

Using Language Models to Identify Relevant New Information in Inpatient Clinical Notes

Rui Zhang, PhD^{1,2}, Serguei V. Pakhomov, PhD^{1,3}, Janet T. Lee, MD, MS², Genevieve B. Melton, MD, MA^{1,2}

¹*Institute for Health Informatics*, ²*Department of Surgery*, and ³*College of Pharmacy*,
University of Minnesota, Minneapolis, MN, USA

Abstract

Redundant information in clinical notes within electronic health record (EHR) systems is ubiquitous and may negatively impact the use of these notes by clinicians, and, potentially, the efficiency of patient care delivery. Automated methods to identify redundant versus relevant new information may provide a valuable tool for clinicians to better synthesize patient information and navigate to clinically important details. In this study, we investigated the use of language models for identification of new information in inpatient notes, and evaluated our methods using expert-derived reference standards. The best method achieved precision of 0.743, recall of 0.832 and F1-measure of 0.784. The average proportion of redundant information was similar between inpatient and outpatient progress notes (76.6% (SD=17.3%) and 76.7% (SD=14.0%), respectively). Advanced practice providers tended to have higher rates of redundancy in their notes compared to physicians. Future investigation includes the addition of semantic components and visualization of new information.

Introduction

Clinical note documentation in Electronic Health Record (EHR) systems provides clinicians with the ability to store and share detailed contextual health information about patients for the primary purposes of communication, documentation, and billing. Most EHR systems allow functionality of “copy-and-pasting” of texts from a previous note to the current clinical note, which shortens the time clinicians spend on documenting encounters. However, an unintended consequence of copy-and-paste practices is creation of large amounts of replicated patient information within the EHR, especially in patients with complicated care or long hospital stays, thus making notes longer and less readable¹⁻³.

Notes with significant amounts of redundant information, combined with a large numbers of notes, increases the cognitive burden of clinicians³⁻⁸. In a time-constrained clinical practice environment, clinicians are limited in their review and synthesis of patient notes. Redundant information in clinical notes creates noise that masks new and clinically relevant information within notes. Moreover, redundant information in clinical notes can also contain a mixture of outdated information or errors in the copied information, making it difficult for clinicians to interpret the data in these notes effectively⁴.

Several studies have reported the effect of copy-and-paste documentation behavior in clinical practice⁹⁻¹¹. Redundant information can also create an integrity problem in clinical notes and create an impression that a note containing significant amounts of copied information is from an author who may not have read or independently constructed the note⁹. For example, in one report, a nurse observed that a historical event that occurred four years prior was subsequently repeated in many clinical notes afterwards⁹. This problem may also result in decreased use of and reliance on the information within clinical notes⁶. Other studies have demonstrated that the combination of redundant information and increased note length results in information overload and difficulties in finding information within notes, thus making narrative communication via notes less effective and efficient for patient care^{10,11}.

Previous studies have found large amounts of redundant information in both inpatient and outpatient notes with automated methods^{12,13}. Wrenn et al. used global alignment techniques to quantify redundancy in inpatient clinical notes¹². They found an average of 78% and 54% information duplicated from previous documents in signout and progress notes, respectively. Zhang et al. modified the Needleman-Wunsch algorithm to quantify redundancy and investigate the redundancy patterns in outpatient clinical notes¹³.

This work demonstrated that redundancy scores appeared to have a cyclic pattern for each individual patient but also that the overall volume of redundant information increased over time.

In this study, we investigated the use of statistical language models to identify clinically relevant new information in progress notes during patient hospital stays, applied a number of discounting models to potentially improve performance of the applied language models, and sought to compare the quantity of redundant information in clinical notes between inpatient and outpatient clinical settings.

