HTL Model: A Model for Extracting and Visualizing Medical Events from Narrative Text in Electronic Health Records

Eddie Paul Hernández[1], Alexandra Pomares Quimbaya[1] and Oscar Mauricio Muñoz[1,2]
[1]Pontificia Universidad Javeriana, Bogotá, Colombia
[2]Hospital Universitario San Ignacio, Bogotá, Colombia

Keywords: Text Mining, Clinical Practice Guidelines, Medical Events, Temporal Expressions, Electronic Health Records.

Abstract: Electronic health records contain important information of a patient and it may serve as source to analyze and audit the process of diagnosis and treatment of a specific clinical condition. This information is registered in narrative text, which generates a limitation to identify medical events like doctor appointments, medications, treatments, surgical procedures, etc. As it is difficult to identify medical events in electronic health records, it is not easy to find a point of comparison between this electronic information with recommendations given by clinical practice guidelines. Such guides correspond to recommendations systematically developed to assist health professionals in taking appropriate decisions with respect to illness. This article presents "Health Text Line Model HTL", a model for extraction, structuring and viewing medical events from narrative text in electronic health records. The HTL model was implemented in a framework that integrates the aforementioned processes to identify and timing medical events. HTL was validated in a general hospital giving good results on precision and recall.

1 INTRODUCTION

With the undeniable progress of information science and the expansion of its use in other areas of knowledge or even in some situations of our lives, we can find a large amount of data produced and stored on a daily basis. Discovering relationships, patterns and knowledge in these large volumes of data is being investigated and has become a great challenge (Oboler et al., 2011). One of the biggest challenges is text mining, which consists of applying a set of techniques to extract relevant and unknown information from large volumes of textual information, usually in unstructured natural language (Bentham and Hand, 2012).

Although there are many advances in the health area, there are still many unexplored fields and unsolved problems. One of these problems is given by the information contained in electronic health records (**EHRs)** due to limitations of applications, tools and models to improve the structuring, analysis and visualization of this information (Bentham and Hand, 2012). This problem is caused by a possible inaccuracy when studying the information recorded in these EHRs, because records are usually written in natural language (Laguna and Zaldumbide, 2007). Therefore, the analysis of textual content in EHRs is expensive in terms of time, and limited when we want to identify medical events, because usually the information from EHRs is stored in an unstructured or narrative text. Due to the limitation of identifying medical events (appointments, medical prescription, treatments, surgical procedures, etc.), it is very difficult to compare the information recorded in the EHRs with treatment guidelines. The recommendations given by clinical practice Guidelines are syntheses of best available evidence that support decision making by clinicians, managers, and policy makers (Gagliardi et al., 2011).To solve these problems, it is presented the "HTL Health Text Line" model, a model for extracting, structuring and visualizing of medical events from narrative text recorded in EHRs. This model helps health professionals to evaluate adherence to clinical practice guidelines, giving clinicians tools to audit and feedback as a good strategy to improve professional practice, either on its own or as a key component of multifaceted quality improvement interventions (Ivers et al., 2012). HTL is based on a model that uses techniques of natural language processing techniques, timing of events and text mining. HTL has three major

Hernández, E., Quimbaya, A. and Muñoz, O.
HTL Model: A Model for Extracting and Visualizing Medical Events from Narrative Text in Electronic Health Records.
In *Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2016)*, pages 107-114
ISBN: 978-989-758-180-9
Copyright © 2016 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

processes (Extraction, Structure and Display). The extraction process is responsible for obtaining facts from the EHR and timing of medical events (corresponding to the structuring of medical events), this process is applied once the information has been collected from medical records, and later there is the visualization process that allows medical events to be represented as a timeline of EHRs.

This article presents in Section 2 an analysis and classification of related works for structuring and viewing of narrative text. Section 3 presents an overview of the "Health Text Line Model HTL" for extraction and structuring of medical events from narrative text in EHRs. Section 4 describes the functionality of the model through an HTL Framework. Later, in Section 5 the model is validated through a case study and Section 6 addresses the conclusions and future work.

