

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA  
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA  
MESTRADO PROFISSIONAL EM GESTÃO E TECNOLOGIA EM  
SISTEMAS PRODUTIVOS

GIOVANA FADINI DE OLIVEIRA

INDETERMINAÇÃO TEMPORAL E DADOS TEMPORAIS EM NARRATIVAS E  
TEXTOS CLÍNICOS – UMA REVISÃO SISTEMÁTICA DA LITERATURA

São Paulo  
Março/2018

GIOVANA FADINI DE OLIVEIRA

INDETERMINAÇÃO TEMPORAL E DADOS TEMPORAIS EM NARRATIVAS E  
TEXTOS CLÍNICOS: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Dissertação apresentado como exigência parcial para a obtenção do título de Mestre em Gestão e Tecnologia em Sistemas Produtivos do Centro Estadual de Educação Tecnológica Paula Souza, no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos, sob a orientação do Prof. Dr. Marcelo T. Okano.

São Paulo

Março/2018

GIOVANA FADINI DE OLIVEIRA

INDETERMINAÇÃO TEMPORAL E DADOS TEMPORAIS EM NARRATIVAS E  
TEXTOS CLÍNICOS: UMA REVISÃO SISTEMÁTICA DA LITERATURA

---

Prof. Dr. Marcelo T. Okano

---

Prof. Dra. Marcia Ito

---

Prof. Dr. José Manoel Souza das Neves

São Paulo, 28 de Março de 2018

Dedico esta dissertação de Mestrado a todos  
que de alguma forma contribuíram para que  
ela fosse concluída com êxito.

## **AGRADECIMENTOS**

Primeiramente, agradeço à orientação e paciência dos meus orientadores Marcelo Okano e Marcia Ito, que me ajudaram a chegar até aqui, com muitas e valiosas orientações e conversas que permitiram com que esse trabalho fosse desenvolvido.

Aos membros da banca, Prof Dr. José Manoel e Profª Dra. Marcia Ito, pelas valiosas recomendações, sendo de grande valia na composição deste trabalho.

Aos colegas e membros da Unidade de Pós-graduação do Centro Paula Souza pelo apoio nessa jornada.

Um agradecimento especial à minha família, pela compreensão nos momentos de ausência, principalmente à minha mãe Adelaide, mulher forte e guerreira, que sempre me apoiou e me incentivou aos estudos, agradeço também imensamente à minha querida avó Antonia e avô Alcides, por terem me criado com tanto carinho e amor e com certeza são responsáveis pelo que sou hoje, um agradecimento especial também ao meu querido e amado irmão Bruno, que não está mais entre nós, mas com certeza foi e sempre será uma pessoa inspiradora para mim, além de um ser humano incrível. Agradeço também a todos os que estiveram envolvidos, diretamente ou indiretamente nesse trabalho, pela paciência e companheirismo nos momentos em que mais precisei de ajuda.

Os meus profundos agradecimentos a todos que de alguma forma contribuíram para que esse sonho se tornasse realidade, aos funcionários da Etec Martinho de Ciero, principalmente à direção e coordenação, que por muitas vezes se desdobraram para que eu pudesse participar dos eventos necessários e nos momentos de ausência necessária para a produção desse trabalho.

Meus imensos agradecimentos ao Centro Estadual de Educação Tecnológica Paula Souza, pelo incondicional apoio, que foi fundamental para que esse mestrado fosse concluído.

“Tudo o que temos de decidir é o que fazer  
com o tempo que nos é dado” (J. R. R.  
Tolkien)

## RESUMO

DE OLIVEIRA, G. F. **Indeterminação Temporal e Dados Temporais em Narrativas e Textos Clínicos**: Uma revisão sistemática da literatura 81f. Dissertação apresentada ao programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos. Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2018.

O presente trabalho tem por objetivo apresentar uma revisão sistemática da literatura, a fim de selecionar pesquisas que apresentem os tipos de tratamento e os métodos computacionais que são realizados para o tratamento da questão temporal e indeterminação temporal em textos da área da saúde, narrativas e textos de autorias de pacientes. A metodologia usada na pesquisa adotou a abordagem quali-quantitativa. A forma de raciocínio lógico utilizada foi o método indutivo. Em relação à natureza das fontes utilizadas a pesquisa se classifica como bibliográfica e experimental. Foi realizada a pesquisa bibliográfica utilizando-se do método de revisão sistemática, a fim de identificar o estado da arte quando se trata de problemas com dados temporais e de indeterminação temporal. Com o resultado da revisão sistemática, foram identificados 18 artigos, sendo que 11 apresentavam tratamentos e métodos realizados em dados temporais e 7 em indeterminação temporal. Foi verificado também que existiam várias ferramentas e softwares que tratavam de problemas relacionados à questão temporal dos dados, mas que não eram eficientes e que nos artigos selecionados, algumas propostas de melhorias, extensões e novos métodos e algoritmos foram apresentados, a fim de proporcionar uma melhor eficiência nessas ferramentas e softwares, além de proporem novas alternativas para o trabalho com dados temporais, possibilitando a indicação de respostas para as questões de pesquisa elencadas, contribuindo assim para a área da saúde.

**Palavras-chave:** Banco de Dados Temporal; Dado Temporal; Indeterminação Temporal; Narrativas Clínicas; Sistemas Produtivos.

## ABSTRACT

DE OLIVEIRA, G. F. **Indeterminação Temporal e Dados Temporais em Narrativas e Textos Clínicos:** Uma revisão sistemática da literatura 81f. Dissertação apresentada ao programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos. Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2018.

The present work aims to present a systematic review of the literature, in order to select researches that present the types of treatment and computational methods that are performed for the treatment of the temporal issue and temporal indetermination in health texts, narratives and texts authorship. The methodology used in the research adopted the quali-quantitative approach. The form of logical reasoning used was the inductive method. In relation to the nature of the sources used the research is classified as bibliographic and experimental. The bibliographic research was carried out using the systematic review method, in order to identify the state of the art when dealing with problems with temporal data and temporal indetermination. With the result of the systematic review, 18 articles were identified, 11 of which presented treatments and methods performed in temporal data and 7 in temporal indetermination. It was also verified that there were several tools and software that dealt with problems related to the temporal question of the data, but that were not efficient and that in the articles selected, some proposals for improvements, extensions and new methods and algorithms were presented in order to provide a better efficiency in these tools and software, and propose new alternatives for working with temporal data, enabling the indication of answers to the research questions listed, thus contributing to the health area.

**Keywords:** Database, Temporal Database; Temporal Data; Temporal Indeterminacy; Clinical Narratives; Productive Systems.