Background

N-gram model

Statistical language modeling (SLM) is widely used for many NLP tasks, such as part-of-speech tagging, parsing, information retrieval, and machine translation^{14,15}. SLM assigns a probability to a set of n words based on a probability distribution from a specific corpus. An n -gram model is a typical language model (LM), which estimates the probability of an i -th word in the context of n previous words. To simplify the calculation of the probability of the word, the *Markov assumption* states that the probability of the word is only based on the prior few words instead of all previous words. One commonly used statistical estimate called Maximum Likelihood Estimates (MLE) is unsuitable for statistical inference in NLP due to the sparseness of the data. MLE assigns zero to unseen events, and the zeros will propagate since the probability of a long string is computed by multiplying probabilities of subparts.

Discounting methods

Discounting is the process of replacing the original counts with modified counts based on the mathematic formula to redistribute the probability in order to avoid assigning zero probability to unseen events due to the sparseness of the training sample texts^{14,15}.

The simplest one is Laplace smoothing (also called add-one method):

$$P_{Lap}(w_1 \cdots w_n) = \frac{C(w_1 \cdots w_n) + 1}{N + B}$$

where $C(w_1 \cdots w_n)$ is the count of the n -gram $w_1 \cdots w_n$, N is the number of training instances, and B is the vocabulary size.

The Good-Turing (GT)^a estimator is an improved method for determining the probability or frequency of n -grams:

$$\text{if } 1 \leq C(w_1 \cdots w_n) \leq \text{gt max}, P_{GT}(w_1 \cdots w_n) = \frac{C(w_1 \cdots w_n)}{C(w_1 \cdots w_{n-1})} \frac{C'(w_1 \cdots w_n) / C(w_1 \cdots w_n) - A}{(1 - A)}$$

$$\text{where } C'(w_1 \cdots w_n) = (C(w_1 \cdots w_n) + 1) \frac{n[C(w_1 \cdots w_n) + 1]}{n[C(w_1 \cdots w_{n-1})]}, A = (\text{gt max} + 1) \frac{n[\text{gt max} + 1]}{n[1]}$$

where $n[a]$ is the notion for the number of unique n -grams that occurred a times. This smoothing method substitutes low frequency n -grams and is quite accurate. It is also suitable for large numbers of observations of data and assumes that the distribution is binomial. The GT estimator works well for n -grams, despite the fact that words and n -grams do not follow a binomial distribution.

Ney and Essen proposed a linear discounting model for estimating frequencies of n -grams:

$$\text{if } C(w_1 \cdots w_n) = r, P(w_1 \cdots w_n) = \begin{cases} (1 - \alpha)r / N & \text{if } r > 0 \\ \alpha / N_0 & \text{otherwise} \end{cases}$$

where α is a constant slightly less than one.

^a <http://www-speech.sri.com/projects/srilm/manpages/ngram-discount.7.html>

These estimates make the probability of unseen events a small number instead of zero and rescale the other probabilities to ensure that the probability mass is equal to 1.0. However, the Ney and Essen linear discounting method does not work as well for higher frequency n -grams. In this study, we directly tested the use of these three discounting methods for building language models as part of our method evaluation.

Methods

Data Collection

EHR notes were retrieved from University of Minnesota Medical Center affiliated Fairview Health Services. For this study, we randomly selected patients in the inpatient clinical setting. These notes were extracted in text format from the EpicTM EHR system^b during a one-year period (05/2011 to 05/2012). For simplicity, we limited the notes to the progress notes authored by the primary team providers including physicians, residents, and advanced practice providers (including physician assistants (PA) and nurse practitioners (NP)). All notes were arranged chronologically for a given individual patients. Institutional review board approval was obtained and informed consent was waived for this minimal risk study.

Manually reviewed annotation as gold standard

Each series of in-patient notes starts with the patient's history and physical (H&P) note, followed by a set of progress notes, and ending with a discharge summary. Starting from the progress notes in a series of patient notes, two 4th-year medical students were asked to identify new and clinically relevant information based on all preceding documents chronologically within the same hospital stay using their clinical judgment. Each medical expert annotated progress notes from ten patients with one patient's set of notes overlapping with both. Annotation of new information in clinical notes was implemented through the publically available software General Architecture for Text Engineering (GATE)^c. GATE allows for annotation of text and XML outputs through a graphical user interface, with a customized annotation schema.