2 STATE OF THE ART

The processing of unstructured or narrative text to extract knowledge applies robust analytical techniques used in Natural Language Processing (NLP) (Mani et al 2005), Information Extraction (IE) (Kianmehr et al., 2009), Sentiment Analysis (SA) (Jeonghee et al., 2003), extraction of temporal information or events (TIE) (Cheng et al, 2007)., such as annotation of entities, regular expressions, automatic generation of summaries, event extraction, textual polarity, clusters, time expressions or temporal relationships. Although significant efforts have been made in the notation, representation and knowledge extraction from narrative texts, in recent years it has increased the interest in the extraction of temporal information in the medical domain (Clark et al., 2011). Considering EHRs contain an overwhelming amount of information stored in narrative text (Savova et al., 2010), to organize this information in a temporal way will lead to, for example, determine whether a drug has been prescribed, if a patient has had some disease symptoms or when an lab test, assessment or treatment is scheduled. Jindal and Roth proposed three tasks to do: Removal of (a) events, (b) temporary expressions and (c) temporal relationships. They use clinical reports (unstructured data) to automatically annotate medical events. Denny et al. developed a system to extract, from EHRs, the clinical concepts, dates to identify the status of colon and rectal cancer, in order to save time and effort in analyzing the medical records of patients with this pathology. Liu et al. built a framework to extract patients who had been exposed to drugs. They identified if a patient had any exposure to the drug warfarin when being admitted to the hospital. Ferro et al., Mani et al. and Zhou et al. scored narrative temporal expressions in EHRs. HeidelTime and SUTime extracted and normalized time expressions through rules, patterns and regular expressions (Strötgen and Gertz, 2012). Although there are studies that focus on these aspects (Savova et al., 2011), it is important to consider the identification of medical events and temporal relationships. This aspect is reflected in TimeML, which provides an annotation scheme for identifying and orienting events on a timeline (Pustejovsky et al., 2003). The Framework TimeML denotes the aspects mentioned above supported by: TIMEX3, EVENT and TLINK (Sauri et al, 2005); TIIMEX3 is mainly used to make explicit temporal expressions, such as times, dates, duration, etc. EVENT identifies events and TLINK tag is used to represent temporal relationships between events or between an event and a temporary expression.

One of the most influential works in the identification of medical events and their relationship to time expressions shows an automatic system to extract events and the relationship of these using an natural language processing architecture. Its scores reach 90% for extracting temporal expressions and 87% for extracting clinical events (Kovacevic et al., 2013).

There are different approaches and works for the extraction and structuring of medical events that have been supported for decision-making and analysis of information in EHRs. However, there are some limitations that should be considered to improve or ensure a complete chronology of events of a medical patient. The criteria and aspects evaluated are: Entity types, temporary expressions identification, medical events, occurrence and temporal relationships as defined below:

i) **Entity Types:** indicates whether the work takes into account the following medical institutions: drugs, diseases, lab tests, medical dosage units.
ii) **Time Expressions:** indicates whether the project studies in depth adverbs of time and time units.
iii) **Medical Events:** indicates whether the project gives importance to medical events. A medical event can be defined as anything that is clinically important and which can also be assigned to a timeline.
iv) **Occurrence:** indicates whether the project considers this criterion to describe if an event actually occurred; that is, the level of uncertainty of the medical event.

Table 1: Related Works.

PAPERS	Entity types	Time expressions	Mediical events	Occurrence	Temporal relations
(Kovacevic et al., 2013)	X	X	X	?	
(Denny et al., 2010)	X	X	?		
(Jindal and Roth, 2013)		X	X		X
(Savova et al., 2011)	X	X	X	?	X
(Wenzina and Kaiser, 2013)	X	X			
(Sun et al., 2013)		X	X		X
(Raghavan et al., 2012)		X	X		X
(Uzuner et al., 2013)	X	X	X		X
(Sauri et al., 2012)	X	X			X
(Reeves et al., 2012)		X	X		

v) **Temporal Relations:** indicates whether the project takes into account if there is a temporary link between two events. The possible events can be: change dosage schedule changes of drugs and / or changes in the characteristics of the tests.

The "X" symbol indicates that the investigation contains the criteria of the table. The symbol "?" Indicates that the investigation does not specify the criteria, but it may have used it. In conclusion, there is no project containing all the criteria and some of them have limitations; for this reason we proposed HTL Model (Health Text Line).

3 HTL HEALTH TEXT LINE

HTL (Health Text Line) is a model that identifies medical events associated with medications, diseases and tests from the narrative texts contained in EHRs. HTL is based on narrative text analysis to identify the medical event and the time when it occurred, to establish the level of certainty of that event. Additionally, it is a model that uses text mining techniques and natural language processing as: Stop words, Tokenization, Splitter, Part of speech (POS), Named Entity Extraction, Gazetteer, among others. The model also takes into account the most relevant criteria for knowledge extraction or removal, structuring and viewing medical events: entity types, temporary expressions, medical events, occurrence, temporal relations and event display.