## LISTA DE QUADROS

<b>Quadro 1:</b> Planejamento da Revisão Sistemática .....	21
<b>Quadro 2:</b> Bases de periódicos utilizados nas revisões sistemáticas em Computação e Medicina .....	40
<b>Quadro 3:</b> Termos sinônimos e grupos de pesquisa.....	41
<b>Quadro 4:</b> Strings de busca .....	41
<b>Quadro 5:</b> Relação dos artigos selecionados.....	44
<b>Quadro 6:</b> Questões para avaliação dos artigos .....	48
<b>Quadro 7:</b> Pontuação por faixa de qualidade .....	49
<b>Quadro 8:</b> Pontuação dos Artigos .....	49
<b>Quadro 9:</b> Tratamentos para dados Temporais .....	56
<b>Quadro 10:</b> Tratamentos para Indeterminação Temporal .....	57

## LISTA DE TABELAS

<b>Tabela 1:</b> Artigos por fonte de busca.....	54
--	----

## LISTA DE FIGURAS

<b>Figura 1:</b> Dados Estruturados e Dados Não Estruturados .....	24
<b>Figura 2:</b> Gastos com Big Data em 2015 .....	26
<b>Figura 3:</b> Origem dos dados em Saúde .....	31
<b>Figura 4:</b> Indicadores de recursos em saúde.....	34
<b>Figura 5:</b> Composição de gastos públicos, planos de saúde e gastos particulares .....	35
<b>Figura 6:</b> Tecnologias utilizadas por empresas de saúde .....	36
<b>Figura 7:</b> Artigos por fase.....	53
<b>Figura 8:</b> Qualidade dos artigos .....	54
<b>Figura 9:</b> Total de artigos por ano .....	55
<b>Figura 10:</b> Artigos Dado Temporal .....	58
<b>Figura 11:</b> Artigos Indeterminação Temporal.....	59
<b>Figura 12:</b> Países de Origem das Pesquisas .....	59
<b>Figura 13:</b> Formulário Padrão .....	60

## LISTA DE SIGLAS

ACID	<i>Atomicity, Consistency, Isolation And Durability</i>
ANS	Agência Nacional de Saúde
BDT	Banco de Dados Temporais
BI	<i>Business Intelligence</i>
CPR	<i>Computer Based Patient Record</i>
DBMS	<i>Database Management Systems</i>
DM	<i>Data Mining</i>
EHR	<i>Eletronic Health Record</i>
EMR	<i>Eletronic Medical Record</i>
EPR	<i>Eletronical Patient Record</i>
IEEE	<i>Institute of Eletrical and Eletronics Engineers</i>
OMS	Organização Mundial da Saúde
OWL	<i>Ontology Web Language</i>
PEPs	Prontuários Eletrônicos dos Pacientes
PHR	<i>Personal Health Record</i>
PLN	Processamento de Linguagem Natural
SGBD	Sistema Gerenciador de Banco de Dados
SQL	<i>Structured Query Language</i>
TI	Tecnologia da Informação
TSQL	<i>Temporal Structured Query Language</i>

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>14</b>
<b>1.1. Objetivo.....</b>	<b>18</b>
1.1.1. Principal.....	18
1.1.2. Específicos.....	18
<b>1.2. Justificativa.....</b>	<b>18</b>
<b>1.3. Método de Pesquisa .....</b>	<b>20</b>
<b>1.4. Organização dos Capítulos.....</b>	<b>21</b>
<b>2. BANCO DE DADOS.....</b>	<b>23</b>
2.1. Big Data.....	25
2.1.1. Big Data em saúde.....	27
2.2. Banco de dados temporais.....	32
<b>3. O SETOR DE SAÚDE.....</b>	<b>34</b>
<b>4. REVISÃO SISTEMÁTICA.....</b>	<b>40</b>
4.1 Etapas da revisão sistemática.....	40
4.2 Visão geral dos estudos.....	44
4.3 Avaliação da qualidade dos artigos.....	47
4.4 Extração dos dados.....	60
<b>5. RESULTADOS E DISCUSSÃO.....</b>	<b>62</b>
<b>6. CONCLUSÕES.....</b>	<b>67</b>
<b>REFERÊNCIAS.....</b>	<b>70</b>

## 1. INTRODUÇÃO

O constante desenvolvimento da área de Tecnologia da Informação (TI) e os avanços em pesquisas para solucionar problemas em dispositivos de armazenamento de dados com grandes capacidades, permitem que cada vez mais dados sejam armazenados, resultando assim em volumosos bancos de dados armazenados em Sistemas Gerenciadores de Bancos de Dados (SGBD).

Os Sistemas Gerenciadores de Bancos de Dados clínicos, armazenam e gerenciam dados, criando um repositório no qual podem ser consultados e utilizados em diversas finalidades (Li et al., 2012), desde uma simples consulta a um registro de pacientes, até complexos relatórios que podem atuar como base para tomada de decisão.

O conteúdo armazenado representa o estado atual de um registro, como por exemplo, de um paciente a sua temperatura, dados relacionados a pressão arterial, monitoramento cardíaco, entre outros.

A mudança no conteúdo armazenado no Banco de Dados, ou seja, a atualização dos valores de seus atributos, implica na perda da informação anterior (SNODGRASS et al., 2010). Isso acontece porque os Sistemas Gerenciadores de Banco de Dados não conseguem trabalhar com o gerenciamento da manutenção dos estados anteriores dos registros. Quando existe essa necessidade, quem realmente é responsável por esse controle é a aplicação ou o usuário.

O aspecto temporal, porém, está naturalmente associado a objetos do mundo real e nas aplicações que envolvem sistemas computacionais, sendo de extrema importância sua manutenção e atualização para a utilização posterior. Visto que essas aplicações possuem alguns atributos cujos valores são modificados durante sua vida e estadia na base de dados, fazendo com que o objeto em si apresente diferentes estados ao longo de sua existência (SAKAI, 2007).

Desta forma, essas modificações precisam ser armazenadas, formando o histórico do objeto que mais tarde pode ser consultado. Por exemplo, o objeto paciente tem o valor de seu atributo temperatura modificado ao longo de sua estadia no hospital e para o diagnóstico de algumas doenças se faz necessário armazenar estes valores anteriores com a finalidade de analisar a diferença entre as temperaturas e relacioná-la com a manutenção de um medicamento ou não.

Muitas vezes, também, é necessário registrar os valores futuros de alguns objetos, identificando situações que ainda não aconteceram, mas que podem vir a acontecer por motivos pré-determinados no presente. Por exemplo, a prescrição de uma dose de um determinado medicamento durante um determinado período futuro.

Além da capacidade de armazenar dados, um Sistema Gerenciador de Banco de Dados, necessita oferecer recursos que possibilitavam manipular estas informações de maneira prática e eficiente, por isso a linguagem de consulta denominada *Structured Query Language* (SQL) foi proposta para tal finalidade e ao longo do tempo está em contínua evolução, mas restringindo-se a apenas aos bancos de dados relacionais (ASTRAHAN; CHAMBERLIN, 1975).

Também é necessário que o SGBD gerencie suas operações garantindo que as informações nunca atinjam estados inconsistentes. Assim, os dados devem ser submetidos à um conjunto de restrições que garantam a integridade e consistência dos mesmos.

Para registrar e gerenciar essa variação de estados dos atributos ao longo do tempo é necessária a implementação e utilização de uma estrutura capaz de associar uma marca temporal aos dados. Para isso, foi proposta uma estrutura tridimensional onde a variável tempo é incluída na estrutura formada por colunas e linhas, sendo as marcas temporais vinculadas aos atributos possibilitando o registro dos momentos em que os valores destes atributos são válidos no mundo real, ou seja, suas alterações ao longo do tempo (SOUZA, 1998). Fazendo com que as estruturas de dados, a linguagem de manipulação e as restrições de integridade devam ser modificadas ou estendidas de maneira a agregar recursos para o tratamento desses aspectos temporais da informação.