We first asked the two medical students to annotate one sample note and then to compare and discuss the annotations with each other to reach a consensus on the annotation categories and standards for new information (definitions and examples shown in Table 1). Each medical student later manually annotated another one set of patient notes based on the same historical notes to measure inter-rater agreement. Cohen's Kappa statistic and percent agreement were used to assess inter-rater reliability at a sentence or statement level.

Overall, longitudinal inpatient clinical notes from 20 sets of patient notes were annotated for this study. Each medical student annotated 50 notes, with a total of 100 annotated progress notes in this study. Fifty of these notes were used for training and system development and another fifty for evaluation. We also asked a 6th-year surgical resident (JL) with clinical practice experience to proofread the annotated notes, particularly around the addition of diagnostic studies (such as radiology reports) and make modifications to ensure the quality of the gold standard. We refereed annotations before and after JL's modification as initial annotation and revised annotation, respectively.

Results of automated methods were compared to the reference standard (initial annotation and revised annotation) and performance reported including accuracy, precision, recall, and F1-measure at a sentence or statement level. Methods were also evaluated comparing redundancy by different author role. For example, measuring redundancy of notes written by physician and residents as well as notes by physician assistants and nurse practitioners. Performance of the best method on identification of new information for different note sections was also tested.

^b <http://www.epic.com>

^c <http://gate.ac.uk>

Table 1. Classification, definition and examples of new clinical information.

Category	Code	Definition	Example
Clinical	Additional history	New information on patient's medical history occurring prior to hospitalization including prior diagnoses, surgeries, labwork/imaging, immunizations, medications, allergies/reactions, family history, and social history (e.g. sexual history, drug use, intake, work history, hobbies, marital status)	Recently admitted elsewhere for a pneumonia.
	Assessment	Changes in the medical diagnosis or differential diagnosis regarding patient's hospitalizing condition as assessed by the note's author	I am thinking this was all related to RF and CHF exacerbation.
	Changes in symptoms	Changes in patient reported symptoms including new complaints or improvement or worsening of existing concerns.	He says he is breathing better.
	Medications	Changes in medication, dosage, or route of administration including returns to pre-hospitalization usage. Also includes changes in fluids/electrolytes/nutrition administration and immediate post-administration clinical status.	Metoprolol 100 mg po bid changed to carvedilol 25 mg po bid
	New imaging and diagnostic studies	Results of imaging previously unmentioned during hospitalization such as X-rays, computerized tomography (CT scans), and solography.	TTE shows grade III diastolic dysfunction with elevated RSVP indicating pulmonary hypertension
	New labwork	Results of labwork previously unmentioned during hospitalization such as the basic metabolic panel, liver function tests, and blood glucose levels. Also includes fluid intake and output.	Wound cx: Gram + cocci UA shows a protein level of 300
	New plan	Changes in clinical care plan for patient care	Send sample for C. Diff studies
	Patient status	Changes in the patient's condition as reported by any healthcare provider	Perhaps a little improved today
	Physical exam	Daily physical examination as performed by clinician	Constitutional: Awake, alert, cooperative, no apparent distress
	Procedure	Procedures performed during hospitalization and related post-operative concerns	C2-4 laminectomy
	Vitals	Daily vitals including temperature, blood pressure, heart rate, and respiratory rate.	Temp: [98.1 F (36.7 C)-99F (37.2 C)] 98.2 F (36.8 C)
Non-clinical	Author	Name and degree of the note's author if previously unmentioned	LastName, FirstName, MD
	Change in service	Patient transfer to other hospital services or clinical sites	Patient will go to 6A for close Neuro-surgical monitoring.
	Date and time	Date and time of note signing by the original author	08/03/12 1148
	Social context	Changes in social history and situation relevant to the patient's condition and care including patient's preference to management	Dialysis discussed with patient, he prefers to wait until July.