HTL is composed by three processes (see Figure 1). The first process is the Extraction of medical event, which consists of a set of threads, methods and techniques to prepare the narrative texts of EHRs for major operations of knowledge discovery. This process allows the identification of entities (Medication, Test and disease) and medical events. The second process is the structuring of medical events, which is in charge of structuring the events and sub-events with their respective medical context; by identifying grammatical connectors, tenses and regular expressions. The third process is the visualization of medical events, whose input is obtained as a result of the above two processes and allowing initiating the visualization of the patient's lifeline.

SCIENCE AND TECHNOLOGY PUBLICATIONS

3.1 Extraction Process

The extraction process is divided into four subprocesses. It has entries (Narrative text HCE and knowledge base Gazetter) and the result of this process is stored in a repository called repository of Extraction of Medical Event). The subprocess that establishes the extraction process includes:

3.1.1 Stop Words

A method or filter process, which main objective is to eliminate the words that provide little information or not represent any important content (Paass G et al.,2005).

3.1.2 Tokenization

Tokenization divides the text into simple tokens, such as numbers, punctuation and words of different types. (Paass G et al., 2005).

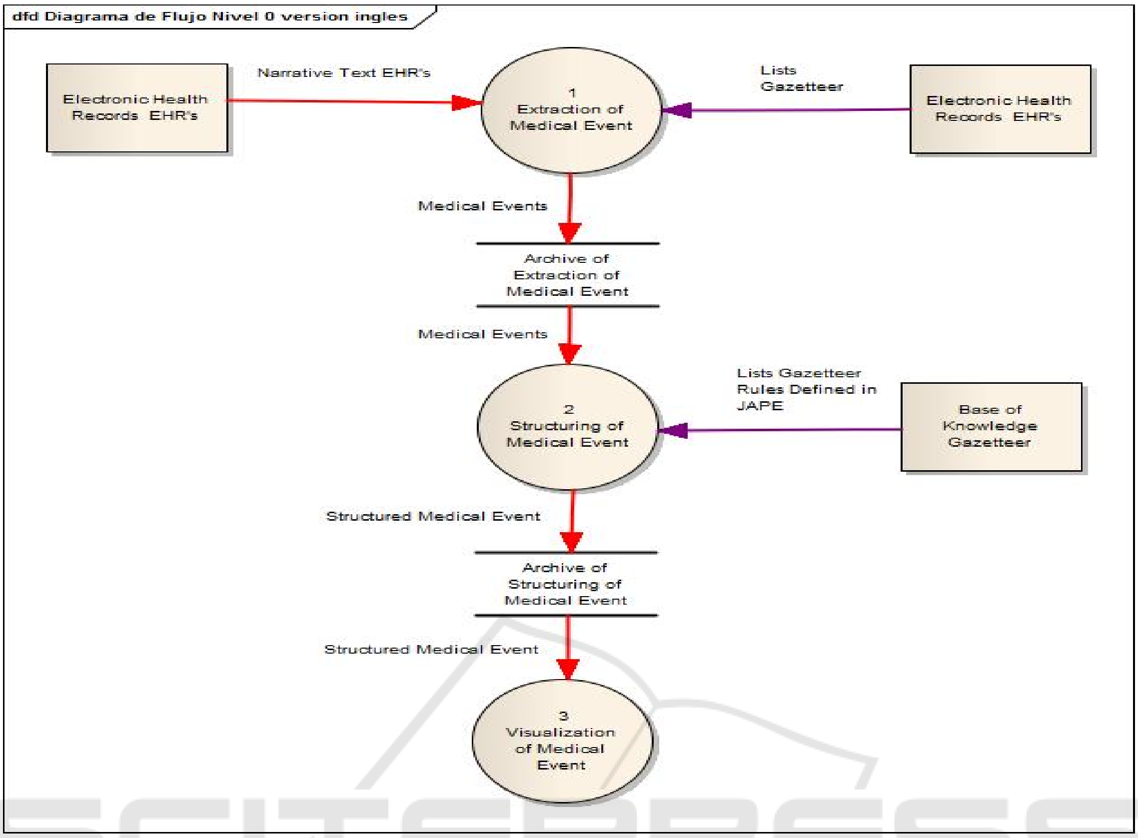

Figure 1: Model Extraction, Structuring and Visualization of Medical Events.