Para o tratamento desses aspectos temporais, os Bancos de Dados Temporais estão sendo pesquisados desde a década de 70, de acordo com Edelweiss (1998), para justamente satisfazer estas necessidades de tratamento dos aspectos temporais e tem tido fortes contribuições com os objetivos de agregar novas modelagem de dados, modelos de banco de dados, também estendendo a linguagem de consulta para manipulação de dados temporais e incorporação das restrições que garantem a consistência da informação (SOUZA, 1998).

De acordo com Snodgrass (2008), a consolidação destas pesquisas resultou em uma nova linguagem de consulta temporal denominada TSQL2, essa linguagem baseia-se em SQL por ser esta linguagem a mais utilizada. As seguintes características foram encontradas na definição do TSQL2 (EDELWEISS, 1998):

- Suporte a novos períodos de tempo, em SQL tínhamos somente *date*, *time*, *timestamp*, *interval*;
- Suporte a múltiplas granularidades, diferentes das existentes em SQL, *year*, *month*, *day*, *hour*, *minute*, *second*.
- Suporte a múltiplas representações, exemplo: quarta semana do mês de abril
- Suporte a múltiplas linguagens, como exemplo: 30 de setembro de 2008;
- Suporte a tempo indeterminado, exemplo: entre 5 e 10 de julho;
- Suporte a tempo Histórico.

De maneira formal, pesquisadores definem Bancos de Dados Temporais como sendo aqueles que armazenam todos os estados de uma aplicação (presente, passado, futuro), registrando e gerenciando sua evolução no tempo (GORALWALLA; OZSU; TANSEL, 1995; OZSOYO-GLU; SNODGRASS et al., 2008; SOUZA, 1998).

Mas, é importante salientar que nem todos os problemas relacionados a questão tempo foram solucionados pela criação do TSQL2. Um dos problemas relatados na literatura é denominado de Indeterminação Temporal, que se refere a informações "não sei quando", ou mais precisamente, "não sei exatamente quando" (JENSEN; SNODGRASS, 2008).

O modificador "temporalmente indeterminado" indica que o objeto modificado tem um tempo associado, mas que o tempo não é conhecido precisamente. O momento em que ocorre um evento, quando um intervalo de tempo começa ou termina, ou mesmo a duração de um período pode ser indeterminada. Por exemplo, o evento de uma aplicação de insulina pode ser "algum tempo na semana passada", o intervalo que um exame pode ficar pronto "sexta a sábado", ou a duração que um pesquisador na área médica se dedica a descobrir uma nova doença "quatro a quinze anos " (JENSEN; SNODGRASS, 2008).

O adjetivo "temporal" permite que tipos paralelos de indeterminação sejam definidos, como a indeterminação espacial. Existe uma diferença sutil entre indeterminado e impreciso. Neste contexto, indeterminado é um termo mais geral do que impreciso, uma vez que a precisão é comumente associada à realização de medições. Normalmente, uma medida precisa é preferida a uma imprecisa. Medições de tempo imprecisas, no entanto, são apenas uma fonte de informações temporariamente indeterminadas (EDELWEISS, 1998).

O gerenciamento e tratamento da questão temporal dos dados também se torna crítica quando trata-se de grandes dados, ou *big data*, termo mais recente, dentre as



metodologias de análise e manipulação de grandes massas de dados existentes e refere-se a arquiteturas de sistemas capazes de lidar com as novas dimensões dos dados (velocidade, variedade, volume e valor) e ganhou bastante importância quando as seguintes tecnologias começaram a ser desenvolvidas (MORGENTHAL, 2012).

Para se gerenciar esses dados, o SGBD tem que ser capaz de controlar o grande volume de dados, bem como sua variedade, velocidade e escalabilidade mantendo a conformidade *ACID* (*atomicity, consistency, isolation and durability*) (SAKR; BATISTA; ALOMARI, 2011). Com essa necessidade, uma nova classe de bancos de dados conhecida como *NoSQL – Not Only SQL*, surgiram e são denominados não relacionais, com foco crescente na escalabilidade e no gerenciamento e armazenamento desses dados (POKORNY, 2011).

E nesse contexto, podemos citar a área da saúde, na qual destaca-se pela geração de muitos dados, impulsionados por requisitos obrigatórios e o potencial para melhorar a qualidade na prestação de serviços, enquanto, reduzindo os custos, esse grande volume de dados (conhecidos como "*Big Data*") prometem apoiar uma vasta gama de funções, tais como apoio à decisão, vigilância de doenças e gestão da saúde da população. (FELDMAN; MARTIN; SKOTNES, 2012).

Percebemos que além do tempo ser primordial no diagnóstico, ele também é um fator crucial na evolução de uma doença, pois os vários eventos relacionados a ela se desencadeiam e evoluem com o tempo, além das várias etapas de tratamento que devem ser realizados de forma precisa, com uma determinada ordem e frequência e por um certo período de tempo para ser efetiva.

Visto a importância do estudo sobre dados temporais e da indeterminação temporal em dados clínicos, as questões de pesquisa que orientam este trabalho são:

Quais são os tipos de tratamento realizados em dados temporais e na indeterminação temporal em textos da área da saúde?

Quais são os métodos computacionais existentes que tratam da questão tempo e a falta dele em narrativas clínicas e textos de autorias de pacientes?

Quais são os métodos computacionais existentes para o tratamento da questão temporal?

Quais são os métodos computacionais existentes para encontrar e tratar a indeterminação temporal?

E para responder as questões de pesquisas, os objetivos deste trabalho são:

## **1.1. Objetivo**

### **1.1.1 Principal**

O objetivo deste trabalho é analisar e avaliar por meio de uma revisão sistemática os métodos computacionais para o tratamento de dados temporais e da indeterminação temporal em narrativas clínicas ou textos de autoria de pacientes.

### **1.1.2 Específicos**

- ✓ Identificar as características e os tipos de tratamento que são feitos em dados temporais e na indeterminação temporal em textos da área da saúde;
- ✓ Selecionar e coletar de forma sistemática (reprodutível) os métodos computacionais que tratem sobre a questão de tempo e a falta dele em narrativas clínicas e textos de autoria de pacientes;
- ✓ Identificar e estudar as formas de avaliação de textos científicos para aplicá-lo na análise dos textos selecionados;
- ✓ Analisar os textos encontrados de acordo com as características e tipos de tratamento identificados;
- ✓ Avaliar os métodos encontrados numa tentativa de classificá-los de acordo com a abordagem adotada para realizar o tratamento do dado temporal ou da indeterminação temporal.

## **1.2 Justificativa**

De acordo com Meeker (2016), um componente tecnológico que será muito importante no futuro da Saúde são os dados, contidos nos registros eletrônicos, em narrativas clínicas, em exames de diagnóstico, dentre outras fontes, pois eles possuem informações valiosas para a área médica, e podem ajudar no auxílio e combate a doenças e também nas pesquisas. Nos próximos anos, ele irá ajudar a transformar a Saúde fazendo com que os 90%

dos dados não-estruturados hoje existentes nesse setor e que não são utilizados, sejam transformados em informação estratégica, por meio de processos cognitivos ou seja, de associações, correlações e interpretações por meio de algoritmos e técnicas de extração de conhecimento.

Essa também é a visão estratégica de Rometty, CEO da IBM, ela afirma que a informação contida nesses dados será muito importante na eficiência operacional das organizações, incluindo as empresas de saúde (ROMETTY, 2015).