Automated methods

We used different *n*-gram models with and without discounting algorithms. We only focused on bigram models since our prior studies¹⁶ have showed the bigram models outperformed than other *n*-gram models. The methods include six steps: 1) text preprocessing, 2) removal of classic stopwords^d and term frequency - inverse document frequency (TF-IDF) stopwords, 3) lexical normalization, 4) baseline modeling, 5) modification with discounting algorithms, and 6) application of heuristic rules to classify clinical relevance. The details of these steps are as follows:

- Step 1:* All progress notes were ordered by time for individual patients and were separated into sentences or sections. We used regular expressions for sentence splitter and word tokenization.
- Step 2:* Remove both classic stopwords and stopwords defined by optimal threshold of TFIDF distribution based on the entire note corpus. This step deemphasizes these less important words for building the language models.
- Step 3:* Use lexical variant generation (LVG)¹⁷ to normalize lexically different forms of the same term as equivalent when building the language models.
- Step 4:* Bigrams were counted in all previous notes for each individual patient. Probability of the bigrams without discounting algorithms was calculated and an optimal threshold value was used to identify new in the target progress notes.
- Step 5:* Discounting algorithms such as Laplace, Good-Turning, and Ney-Essen (details in Background) were used to calculate the probability of bigrams. An optimal threshold probability value was used to identify new versus redundancy information.
- Step 6:* Develop heuristic rules to judge the clinically relevance on section content, clinical note headers, signatures. For example, vitals in all notes are judged as relevant new information. All note headers and footer are non-relevant information.

Comparison of information redundancy between inpatient and outpatient notes

To investigate the difference of redundancy in outpatient and inpatient clinical notes, we calculated the information redundancy based on the reference standards using the below equation:

$$\text{Redundancy percentage} = 100 \times \left(1 - \frac{\# \text{ sentences with relevant new information}}{\# \text{ sentences}} \right)$$

Note that the medical experts and residents only annotated the new and clinically relevant information. Thus non-clinical new information such as header of the notes, signatures *etc.* was excluded. Averages, standard deviations, interquartile ranges of redundant and irrelevant information percentages for clinical notes in different clinical settings and note types were calculated.

Results

Annotation evaluation and method performance

Two medical students showed a good agreement on initial set of annotations for identifying new information in the overlapping annotations (Cohen's Kappa coefficient of 0.83 and percentage agreement of 92%). On subsequent review, JL found additional new information (average 3.80 sentences per note) and incorrectly annotated information (average 0.07 sentences per note). This resulted in two reference standards – the initial and the revised one. The performance characteristics of various algorithms on both reference standards are listed in Table 2. Generally, all discounting methods performed better than the baseline, although three discounting algorithms did not change significantly. Compared with revised annotation, the methods' precision increased significantly with small drop in recall. Recall of methods in notes written by advanced practice providers were higher than notes by physician or residents, while the precision was much lower. The bigram model with the Ney-Essen algorithm performed the best among these methods, achieving a recall of 0.832, a precision of 0.743, and F1-measure of 0.784 for all notes after JL annotation.

^d <http://www.textfixer.com/resources/common-english-words.txt>

Table 2. Performance of algorithms on identification of clinically relevant new information. Precision = TP/(TP+FP), Recall = TP/(TP+FN), F1-Measure = 2×Precision×Recall/(Precision+Recall).