3.1.3 Splitter

The Splitter is a cascade of finite-state transducers that segments the text into sentences. The divider uses a list of abbreviations gazetteer to help distinguish a new paragraph. Each phrase is annotated with the type 'Sentence' (Paass G et al., 2005).

3.1.4 Part of Speech (POS)

It is the process of assigning (or tag) to each of the words in a text their grammatical category. This process can be executed in accordance with the definition of the word or the context in which it appears (Paass G et al., 2005).

3.2 Structuring Process

The structuring process consists of three subprocesses: recognition of the organization, identification of subevents and identification of context). It takes the information from the repository of Extraction of Medical Event and Knowledge Base Gazetter to generate a repository of Structuring of Medical Event. The subprocesses included are:

3.2.1 Entity Recognition

This process is responsible for identifying the entity where the medical event comes from. It relies on the various gazetteers lists of entities that conforms the base of knowledge. The lists of entities are: diseases, lab test, medical units, drug, adverbs of time, connector's grammar, verbs and time units.

3.2.2 Identification of Subevents

This is the process for identifying subevents (dosage changes, changes of medication schedules and / or changes in the lab tests characteristics) corresponding to one entity and a medical event. This process consists of the following steps:

i) **Identification of Grammatical Connectors:** This process is responsible for identifying the different grammatical connectors that exist on EHRs, which set the pattern for an event split in sub-events.

ii) **Recognition of Verbs and Tenses**: Process responsible for identifying from a list of verbs, the verbs commonly used in EHRs. Verbs in past, present and future due to the need to recognize whether the sub-events occur in the past, or are a current event or are an event that will happen in the future.

iii) Identification of Regular Expressions: A process supported by a set of rules or patterns defined in JAPE (Java Annotation Patterns Engine).

3.2.3 Context Identification

Process supported by defined characteristics as Quantum (entity, characteristic of quantity, time characteristic and adverb of time), which can identify the context of a medical subevent in EHRs. The steps for making the identification process of the context and Quantum characteristics are described below:

i) Extraction of Adverb of Time: Process responsible for identifying the different adverbs of time that exist in the EHRs, to be applied to the sub-events (complement the process initiated by the identification of verbs and verbal tenses).

ii) Characteristic Identification Associated to Quantity: Process responsible for identifying the characteristics associated with an amount (dosage or intensity) of an entity in a sub-event that means, the change of the quantity unit in an entity

iii) Identification of Time Characteristics: The process responsible for identifying the characteristics associated with time (days, day, night, fasting, month, week) of an entity in a sub-event, whether an entity changes the unit of time.

iv) Occurrence Identification or Medical Event Uncertainty: The process classifies the level of uncertainty of the medical event in Medical probable event and Medical uncertain Event; it can identify whether a medical event has a high probability of occurrence or not (see Figure 2).

- **Medical Probable Event:** It is identified when there is an important probability of occurrence of a medical event; although there is some uncertainty about the date of the medical attention, there is an approach about the reference date. Using the previous EHRs.

- **Medical Uncertain Event:** It is identified when there is very little certainty of the occurrence of a medical event; unlike the probable date there is a very high level of uncertainty (impossibility of identification relate to the reference date).

3.3 Visualization Process

This process of the model HTL is aimed at displaying medical events that have been previously extracted and structured. It allows health professionals to have a point of comparison between that recorded in the EHRs and clinical practice guidelines. The visualization process consists of three threads: Align, Rank and Filter. The lifeline was made by the University of Maryland and the Institute for Advanced Computer Studies (UMIACS) (Plaisant, et al., 1998). The subprocesses that integrate the Visualization process are:

3.3.1 Align

Process that allows aligning all EHRs for a specific event type (for example diabetes).

3.3.2 Rank

Process that classifies EHRs depending on the number of occurrences of a particular event. For example, the number of surgeries to a patient or the number of abnormal results in taking blood pressure. It can also be ordered considering the judgment of the medical specialist, the most relevant or important events of this classification.

3.3.3 Filter

Interactive process that allows health professionals to filter EHRs to find temporal patterns of medical events. For example, high glucose, diabetes or ibuprofen, headache. That is, the process of filtering enables searching for particular sequences of events, including both the presence and absence of events.