E para que esses dados sejam analisados por completo, se faz necessário que o mesmo tenha um tratamento adequado e que seja um dado conhecido, principalmente quando se trata de dado temporal. É observado que o tempo exerce grande influência e importância na área médica e saber o exato tempo e ocorrência se faz necessário para um melhor e rápido diagnóstico, mas é sabido também que por muitas vezes esse dado é incerto (LI; PATRICK, 2012).

Em indeterminação temporal, sabe-se que um evento armazenado em um banco de dados temporal realmente ocorreu, mas não se sabe exatamente quando o evento ocorreu. Muitas vezes, um usuário sabe apenas aproximadamente quando um evento aconteceu, por exemplo, ela pode saber que aconteceu "entre as 2 da manhã e as 4 da tarde", "algum dia na primeira semana de junho" ou "em torno de meados de agosto", estes são alguns exemplos de indeterminação temporal (SNODGRASS, 2012).

De acordo com Li et al. (2012), o tempo é um atributo essencial na área médica, durante um diagnóstico, conhecer a ordem temporal de alguns fatos, ocorrências e sintomas podem fazer uma diferença substancial no processo de atendimento ao paciente. Saber quais e por quanto tempo os sintomas duraram permite com que o diagnóstico seja mais preciso e que muitas doenças sejam descartadas, sendo a temporalidade dos dados algo de extrema importância.

Além do tempo ser primordial no diagnóstico, ele também é um fator crucial na evolução de uma doença, pois os vários eventos relacionados a ela se desencadeiam e evoluem com o tempo, além das várias etapas de tratamento que devem ser aplicados de forma precisa, com uma determinada ordem e frequência e por um certo período de tempo para ser efetiva (SUSAN et al., 2013).

Por isso identificar os problemas envolvendo a temporalidade dos dados e a indeterminação temporal em dados estruturados e não estruturados é de extrema importância,

visto a relevância do atributo tempo nos registros clínicos e nas narrativas, saber quais são os tipos de tratamentos existentes para esses problemas e descrevê-los, agrupando-os em tratamentos aos problemas de ordem temporal e na indeterminação do tempo dos dados temporais, de forma a contribuir à comunidade médica.

### **1.3 Método de Pesquisa**

De acordo com o objetivo e a finalidade deste projeto, a pesquisa adotou a abordagem quali-quantitativa. A forma de raciocínio lógico utilizada foi o método indutivo. Em relação à natureza das fontes utilizadas a pesquisa se classifica como bibliográfica e experimental.

Para identificar o estado da arte em soluções para problemas de ordem temporal ou a indeterminação temporal foi realizada a pesquisa bibliográfica utilizando-se do método de revisão sistemática, que de acordo com Sampaio e Mancini (2007), trata-se de um método formal para a síntese de informações disponíveis em estudos primários que sejam relevantes para o conjunto de questões de pesquisa.

Segundo Kitchenhan et al. (2008), a revisão sistemática caracteriza-se pela adoção de uma metodologia para a identificação, análise e interpretação de todas as evidências disponíveis durante a realização da pesquisa, de modo que a mesma seja livre de viés e possa ser repetida em caso de necessidade.

Sampaio e Mancini (2007) apresentam três etapas para a revisão sistemática: planejamento, execução e relatório. A etapa de planejamento consiste basicamente na definição das questões de pesquisa e elaboração do protocolo de revisão. O protocolo de revisão define exatamente como cada etapa da pesquisa é executada de modo que seja possível reproduzi-la.

Diferentes autores apresentam várias formas e variações no que se diz respeito aos estágios que devem ser seguidos para elaborar uma revisão sistemática. Para os propósitos dessa dissertação, replicamos o modelo desenvolvido por Cooper (2010). O quadro 1 apresenta as etapas do estudo.

**Quadro 1:** Planejamento da Revisão Sistemática

<b>Estágio</b>	<b>Descrição</b>
<b>1</b>	Identificação/formulação do problema de pesquisa
<b>2</b>	Pesquisa da literatura (artigos científicos)
<b>3</b>	Seleção dos estudos que respondiam às questões de pesquisa
<b>4</b>	Avaliação da qualidade dos estudos
<b>5</b>	Análise e síntese dos resultados dos estudos
<b>6</b>	Interpretação dos dados coletados
<b>7</b>	Apresentação dos resultados da pesquisa

**Fonte:** Cooper (2010)

Após a conclusão dos 7 estágios, a análise quantitativa dos resultados encontrados com a revisão sistemática possibilitou uma visão geral dos estudos existentes dos problemas temporais e a indeterminação temporal em textos da área médica e nas narrativas clínicas e de pacientes, bem como os métodos existentes quando trata-se dessa temática.

#### **1.4 Organização dos Capítulos**

O Capítulo 1 da Introdução, apresenta a questão de pesquisa, os objetivos, a justificativa e o método utilizado para a construção desse trabalho.

No capítulo 2 encontra-se a fundamentação teórica necessária para o maior entendimento do tema, Fundamentos sobre Banco de Dados, Modelos Relacionais, Modelos Não relacionais, dados estruturados e não estruturados, bem como a questão temporal dos dados são alguns tópicos abordados nesse capítulo.

No capítulo 3, são apresentados conceitos relacionados a área da saúde, bem como os problemas encontrados na literatura sobre dados temporais e da indeterminação temporal em narrativas e textos clínicos e de autorias de pacientes.

No capítulo 4, são apresentadas as etapas da revisão sistemática aplicada aos textos selecionados.

No capítulo 5, são apresentados os resultados obtidos com a revisão sistemática, descrevendo todos os tratamentos e métodos que foram encontrados na literatura no tratamento de dados temporais e da indeterminação temporal em textos e narrativas clínicas e de autorias de pacientes.

No capítulo 6, encontra-se as conclusões desse trabalho, que reúne todas as evidências e respostas para as questões de pesquisa que orientaram esse trabalho. Também são apresentados alguns possíveis trabalhos futuros.

No capítulo a seguir, temos a fundamentação teórica utilizada como base da construção dessa dissertação.

## 2. BANCO DE DADOS

O banco de dados, por si só pode ser considerado como um repositório ou recipiente para uma coleção de arquivos de dados computadorizados. Já um sistema de banco de dados é apenas um sistema computadorizado de manutenção de registros (dados) que por sua vez podem ser inseridos, selecionados, excluídos ou até mesmo alterados pelos seus usuários. Esses dados são gerenciados através de um SGBD ou Sistema de Gerenciador de Banco de Dados.

Os sistemas de bancos de dados são integrados e podem ser compartilhados, seus dados são representados e organizados em tabelas possuindo assim linhas e colunas que se relacionam, a este modelo é dado o nome de modelo relacional que através de uma linguagem denominada *SQL – Structure Query Language* (Linguagem de consulta estruturada), é possível a realização de inserções, atualizações, consultas nos dados do Sistema de Banco de Dados (DATE, 2000).

Com a crescente demanda do volume de dados, surge um novo modelo de SGBD, chamado *NoSQL*, ou seja, não apenas SQL (Linguagem de Consulta Estruturada). Este novo modelo foi proposto com o objetivo de atender ao gerenciamento de grandes volumes de dados semiestruturados ou não estruturados, que necessitam de alta disponibilidade e escalabilidade (LÓSCIO et al., 2011).