Author type: All

Algorithms	Initial Annotation			Revised Annotation		
	Recall	Precision	F1-measure	Recall	Precision	F1-measure
Baseline	0.812	0.572	0.671	0.807	0.645	0.717
LapLace	0.827	0.654	0.730	0.826	0.728	0.774
Good-Turning	0.834	0.669	0.742	0.829	0.735	0.779
Ney-Essen	0.841	0.680	0.752	0.832	0.743	0.784

Author type: Physician & Resident

Algorithms	Initial Annotation			Revised Annotation		
	Recall	Precision	F1-measure	Recall	Precision	F1-measure
Baseline	0.800	0.587	0.677	0.800	0.667	0.733
LapLace	0.817	0.670	0.707	0.812	0.746	0.762
Good-Turning	0.824	0.681	0.746	0.820	0.758	0.788
Ney-Essen	0.830	0.692	0.755	0.825	0.767	0.795

Author type: Physician Assistant & Nurse Practitioner (Advanced Practice Providers)

Algorithms	Initial Annotation			Revised Annotation		
	Recall	Precision	F1-measure	Recall	Precision	F1-measure
Baseline	0.861	0.506	0.637	0.857	0.553	0.651
LapLace	0.918	0.517	0.662	0.917	0.576	0.707
Good-Turning	0.923	0.522	0.667	0.920	0.584	0.714
Ney-Essen	0.931	0.531	0.677	0.927	0.589	0.720

Performance on identification of new information in different sections

Percentages of new information identified based on the revised annotations and separated by section, are also shown in Table 3. Top three sections with most new information are Physical Exam (33%), Assessment & Plan (27%), and Medication (14%). Performance of the best method (Ney-Essen algorithm) was also evaluated in different sections. Recall of this discounting method in the Medication section was the highest, and precision and F1-measure in the Vitals section were the best.

Table 3. Identification of new information in different sections on revised annotation.

Sections	Percentages of new information in notes	Recall	Precision	F1-measure
Physical Exam	33%	0.860	0.820	0.839
Assessment & Plan	27%	0.910	0.612	0.732
Medication	14%	0.982	0.764	0.859
Vitals	10%	0.939	0.957	0.948
Imaging	5%	0.723	0.933	0.815

Redundant and irrelevant information in inpatient versus outpatient clinical notes

Statistical descriptions of information redundancy for clinical notes are listed in Table 4. It is surprising that outpatient clinical notes contain redundant information at the same level (about 76%) as those in inpatient clinical notes. Notes written by advanced practice providers contain more redundant information than those by physicians and residents. Notes written by physicians had a smaller arithmetic mean and a larger standard deviation of redundancy than the notes by residents, fellows, and advanced practice providers.

Table 4. Redundant and irrelevant information for different clinical settings. PA, physician assistant; NP, nurse practitioner. Redundancy includes the non-clinical but new information.

Clinical Setting	Author Type	# Notes	Redundancy/irrelevancy mean (standard deviation)	Redundancy/irrelevancy (interquartile range)
Inpatient	All	100	76.6% (17.3%)	(70.6%, 87.9%)
	Physician	57	73.3% (19.1%)	(63.0%, 86.8%)
	Resident or Fellow	15	84.4% (10.0%)	(66.3%, 87.3%)
	PA or NP	27	84.5% (7.3%)	(82.6%, 89.4%)
Outpatient	Physician	90	76.7% (14.0%)	(72.4%, 86.2%)

Discussion

Our investigation of patterns of relevant new information in the inpatient clinical practice highlights the issue of redundancy of clinical information in EHR documentation, which has been increasingly gaining the interest of clinicians and informaticians in recent years. Automated methods to accurately identify and visualize relevant new information represent a potential way to improve the clinicians' reviewing process. Although researchers have developed some preliminary methods to deal with redundancy, most previous evaluations do not include the clinicians' own judgments based on clinical experience as a gold standard and thus fall short in evaluating these methods. However, it is vital to include clinicians' views on redundant information for the development of the methods since they will be potential end users of any future system. In this study, we focused on the development of automated methods to identify relevant new information in inpatient clinical notes as well as evaluation for the methods by comparing with the reference standards annotated by the end users - clinicians.

