them as such. The following three metrics were used to evaluate system's performance.

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1-score = \frac{2\times precision\times recall}{precision+recall} \quad (8)$$

4. Research Results and Analysis

4.1. Result of Method 1

The detailed results for the three parameter settings of Method 1 are presented in Table 5. The table reveals that these settings yield better Recall compared to Precision in labeling performance. This can be attributed to the fact that most of the new information deemed important by doctors is effectively captured by the concept dichotomy method, although this method also tags some less relevant information. Among the evaluated cases, Patients 5 and 8 demonstrate the highest and lowest F1-scores, respectively.

Table 5. All patients in the three parameter settings annotations of Method 1.

Patient ID	#1			#2			#3		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	80.26	92.85	86.10	80.53	88.27	84.23	80.61	88.27	84.27
2	85.58	79.79	82.58	84.76	72.87	78.37	84.86	72.69	78.31
3	90.03	92.92	91.45	89.61	89.56	89.58	89.21	89.68	89.44
4	76.50	91.32	83.25	75.23	86.60	80.51	76.14	86.31	80.91
5	93.49	92.08	92.78	93.69	90.25	91.93	93.86	90.25	92.02
6	69.97	88.97	78.33	70.27	86.09	77.38	70.72	86.59	77.86
7	61.40	89.80	72.93	61.65	87.29	72.26	60.87	86.82	71.57
8	44.34	95.42	60.54	43.17	90.14	58.38	43.07	89.22	58.10
9	72.36	78.33	75.23	70.65	73.46	72.03	70.90	73.74	72.29
10	55.60	91.76	69.24	57.89	89.59	70.34	58.12	89.59	70.50
Average	72.95	89.32	**80.31**	72.74	85.41	78.57	72.84	85.32	78.58

Through the results of Method 1, it can be found that the patients with the best and worst F1-Score are 5 and 8 respectively. After observation, it can be found that there are significant differences in the text content of these two patients:

- The medical records of patient 5 contain more descriptive short sentences about the condition. Studies have shown that MetaMap performs better on short sentences than on long sentences in general.
- Patient 8's records are replete with test data, which creates noise during MetaMap's mapping stage, even after preprocessing. This disrupts the quality of the mapping results. Such test information typically consists of the patient's test results from that day or data copied from past medical records, and it is interspersed throughout various sections of the text.

4.2. Result of Method 2

Method 2 used six similarity calculation methods to annotate the effects and integrated them to achieve the best F1-score, as shown in Table 6. The Batet approach achieved the highest F1-score of 63.08% and provided similarity scores without incurring significant computational costs. In contrast, the Vector approach was the least effective. One possible explanation is that the Vector approach relies on the WordNet semantic network structure, which doesn't encompass all the medical terms found in the medical

records used in this study, especially when compared to UMLS. As a result, it fails to accurately represent the relationships between terms and their hierarchical structures.

Table 6. Optimal thresholds and F1-scores for each similarity calculation method.

Method	Optimal threshold	Best F1-score
Path	0.5	60.91%
Lch	3.5	60.91%
Sanchez	0.9	62.31%
Batet	0.975	63.08%
Vector	0.9	57.02%

4.3. Summary

From the above experimental results, it can be found that simply labeling with concept dichotomy has a better effect and the steps are relatively simple, and different patients also have different labeling effects. We further summarize the experimental results as following:

- Records that primarily consist of narrative information written by medical staff tend to perform better than those filled with laboratory data. Test data, scattered across different sections of each paragraph, can't be entirely eliminated through preprocessing, leading to noise. If medical information systems could segregate this test data into a separate paragraph and include time-series features, this would likely improve the accuracy of labeling.
- The calculation of concept similarity allows for a semantic-level comparison between two Concept Unique Identifiers (CUIs), categorizing them as either new or old information based on their similarity scores. However, experiments reveal that the optimal thresholds for all similarity calculation methods tend to fall at the upper limits of their score ranges. Zhang et al. (2017) employed a Path-based similarity calculation method combined with Method 2, yielding results similar to those of this study. This suggests that the concept dichotomy method is effective in marking most of the new information deemed important by physicians, although it also tags some information considered less relevant.
- Even when CUIs belong to different semantic types, similarity scores still exist between them. Once filtered by the threshold, very few concepts are designated as new information, resulting in a low annotation rate and diminished effectiveness.

5. Conclusion

EMRs are increasingly becoming popular due to their digital advantages, leading to more extensive sharing of patient information across clinical organizations. However, the accumulation of repetitive test data in EHRs causes information overload, demanding significant time from medical staff to comprehend. While researchers have used statistical language models and tools like PubMed or UMLS to identify new EHR information, few have compared the effectiveness of various similarity calculation methods or adjusted text mapping parameters in MetaMap, using CUI information for diverse day medical record retrieval.

In this study, EHRs of stroke patients from a case hospital were processed using MetaMap to map medical term concepts. New information was identified based on these concepts' historical appearances and their similarity scores. This system's findings were cross-referenced with a Gold Standard, co-confirmed by two physicians. Experimental evaluation showed our method attained an F1-score of 78.57-80.31% under different parameter sets. The Batet approach, in another test, was more efficient, achieving an F1-score of 63.08%. Clinically, this approach helps pinpoint new information in EHRs, easing the reading process for medical staff, aiding swift patient health comprehension, and fostering timely, informed decisions.