A necessidade de um novo conceito em banco de dados surgiu como consequência da ineficiência de SGBDs relacionais em trabalharem com o atual volume de informações de diversas aplicações. Os SGBDs relacionais foram propostos na década de 70, quando as aplicações trabalhavam apenas com dados estruturados, ou seja, dados que possuem uma estrutura fixa e bem definida. Além disso, o volume de dados gerado por tais aplicações é muito menor do que o volume gerado pelas aplicações atuais (LEAVITT, 2010).

Esses dados gerados podem ser classificados em dados estruturados, não estruturados e semiestruturados, onde de acordo com ELMASRI e NAVATHE (1989), dados estruturados são dados organizados em blocos semânticos possuindo assim relações entre eles, normalmente pertencem a um mesmo grupo, possuindo assim os mesmos atributos e que são armazenados em formatos pré-existentes e mantidos em um SGBD – Sistema Gerenciados de Bancos de Dados com uma estrutura e esquema previamente definidos.

Como mantém uma organização, são mais fáceis de serem manipulados, por exemplo, cada registro (linha) em uma tabela de banco de dados relacional, já os dados não

estruturados são quaisquer documentos, arquivos, gráficos, imagens, textos, relatórios, formulários ou gravações de vídeo ou áudio que não tenha sido codificado, ou de outra forma estruturados em linhas e colunas ou registros (SHUXIA et. al, 2009).

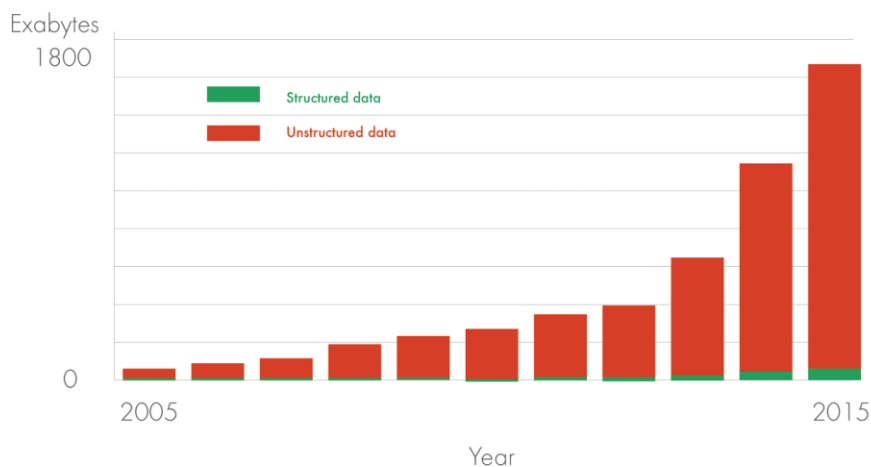
Os dados semiestruturados apresentam uma representação estrutural heterogênea, não sendo nem completamente não estruturados nem estritamente estruturados. Afirmar-se, que dados semiestruturados são dados nos quais o esquema de representação está presente (de forma explícita ou implícita) juntamente com o dado, ou seja, o mesmo é auto descritivo. Isto significa que uma análise do dado deve ser feita para que a sua estrutura possa ser identificada e extraída (SOUZA, 1998).

De acordo com pesquisas realizadas pela IBM (2016), cerca de 80% de todos os dados armazenados são mantidos fora de bancos de dados relacionais. Dados não estruturados se apresentam como informações armazenadas em contexto e são gerenciados por bancos de dados não relacionais, não utilizando a linguagem de consulta estruturada *SQL* e sim o *NoSQL*. A figura 1, mostra o crescimento dos dados não estruturados a partir de 2005:

**Figura 1:** Dados Estruturados e Dados Não Estruturados

## The Rise of Unstructured Data

Merrill Lynch: Unstructured data accounts for 80%+ of all data in organizations and 95% of data generated daily online.



**Fonte:** <https://www.ml.com/articles/health-care-and-the-data-revolution.html#financial-research-and-insights>



Com o crescimento dos dados não estruturados e semi-estruturados, o conceito de *Big Data*, passa a ser utilizado, conforme descreve o item 2.2.

## 2.2 BIG DATA

O termo *Big Data* é mais recente, dentre as metodologias de análise e manipulação de grandes massas de dados existentes e refere-se a arquiteturas de sistemas capazes de lidar com as novas dimensões dos dados: velocidade, variedade, volume, valor e ganhou bastante importância quando as seguintes tecnologias começaram a ser desenvolvidas (MORGENTHAL, 2012):

- Baixo custo dos dispositivos de armazenamento que até então obrigavam as empresas a armazenar apenas a informação essencial.
- Virtualização que aumentou a possibilidade de virtualizar ou emular, por meio de máquinas virtuais quase todo o tipo de hardware necessário.
- *Cloud computing* que proporcionou às empresas uma solução aos problemas de armazenamento, à medida das suas necessidades.
- Bases de dados NoSQL que são desenhadas para trabalharem com grandes volumes de informação em tempo útil e com suporte para processamento distribuído.

Dentro de bases de dados volumosas encontramos muito mais do que “coisas” armazenadas, e é inviável a análise desses dados sem o auxílio de alguma ferramenta computacional apropriada. De acordo com Goldschmidt e Passos (2005):

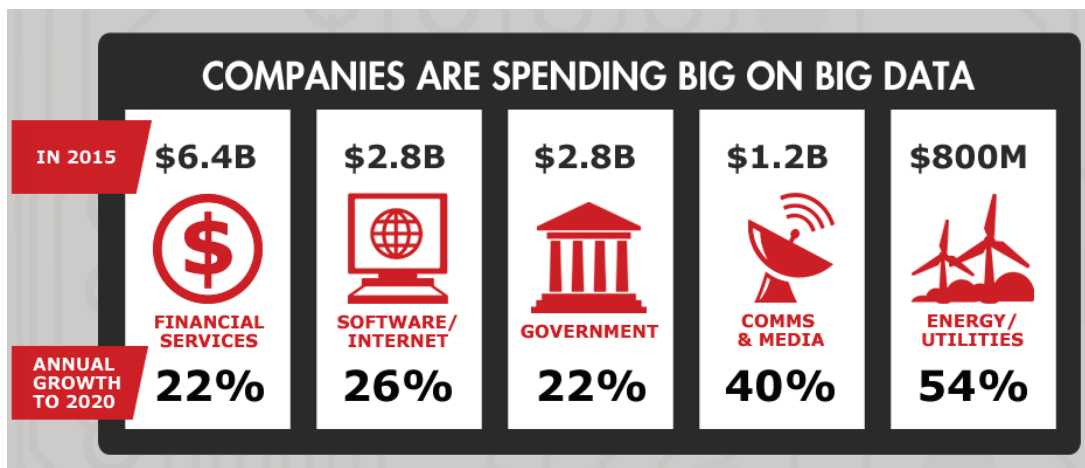
É imprescindível o desenvolvimento de ferramentas que automatizem a tarefa de analisar, interpretar e relacionar esses dados para que possamos desenvolver e selecionar estratégias de ação que tragam benefícios e agreguem valor às instituições (GOLDSCHMIDT; PASSOS, 2005).

E para facilitar a leitura, interpretação e manipulação desses dados, é possível identificar várias técnicas que podem ser aplicadas para se obter melhor compreensão dos mesmos, dentre elas ferramentas de *Business Intelligence*, Mineração de Dados (*data mining*), estatísticas, algoritmos de similaridades e Visualização de Informação e *Big Data*.