Constructing a quality reference standard is an important but challenging task to support the development of robust automatic methods. We followed the same process as our previous study¹⁸ for reference standard: smaller sample annotation, discussion to reach a consensus, and then larger sample annotation. Although the annotation code book (Table 1) was meant to help with consistency of our medical student coders, we still found that they missed new information and our modified standard improved the consistency of these annotations. For instance, in the original set of annotations, coders sometimes ignored the changes of current medication list, including addition of a new drug or discontinuation of a current medication ("DISCONTD: sodium chloride 0.9 % flush 10 mL" as an example), and the possible reason is that it is difficult to find small changes in a long and tightly laid out section, such as a medication section. Other disagreements were from the different identification of new information boundary. For example, one annotator included the section title (e.g, objective, exam, assessment and plan) as new information if there was new information within the section; another annotator chose not to identify the title of section as new information.

After comparing the results produced by automated methods with reference standards, we found that discounting algorithms help to improve the performance of the methods. All methods did not perform very well on precision as our methods were developed on the lexical level of the texts and did not consider the semantic meaning of the sentences. For example, the sentence "continue to hold all nephrotoxic meds" in the target note was not identified as redundant by comparison with the sentence "hold lasix, lisinopril and spironolactone given acute kidney injury" in a previous note. Due to the limitations of the methods, they cannot recognize that specific drugs such as lasix, lisinopril and spironolactone are nephrotoxic medications.

The text formats in different sections result in variability in performance on different sections. “Assessment/Plan” (A&P) is one of the longest sections in the notes, where clinicians input their thoughts and tend to rephrase sentences even for the same meaning from the previous notes. For example, in the above example, the physician typed “nephrotoxic medications” instead of specific medication names “Lasix, lisinopril and spironolactone”. Thus, this probably is the main reason why the precision in the A&P section was the lowest. As for the Vitals section, the format is unique as “Temp: 97.2 °F (36.2 °C) | BP: 100/75 | Resp: 20...” in most notes, allowing the methods to easily recognize the pattern, resulting in a higher precision and recall compared with other sections. Similarly, the “Physical Exam” section usually contains the short statement for each part such as “General: Lying nearly flat in bed, comfortable, NAD, Interactive”.

Analysis indicated a high level of redundant and irrelevant information in inpatient progress notes (average 76.6% for 100 notes), although we included the irrelevant note format or noise in this calculation. This number may be different if we were to consider a larger dataset. To our knowledge, there is no prior study reporting the percentage of redundant information in outpatient notes. Surprisingly, the redundancy in outpatient notes contains the same amount of redundant information. One reason for this is that we only selected chronically diseased patients, allowing the larger sets of longitudinal clinical notes for our previous study. Another reason is that all those notes for calculating redundancy in outpatient clinical setting were the last three notes from each set of individual patients. In our previous study¹³, we found that the information redundancy of longitudinal outpatient notes was increasing over time. In other words, the last three notes tend to contain relatively higher redundancy than the earlier notes averagely. Therefore, the actual percentages of information redundancy for the entire set of outpatient notes could be lower than the reported number (76.7%) here in Table 3. In addition, we observed that advanced practice providers had higher levels of redundancy and less variability in this (lower standard deviation) than physicians. While it is unclear why this was the case, we speculate that physician providers have more diagnostic and case-based reasoning in notes with significant clinical events, and advanced practice providers are more prescriptive in their narrative. Future studies are needed to both confirm these findings on a larger corpus, as well as perform an analysis of why these differences may exist.

Our methods have certain limitations. All methods focused only on the lexical level. Semantic level issues were out of the scope of this paper, such as co-reference (e.g., “it”, “this”) and experiencer detection (e.g., “patient”, “sister”). For example, “Pt has diabetes” and “His mother has diabetes” shared most of the words, but they are semantically different as the experiencers are changed. Acronym and symbol disambiguation were also not included in the study. Moreover, relevant new information was only limited to the addition of information in the newer notes. The deletion of relevant new information in the more recent clinical notes was not considered in this study. Due to the asymmetric nature of the new information identification process, deletion of relevant information can only be obtained by comparing the object notes and target note in reverse and warrants additional investigation. Future research will add more semantic components to make the system more accurate and comprehensive, and design the ways to visualize the relevant new information by incorporating within existing EHR systems. This implementation will ultimately enhance the efficiency of reviewing and using clinical documentation, and improve the satisfaction of clinicians with EHR systems.