4 IMPLEMENTATION

To perform the validation of the strategy proposed by the HTL model, a software tool that follows each of its phases and processes was implemented. This software tool is called HTL Framework, a software that allows the extraction of medical events, structuring of medical events and visualization of medical events from the narrative text in EHRs.

As it can be seen in Figure 2, the identification of medical events takes into account whether the occurrence of these events are uncertain or probable. These are placed in the lifeline depending on the

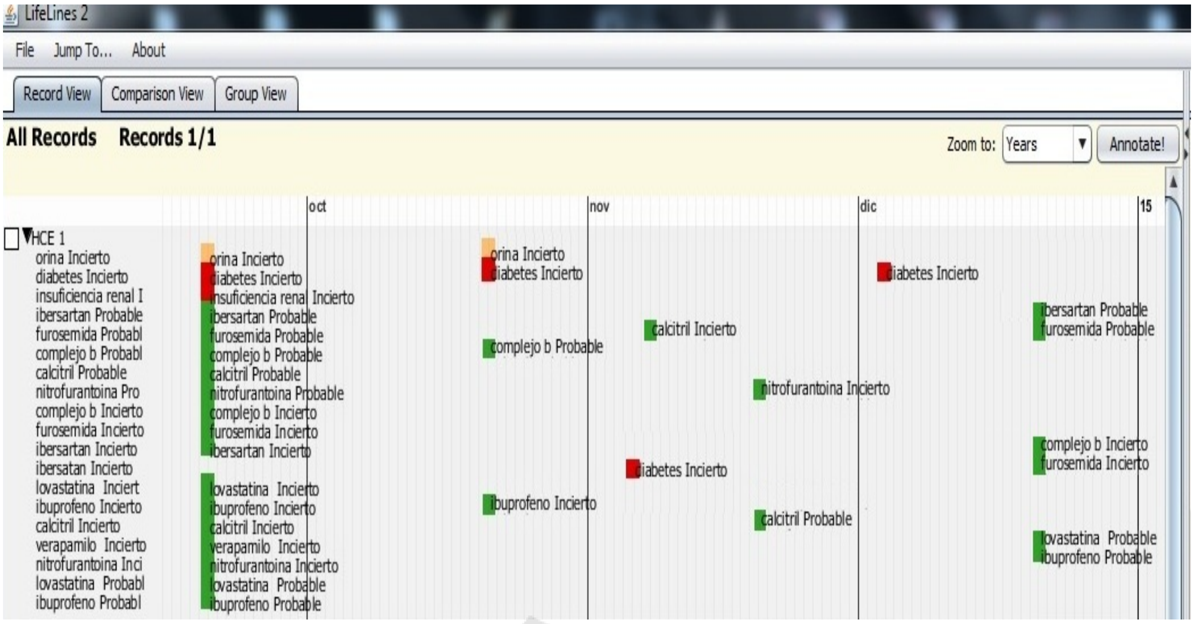

Figure 2: Lifeline2.

date calculated in the extraction process. Besides, medical events are represented by the type of entity by a specific color:

A Software Engineering process was used to implement the HTL framework applying the Agile Model Driven Development (AMDD) (Scott et al., 2008). The development of the Framework was made using the programming language Java JDK7.1.

5 VALIDATION

This evaluation and validation is performed through a case study applied to the narrative of EHRs. These EHRs stored information of diagnosis, treatment and monitoring of patients in a general hospital. In the study case functionality tests were performed comparing the results generated manually against all HTL generated results (response). To perform this task the metrics precision and recall were measured.

5.1 Precision Metric

The precision metric (P) is defined as the proportion of relevant retrieved events among all retrieved events.

5.2 Recall Metric

The Recall metric (R) is defined as the proportion of the relevant events that were recovered from all the relevant events available.

5.3 Results Analysis

Three initial iterations were performed to the model formed of a set of 40 EHRs. These EHRs were randomly selected in a period of one year.

When performing the iterations with the three initial data sets, there were able to obtain percentages of 81% and 72% related to the identification of entities and medical events. Although these data are satisfactory, it was necessary to optimize the rules and refine the knowledge base to increase the percentage of metrics. The results of the precision and recall metrics obtained by the model for the set of evaluation are shown below in Table 2.