McAFEE (2012) afirma ainda que soluções de *Big Data* possuem um potencial maior do que as soluções analíticas tradicionais existentes para trazer benefícios e aumentar a

competitividade frente aos concorrentes. Grandes empresas e até governos têm feito o uso de *Big Data* para auxiliar consumidores e até mesmo cidades inteiras a planejar suas atividades para alcançar um futuro melhor (MANYIKA et al., 2011). A figura 2 mostra o quanto as empresas estão investindo em *Big Data*.

**Figura 2:** Gastos com *Big Data* em 2015



**Fonte:** [http://www.bain.com/publications/articles/big\\_data\\_the\\_organizational\\_challenge.aspx](http://www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx). Acesso em: 30/05/2016.

Davenport (2013) conceitua que os dados de um *Big Data*, podem ser provenientes dos mais diversos locais tais como, voz de um *call center* até dados de genoma de pesquisa biológica e medicinal). Nota-se que algumas das maiores fontes de dados existentes, são relativamente novas tais como as redes sociais, dentre as quais se destacam o *Facebook* lançado em 2004, *Twitter* em 2006 e o *Youtube*, que aliadas a grande massificação de vendas de *smartphones* e *tablets*, redes wireless e redes de dados 3G e 4G, permitem às pessoas permanecerem conectadas em praticamente todo o tempo, gerando conteúdo, ideias e opiniões.

De acordo com Pearson e Wegener (2013), outros diferentes setores de negócio vêm utilizando as ferramentas de *Big Data* para garantir avanços e melhorias aos negócios, tais setores são Bancos, Ensino, Governo, Saúde e Varejo.

Não é surpresa que a Gartner Research identifique levar “a informação certa para a pessoa certa no momento certo” como uma tendência tecnológica estratégica. A Accenture

identificou este ponto como uma prioridade máxima para CIOs (RANSBOTHAM; KIRON; PRENTICE, 2016).

### 2.2.1 BIG DATA EM SAÚDE

Segundo Raghupathi (2010), o setor de saúde historicamente tem gerado grandes quantidades de dados, pela necessidade da manutenção de registros, conformidade e requisitos regulamentares, e atendimento ao paciente. Enquanto a maioria dos dados é armazenada em forma de cópia impressa, a tendência atual é para a digitalização rápida desse montante de dados.

Lobo (2017) afirma que o *Big data* está sendo gradualmente introduzido no sistema de atenção à saúde. Dados de prevalência, incidência e evolução de enfermidades permitiriam gerar dados estatísticos, antecipar surtos epidemiológicos e prescrever ações preventivas e que dados de pacientes, como idade, sexo, etnia, local de residência, antecedentes pessoais e familiares, sintomas e sinais apresentados, exames realizados ou obtidos por meios eletrônicos (*wearable devices*), diagnósticos feitos, tratamento e evolução coletados, permitiriam estabelecer uma base de dados e aprimorar condutas estabelecidas.

Impulsionados por requisitos obrigatórios e o potencial para melhorar a qualidade da prestação de cuidados de saúde, enquanto, reduzindo os custos, esse grande volume de dados promete apoiar uma vasta gama de funções médicas e de saúde, Apoio à decisão, vigilância de doenças e gestão da saúde da população. (FELDMAN; MARTIN; SKOTNES, 2012).

De acordo com pesquisas realizadas pela IBM (2016), os dados do sistema de saúde dos EUA atingiram, em 2011, 150 *exabytes*. A essa taxa de crescimento, *Big Data* na área da saúde nos EUA em breve atingirá a escala *zettabyte* (1021 gigabytes) e, não muito tempo depois, o *yottabyte* (1024 gigabytes). Kaiser Permanente, uma operadora de saúde americana com base na Califórnia, que possui mais de 9 milhões de membros, acredita-se ter entre 26,5 e 44 *petabytes* de dados de *EHRs* (*Eletronic Health Record*), incluindo imagens e anotações.

Frost e Sullivan (2010) definem *Big Data* em saúde como a um conjunto de dados eletrônicos tão grandes e complexos que são difíceis de gerenciar com software e / ou hardware tradicionais; nem podem ser facilmente geridos com ferramentas tradicionais ou comuns de gestão de dados.

Os volumes de dados gerados na área da saúde são grandes não só devido ao seu volume, mas também devido à diversidade de tipos de dados e à velocidade com que devem ser geridos.

A totalidade dos dados relacionados à saúde e bem-estar dos pacientes constituem o *Big Data* no setor de saúde. Inclui dados clínicos e sistemas de apoio à decisão clínica (notas e prescrições escritas do médico, imagens médicas, laboratório, farmácia, seguros e outros dados administrativos); Dados do paciente em registros eletrônicos de pacientes (EPRs); Segundo Bian et al. (2012) tem-se dados gerados por equipamentos, tais como dados de sensor, monitoramento de sinais vitais; Posts em mídias sociais, incluindo feeds do *Twitter* (*tweets*), *facebook*, dentre outros.

Nestes dados, pode-se encontrar informações valiosas para tomada de decisão no setor de saúde. Ao descobrir associações e padrões de compreensão e tendências nos dados, a análise em um *Big Data* conforme relata Smith (2012) tem o potencial de melhorar os cuidados, salvar vidas e reduzir os custos. Assim, grandes aplicações de análise de dados em saúde se aproveitam da explosão de dados para extrair insights para tomar decisões.

Quando *Big Data* são sintetizados e analisados extraíndo assim padrões e tendências, os envolvidos no setor de saúde bem como as operadoras de saúde, podem desenvolver diagnósticos e tratamentos mais aprofundados e perspicazes, resultando assim em melhores custos e melhorias no sistema como um todo (RAGHUPATHI, 2013).

O potencial para a análise do *Big Data* em saúde pode revelar melhores resultados em muitos cenários, por exemplo: analisando as características do paciente, o custo e os resultados de cuidados para identificar os tratamentos mais clinicamente eficazes.

Pode-se também realizar análise de perfis de pacientes (por exemplo, segmentação e modelagem preditiva) para identificar quais pessoas se beneficiariam com cuidados preventivos ou mudanças no estilo de vida.

Além do perfil dos pacientes, também é importante traçar o perfil de doenças existentes em larga escala para identificar eventos preditivos e apoiar iniciativas de prevenção e combate às mesmas.

Através de sistemas analíticos também seriam possíveis a identificação e a prevenção de possíveis fraudes, verificando assim a exatidão e coerência das alegações por parte dos pacientes ou familiares.

Segundo Dembosky (2012), as organizações de saúde que vão desde consultórios médicos até grandes redes de hospitais podem obter benefícios significativos, ao utilizarem o *Big Data*. Os benefícios potenciais incluem a detecção de doenças em estágios anteriores, quando eles podem ser tratados com mais facilidade e eficácia.

Numerosas questões podem ser abordadas analisando o *Big Data*, como exemplo, certos desenvolvimentos ou resultados podem ser previstos e / ou estimados com base em dados históricos, tais como duração da estadia.

Ao analisar dados históricos dos pacientes, é possível também identificar os que provavelmente escolherão a cirurgia eletiva, os que não se beneficiarão da cirurgia.