Conclusion

We developed language models with discounting algorithms to identify relevant new information in inpatient progress notes, and evaluated the performance by building up and comparing with a medical expert-derived reference standard. Inpatient clinical notes have approximately the same amount (76%) of redundant or irrelevant information as outpatient clinical notes. Further investigation is needed to improve the performance of the system and visualize the information in EHR systems to enhance the efficiency of using clinical documentation.

Acknowledgments

This research was supported by the Agency for Healthcare Research & Quality grant (#1R01HS022085-01) (GM), and University of Minnesota Clinical and Translational Science Award (#8UL1TR000114-02) (Blazer). The authors thank Fairview Health Services for support of this research.

References

1. Markel A. Copy and paste of electronic health records: a modern medical illness. *Am J Med.* 2010 May;123(5):e9.
2. Hirschtick RE. A piece of my mind. Copy-and-paste. *JAMA.* 2006 May 24;295(20):2335-6.
3. Yackel TR, Embi PJ. Copy-and-paste-and-paste. *JAMA.* 2006 Nov 15;296(19):2315; author reply -6.
4. Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. *AMIA Annu Symp Proc.* 2003:269-73.
5. Weir CR, Hurdle JF, Felgar MA, Hoffman JM, Roth B, Nebeker JR. Direct text entry in electronic progress notes. An evaluation of input errors. *Methods Inf Med.* 2003;42(1):61-7.
6. Hripcsak G, Vawdrey DK, Fred MR, Bostwick SB. Use of electronic clinical documentation: time spent and team interactions. *J Am Med Inform Assn.* 2011 Mar;18(2):112-7.
7. Patel VL, Kaufman DR, Arocha JF. Emerging paradigms of cognition in medical decision-making. *J Biomed Inform.* 2002 Feb;35(1):52-75.
8. Reichert D, Kaufman D, Bloxham B, Chase H, Elhadad N. Cognitive analysis of the summarization of longitudinal patient records. *AMIA Annu Symp Proc.* 2010;2010:667-71.
9. Embi PJ, Weir CR, Efthimiadis EN, Thielke S, Hadeen A, Hammond K. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. *J Am Med Inform Assoc.* 2013 Jul-Aug;20(4):718-26.
10. Embi PJ, Yackel TR, Logan JR, Bowen JL, Cooney TG, Gorman PN. Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians. *J Am Med Inform Assoc.* 2004 Jul-Aug;11(4):300-9.
11. Weir CR, Nebeker JR. Critical issues in an electronic documentation system. *AMIA Annu Symp Proc.* 2007:786-90.
12. Wrenn JO, Stein DM, Bakken S, Stetson PD. Quantifying clinical narrative redundancy in an electronic health record. *J Am Med Inform Assoc.* 2010 Jan-Feb;17(1):49-53.
13. Zhang R, Pakhomov S, MaInnes BT, Melton GB. Evaluating Measures of Redundancy in Clinical Texts. *AMIA Annu Symp Proc.* 2011:1612-20.
14. Manning CD, SchÜtze H. Foundations of Statistical Natural Language Processing. Cambridge, Massachusetts: The MIT Press; 2003.
15. Jurafsky D, Martin JH. Speech and Language Processing. Upper Saddle River, NJ: Prentice Hall; 2009.
16. Zhang R, Pakhomov S, Melton GB. Automated identification of relevant new information in clinical narrative. *2nd ACM SIGHIT Inter Health Inform (IHI) Symp Proc.* 2012: 837-41.
17. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Compt Appl Med Care* 1994:5.
18. Zhang R, Pakhomov S, Lee JT, Melton GB. Navigating longitudinal clinical notes with an automated method for detecting new information. *Studies in health technology and informatics.* 2013;192:754-8.