The values of accuracy and recall criteria for identification of Temporal Expressions and Medical Events were compared with the works that achieved better results in the analysis accomplished in the literature, Tang et al., Raghavan et al. and Kovacevic et al. HTL yields better results when identifying temporal expressions and the narrative of EHRs.

Table 2: Accuracy and Recall of HTL.

Criteria	P (%)	R (%)
Type of Entity	96%	93%
Temporal Expressions	94%	92%
Clinical Events	92%	89%
Occurrence	85%	82%
Temporal Relations	84%	80%

It is concluded that the results of accuracy and completeness to identify medical events are satisfactory for refining the knowledge base and optimizing the rules. However, not closer to 100% accuracy values are reached due to the failure to identify certain entities within the narrative were misspelled, and therefore did not belong to the knowledge base.

6 CONCLUSIONS AND FUTURE WORK

There is a wide range of benefits in the different areas of application of text mining, from a greater understanding of customer needs until the discovery of fraud in banks. A clear example is the many advances in the health area and more specifically in the study of the information recorded in EHRs. However, there are still limitations in identifying medical events because usually the information from EHRs is stored in an unstructured or narrative text, which leads to the loss of much potentially important information. From the analysis realized of the related literature it was established that although there are different approaches and works for structuring and visualization of medical events from EHRs, there are still some limitations that should be considered to improve or ensure a complete chronology of events of a medical patient record.

HTL is a model that identifies medical events associated with medications, diseases and tests from the narrative contained in electronic health records. It is based on the analysis of narrative text to identify the medical event and the time at which it occurred. HTL includes the extraction, structuring and visualization process of medical events which are of vital importance when performing tasks or procedures of medical reasoning. Likewise, in the process of viewing it creates a lifeline easier to understand for health professionals to have a point of comparison between what is recorded in the EHRs and clinical practice guidelines. The strategy proposed by the HTL model, was implemented in a Framework. HTL gives to the health professional a tool to evaluate the occurrence of some medical events.

The model and framework were validated through a case applied to the narrative text of EHRs of a general university hospital. In the study case, functional tests were performed using the precision and recall metrics, which returned values of 94%, 92%, 92% and 89% respectively for the identification of temporal expressions and medical events. These values exceeded the digits obtained by others research. It is important to stand out that HTL model results can be strategically used to more easily understand and analyze the overall structure for EHRs, due to the benefits in delivering structured information and its visual display on a lifeline. By understanding and exploiting this lifeline, the time can be reduced, improve accuracy in the results of medical research and thus discover unknown information.

As future work HTL could be used and validated with other health professionals (nutritionists, veterinarians, and psychologists), although they do not use EHRs yet, they store information in narrative form. In addition, it would be created a module in the framework to allow users the option of create and modify regular expressions or patterns contemplate by HTL in order to increase the accuracy of the model and narrative analysis of EHRs.

ACKNOWLEDGEMENTS

This work is part of the projects funded by Hospital Universitario San Ignacio and Pontificia Universidad Javeriana for improving the analysis of electronic health records.

REFERENCES

Oboler, A., Welsh, K., Cruz, L. (2012). The danger of big data: *Social media as computational social science. Scopus*.

Bentham, J., Hand, D. (2012). Data mining from a patient safety database: *The lessons learned Data Mining and Knowledge Discovery*. Scopus.

Laguna, J., Zaldumbide, Y. (2007). Documentación clínica: Aspectos Legales y fuente de información para las bases de datos hospitalarias. *Revista Española de Cardiología suplementos*.

Gagliardi, R., Brouwers, C., Palda, V., Lemieux-Charles, L., and Grimshaw, J. (2011). How can we improve guideline use? A conceptual framework of implementability. *Implementation Science : IS, 6(1), 26.*

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J., Odgaard-Jensen, J., French, S., Johansen, M. (2012). Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews 2012, (6),* 1–227.

Mani, I., Pustejovsky, J., Gaizauskas, R. (2005). *The language of time: a reader*. Oxford University Press.

Kianmehr, K., Gao, S., Attari, J., Rahman, M. M., Akomeah, K., Alhajj, R., Rokne, J., and Barker, K.

(2009). Text summarization techniques: SVM versus neural networks. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, iiWAS '09, pages 487–491, New York, NY, USA. ACM.

Jeonghee, Yi., Tetsuya, N., Razvan, B., and Wayne, N. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*.