Além de prever os prováveis pacientes que possuem risco de complicações médicas, tais como sepse (infecção generalizada) ou outra doença adquirida no ambiente hospitalar (FROST; SULLIVAN, 2010). McKinsey (2014) estima que a análise em *Big Data* pode permitir mais de US\$ 300 bilhões de dólares em economia por ano nos EUA, dois terços disso por meio de reduções de aproximadamente 8% nas despesas nacionais com a saúde. McKinsey (2014) acredita que os grandes dados poderiam ajudar a reduzir o desperdício e a ineficiência nas três áreas a seguir:

a) **Operações clínicas:** Pesquisa comparativa de eficácia para determinar formas mais clinicamente relevantes e rentáveis de diagnosticar e tratar pacientes.

b) **Pesquisa e desenvolvimento:** modelagem preditiva para diminuir o desgaste e produzir um *pipeline* de P&D mais enxuto, mais rápido e mais direcionado em medicamentos e dispositivos; ferramentas estatísticas e algoritmos para melhorar a concepção de ensaios clínicos e o recrutamento de pacientes para melhor combinar os tratamentos aos pacientes individuais, reduzindo assim as falhas dos ensaios e acelerando novos tratamentos no mercado; Analisar ensaios clínicos e registros de pacientes para identificar indicações de seguimento e descobrir efeitos adversos antes que os produtos cheguem ao mercado.

c) **Saúde pública:** analisar os padrões de doenças e acompanhar os surtos de doenças e a transmissão para melhorar a vigilância da saúde pública e acelerar a resposta; desenvolvimento mais rápido de vacinas mais especificamente direcionadas; E transformar grandes quantidades de dados em informações que possam ser usadas para identificar necessidades, prestar serviços e prever e prevenir crises, especialmente em benefício das populações. Além disso, sugere que a análise em dados de saúde pode contribuir para Medicina baseada em evidências: Combinar e analisar uma variedade de dados estruturados e não-

estruturados, dados financeiros e operacionais, dados clínicos e dados genômicos para combinar os tratamentos com os resultados, além de prever pacientes em risco de alguma possível doença ou reincidência e fornecer cuidados mais eficientes.

O volume de dados em saúde existentes é extremamente grande e inclui registros médicos pessoais, imagens de radiologia, dados de ensaios clínicos FDA submissões, genética humana e dados populacionais sequências genômicas, etc. Mais recentes formas de dados grandes, como imagens 3D, genômica e leituras de sensores biométricos também alimentam esse crescimento exponencial.

Segundo Feldman (2012), os avanços no gerenciamento de dados, particularmente na virtualização e na computação em nuvem, estão facilitando o desenvolvimento de plataformas para captura, armazenamento e manipulação mais eficientes de grandes volumes de dados. Os dados são acumulados em tempo real com rapidez. O fluxo constante de novos dados que se acumulam a taxas sem precedentes apresenta novos desafios. Assim como o volume e a variedade de dados que são coletados e armazenados mudou, também tem a velocidade em que é gerado e que é necessário para recuperar, analisar, comparar e tomar decisões.

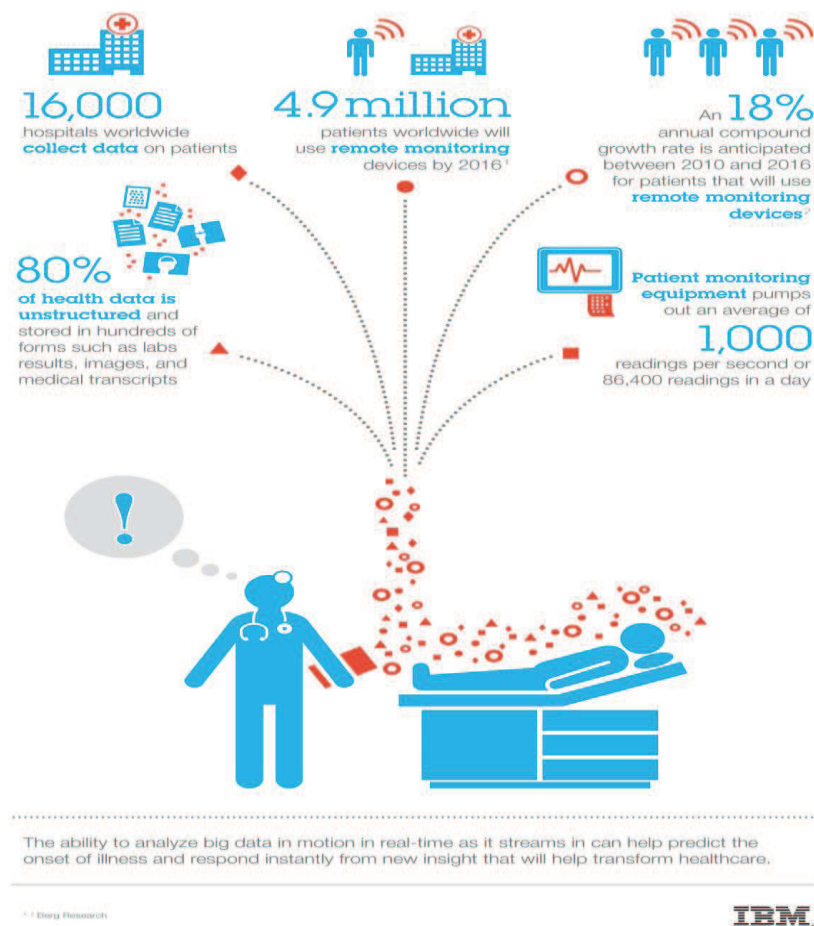
A maioria dos dados de saúde advém tradicionalmente de arquivos de papel, filmes de raio-x e scripts. A velocidade aumenta com os dados que representam o monitoramento regular, como várias medições diárias de glicose diabética (ou controle contínuo por bombas de insulina) e leituras de pressão arterial. Enquanto isso, em muitas situações médicas, dados constantes em tempo real (monitoramento de trauma para pressão sanguínea, monitores de sala de cirurgia para anestesia, monitores de coração à beira do leito, etc.) podem significar a diferença entre a vida e a morte (FELDMAM, 2012).

A Figura 3, mostra a quantidade de dados advindos da área da saúde, bem como as diversas fontes de origem desses dados.

**Figura 3: Origem dos dados em Saúde**

## Big Data in Healthcare: Tapping New Insight to Save Lives

Healthcare is challenged by large amounts of data in motion that is diverse, unstructured and growing exponentially. Data constantly streams in through interconnected sensors, monitors and instruments in real-time faster than a physician or nurse can keep up.



**Fonte:** IBM (2016)

De acordo com Zenger (2012) à medida que a natureza dos dados de saúde evoluiu, também evoluíram as técnicas analíticas existentes, agora são utilizadas em análises complexas e sofisticadas necessárias para acomodar volume, velocidade e variedade. Ficou no passado as coletas exclusivamente em registros eletrônicos de saúde e outros formatos estruturados. Cada vez mais, os dados estão em formato multimídia e não estruturados.

Dados estruturados são dados que podem ser facilmente armazenados, consultados, recuperados, analisados e manipulados por máquina. Historicamente, em saúde, os dados



estruturados incluem leituras de instrumentos e dados gerados pela conversão contínua de registros em papel para registros eletrônicos de saúde e médicos (BORKAR, 2012).