This study is not without limitations. First, the calculation and analysis work before annotation still time consuming. It is still a challenge to deploy it in the system of medical institutions. In the future, it may be possible to deploy the UMLS corpus or the calculation similarity method on the cloud, hoping to reduce the calculation cost. Second, The habits of physicians in writing medical records may affect the system's judgment performance. In the case hospital, a small number of physicians do not use punctuation but use blank or newline keys, which may cause some errors in sentence or word segmentation. Different ways of segmenting sentences and words and throwing them into MetaMap will also cause different results when mapping. Third, some medical records appear in two languages (Chinese and English). Therefore, this study first translated the Chinese into English before preprocessing the medical text. In future applications in case hospitals, an additional preprocessing step will be needed for clinical use.

References

Aronson, A. R., & Lang, F. M. (2010). An overview of MetaMap: Historical perspective and recent advances. *Journal of the American Medical Informatics Association*, *17*(3), 229–236.

Batet, M., Sánchez, D., & Valls, A. (2011). An ontology-based measure to compute semantic similarity in biomedicine. *Journal of Biomedical Informatics*, 44(1), 118–125.

Batley, N. J., Osman, H. O., Kazzi, A. A., & Musallam, K. M. (2011). Implementation of an emergency department computer system: Design features that users value. *Journal of Emergency Medicine*, *41*(6), 693–700.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., … Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine*, *144*(10), 742–752.

Chen, L., Song, L., Shao, Y., Li, D., & Ding, K. (2019). Using natural language processing to extract clinically useful information from Chinese electronic medical records. *International Journal of Medical Informatics*, *124*, 6–12.

Hingle, S. (2016). Electronic Health Records: An Unfulfilled Promise and a Call to Action. *Annals of Internal Medicine*, *165*(11), 818–819.

Hirschtick, R. E. (2006). A piece of my mind. Copy-and-paste. *JAMA*, *295*(20), 2335–2336.

Hu, Q., Huang, Z., ten Teije, A., & van Harmelen, F. (2015). Detecting new evidence for evidence-based guidelines using a semantic distance method. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9105*, 307–316.

Liang, J., Tsou, C.-H., & Poddar, A. (2019). A Novel System for Extractive Clinical Note Summarization using {EHR} Data. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 46–54. Minneapolis, Minnesota, USA: Association for Computational Linguistics.

Markel, A. (2010, May). Copy and paste of electronic health records: a modern medical illness. *The American Journal of Medicine*, Vol. 123, p. e9. United States.

Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk Management and Healthcare Policy*, *4*, 47–55.

Moradi, M., & Ghadiri, N. (2017). Quantifying the informativeness for biomedical literature summarization: An itemset mining method. *Computer Methods and Programs in Biomedicine*, *146*, 77–89.

Overhage, J. M., & McCallie Jr, D. (2020). Physician time spent using the electronic health record during outpatient encounters: a descriptive study. *Annals of internal medicine*, *172*(3), 169-174.

Patwardhan, S, & Pedersen, T. (2006). Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts. *Proceedings of the EACL 2006 Workshop Making Sense of Sense - Bringing Computational Linguistics and Psycholinguistics Together*, *1501*, 1–8. Trento, Italy.

Pivovarov, R., & Elhadad, N. (2015). Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, *22*(5), 938–947.

Sanchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. *Expert systems with applications*, 39(9), 7718-7728.

Shoolin, J., Ozeran, L., Hamann, C., & Bria, W. 2nd. (2013). Association of Medical Directors of Information Systems consensus on inpatient electronic health record documentation. *Applied Clinical Informatics*, *4*(2), 293–303.

Wang, Y., & Fang, H. (2016). *Extracting Useful Information from Clinical Notes*. 1–5.

Wrenn, J. O., Stein, D. M., Bakken, S., & Stetson, P. D. (2010a). Quantifying clinical narrative redundancy in an electronic health record. *Journal of the American Medical Informatics Association : JAMIA*, *17*(1), 49–53.

Wrenn, J. O., Stein, D. M., Bakken, S., & Stetson, P. D. (2010b). Quantifying clinical narrative redundancy in an electronic health record. *Journal of the American Medical Informatics Association*, *17*(1), 49–53.

Yadav, P., Steinbach, M., Kumar, V., & Simon, G. (2017). Mining electronic health records (EHR): A survey. *ArXiv*, *50*(6), 1–40.

Zhang, R., Pakhomov, S., McInnes, B. T., & Melton, G. B. (2011). Evaluating measures of redundancy in clinical texts. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2011*, 1612–1620. Retrieved from

Zhang, R., Pakhomov, S., & Melton, G. B. (2012). Automated identification of relevant new information in clinical narrative. *IHI'12 - Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 837–841.

Zhang, R., Pakhomov, S., & Melton, G. B. (2014). Longitudinal analysis of new information types in clinical notes. *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science*, *2014*, 232–237. Retrieved from

Zhang, R., Pakhomov, S. V. S., Arsoniadis, E. G., Lee, J. T., Wang, Y., & Melton, G. B. (2017). Detecting clinically relevant new information in clinical notes across specialties and settings. *BMC Medical Informatics and Decision Making*, *17*(Suppl 2). https://doi.org/10.1186/s12911-017-0464-y