Cheng, Y., Asahara, M., Matsumoto, Y. (2007). NAIST.japan: temporal relation identification using dependency parsed tree. *Proceedings of the 4th. International Workshop on Semantic Evaluations. Association for Computational Linguistics*.

Clark, K., Sharma, D., Chute, C. (2011). Application of a temporal reasoning framework tool in analysis of medical device adverse events. *AMIA Annual Symposium Proceedings*. American Medical Informatics Association,

Savova, K., Masanz, J., Ogren, P., Sunghwan, J., Karin, S., Chute, C. (2010). Clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*.

Sun, W., Rumshisky, A., Uzuner, Ö. (2013). *Evaluating Temporal Relations in Clinical Text: 2012 i2b2 Challenge Overview*. J Am Med Inform Assoc.

Jindal, P., Roth, D. (2013). Extraction of events and temporal expressions from clinical narratives, *Journal of Biomedical Informatics*, 46, p.S13-S19.

Kovacevic, A., Dehghan, A., Filannino, M., Keane, J., Nenadic, G. (2013). Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. J Am Med Inform Assoc.

Denny, J., Peterson, J., Choma, N. (2010). *Extracting timing and status descriptors for colonoscopy testing from EHRs*. J Am Med Inform Assoc.

Liu, M., Jiang, M., Kawai, V. (2011). Modelling drug exposure data in EHRs: an application to warfarin. *AMIA Annual Symposium Proceedings*. American Medical Informatics Association.

Mani, I., Wilson, G., Sundheim, B., and Ferro, L. (2001). Guidelines for Annotating Temporal Information. In Proceedings of HLT 2001, *1st International Conference on Human Language Technology Research*, J. Allan, ed., Morgan Kaufmann, San Francisco.

Raghavan, P., Fosler, E., Lai, A. (2012). Inter-annotator reliability of medical events, coreferences and temporal relations in clinical narratives by annotators with varying levels of clinical expertise. In: *AMIA Annual Symposium Proceedings*, p. 1366. American Medical Informatics Association.

Uzuner, Ö. Stubbs, A., Sun, W. (2013). Chronology of your health events: Approaches to extracting temporal relations from medical narratives. *Journal of biomedical informatics*.

Wenzina, R., Kaiser, K. (2013). Identifying Condition-Action Sentences Using a Heuristic-based Information Extraction Method. *Proc. Jt. Int. Work. KR4HC/ProHealth*.

Reeves, R., Ong, F., Matheny, M., Denny, J., Aronsky, D., Gobbel, G., Montella, D., Speroff, T., Brown S. (2013). *Detecting temporal expressions in medical narratives*. Int J Med Inform.

Sun, W., Rumshisky, A., Uzuner, Ö. (2012). Evaluating Temporal Relations in Clinical Text: 2012 i2b2 Challenge Overview. *J Am Med Inform Association*.

Ferro, L. (2001). *Instruction Manual for the Annotation of Temporal Expressions*. MITRE Technical Report MTR 01W0000046. McLean, Virginia: The MITRE Corporation.

Strötgen, J., Gertz, M. (2012). Multilingual and cross-domain temporal tagging. Lang Resour Eval.

Chang, A., Manning, C. (2012). SUTIME: a library for recognizing and normalizing time expressions. In: *8th International conference on language resources and evaluation (LREC 2012).*

Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., and Katz, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. *IWCS-5, Fifth International Workshop on Computational Semantics*.

Sauri, R., Littman, J., Knippen, B. (2005).TimeML annotation guidelines, V.1.2.1. Available at: http://www.timeml.org/timeMLdocs/AnnGuide14.pdf.

Chen, E., Hripcsak, G., Xu, H., Markatou, M., Friedman, C. (2007). Automated Acquisition of Disease-Drug Knowledge from Biomedical and Clinical Documents: An Initial Study. J Am Med Inform Assoc.

Meystre, S., Savova, G., Kipper-Schuler, K. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform. 35:128–44.

Savova, G., Chapman, W., Zheng, J., Crowley, R. (2011). *Anaphoric relations in the clinical narrative: corpus creation*. JAMIA; 18(4): 459-465.

Paass, G., Hotho, A., Nurnberger, A. (2005). *A brief survey of text mining. IEEE,*

Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., Shneiderman, B. (1998). *LifeLines Using visualization to enhance navigation and analysis of patient records*. Proc. AMIA Fall Symposium.

Scott, A. (2008). Agile Model Driven Development (AMDD).*The Key to Scaling Agile Software Development.*