Historicamente, os postos de atendimento, clínicas e hospitais geraram dados não estruturados: registros médicos de escritório, notas manuscritas de enfermeira e médico, registros de admissão e alta hospitalar, prescrições em papel, filmes radiográficos, ressonância magnética, tomografia computadorizada e outras imagens. Mas relativamente pouco destes dados podem ser capturados, armazenados e organizados de forma que possam ser manipulados por computadores e analisados para obter informações úteis. Os aplicativos de saúde, em particular, precisam de formas mais eficientes de combinar e converter variedades de dados, incluindo a automatização da conversão de dados estruturados para dados não estruturados (BURGHARD, 2012).

Zenger (2012) afirma que dados estruturados advindos dos EMR – *Electronic Medical Records* (Registros Médicos Eletrônicos) e EHR – *Electronic Health Records* incluem alguns campos de preenchimento que são familiares e são fáceis de tratar, tais como nome do paciente, dados de nascimento, endereço, nome do médico, hospital, códigos de exames se necessário realizá-los, dados da operadora de saúde, que podem ser facilmente manipulados e automatizados por banco de dados, mas por sua vez, muitas vezes esses mesmos dados são manuscritos, dificultando o entendimento e facilitando a indução de erros na prescrição de medicamentos e até mesmo de exames.

O potencial do *Big Data* em saúde reside na combinação de dados tradicionais (estruturados) com novas formas de dados (não estruturados), tanto individualmente como a nível da população (RAGHUPATHI, 2013).

## **2.3 BANCO DE DADOS TEMPORAIS**

Muitas aplicações, nas mais diversas áreas do conhecimento, têm a necessidade de manipular de alguma forma informações históricas, dados relacionados a estados passados da aplicação. Os SGBD convencionais, no entanto, não proporcionam suporte a estas informações (EDELWEISS, 1994).

A necessidade de suprir essa deficiência fez com que nas últimas décadas muitas pesquisas tenham sido realizadas na área de Banco de Dados Temporais, com o objetivo de definir conceitos e estratégias para tratar de informações históricas (WU, 1997).



Banco de Dados Temporais permitem armazenar os vários estados de uma aplicação presente, passado e futuro), registrando sua evolução com o passar do tempo. Para que isto seja possível, informações temporais são associadas aos dados armazenados, identificando quando a informação foi definida ou o tempo de sua validade (ZANIOLO, 1997).

Jensen et. al (2008) afirma que a noção de tempo, como datas, períodos, duração de validade de informações, intervalos temporais, surge em diferentes níveis, tais como na modelagem de dados, na linguagem de recuperação e manipulação, e no nível de implementação do SGBD.

De acordo com Cameron et. al (2012), os bancos de dados tradicionais relacionais tem-se duas dimensões, representado pelas linhas de uma tabela e pelas colunas dela, sendo que cada atributo de uma linha apresenta um só valor, caso seja realizada uma alteração deste valor, o anterior é perdido, por exemplo um atributo que representa a pressão arterial do paciente, neste caso o banco de dados armazena somente o último valor para o atributo pressão arterial.

Já nos modelos temporais tem-se uma nova dimensão, que é acrescida aos modelos tradicionais, a chamada dimensão temporal, onde a cada alteração do valor de um atributo, o valor anterior não se perde, mantendo todos os estados ou estágios do atributo, sendo possível posteriormente analisar a sua evolução temporal CAMERON (2012).

Lohr (2017) afirma que apesar do gerenciamento temporal dos dados ser fundamental a área médica, ainda há falta de um Sistema de Gerenciamento Temporal adequado para dados temporais, alguns problemas no gerenciamento do volume, a variedade e a velocidade dos dados, quando se fala em grandes dados ou *Big Data*.

Com a implementação de registros médicos eletrônicos (EMRs) e as narrativas clínicas, são produzidas enormes quantidades de textos clínicos não estruturados, com vários dados de natureza temporal, com dados temporais por muitas vezes implícitos, desestruturados e sem qualidade na escrita afirma Shah et. al (2017).

No capítulo a seguir, são apresentados conceitos relacionados a área da saúde, bem como os problemas relatados em relação a dados temporais e a indeterminação temporal em narrativas clínicas e textos de autoria de pacientes.

### 3 O SETOR DE SAÚDE

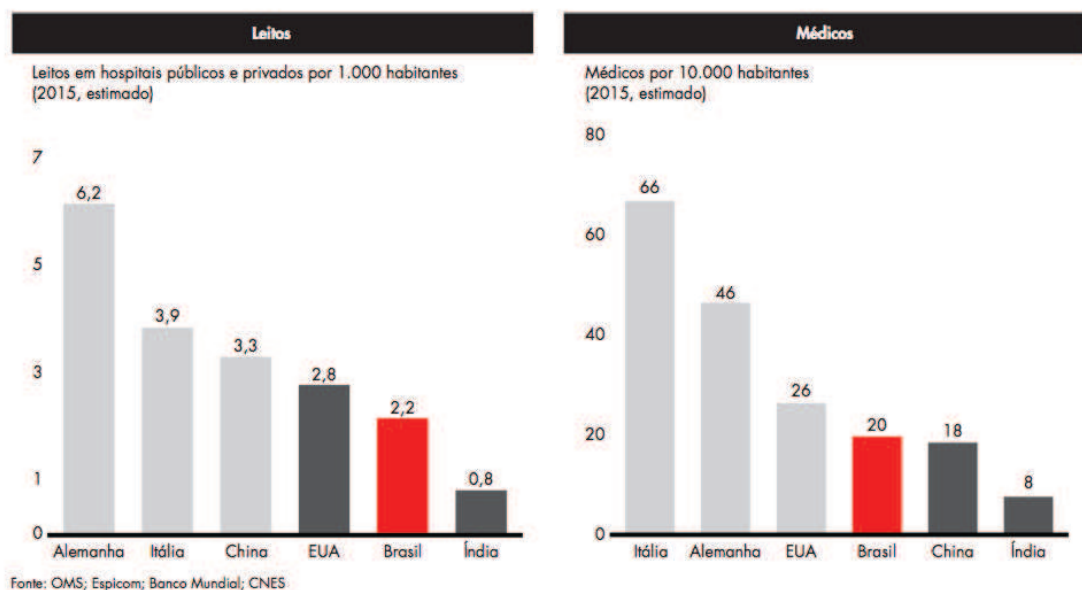
A saúde, juntamente com a educação, a alimentação e a liberdade são direitos fundamentais do ser humano e são essenciais para se ter qualidade de vida. Ela é resultante da combinação entre fatores sociais, econômicos, políticos e culturais, que se apresentam de forma particular em cada sociedade.

Alguns países gastam com o setor de saúde, um percentual significativo de seu PIB (Produto Interno Bruto), chegando a 15%, como ocorreu em alguns países desenvolvidos em 2015, de acordo com a OMS.

Em 2015 no Brasil, aproximadamente 10% do seu PIB, foi destinado ao setor de saúde, porém apesar dos gastos elevados, o país apresenta baixos índices no setor de saúde, sendo comparado à países que estão em desenvolvimento tais como Índia e China, de acordo com relatórios disponíveis no site da Organização Mundial da Saúde (OMS).

De acordo com Fiorentino et. al. (2016), os indicadores de disponibilidade de recursos, tais como o número de médicos e de leitos hospitalares, mostram que a infraestrutura do Brasil está distante da oferecida pelos países desenvolvidos, de acordo com relatórios disponíveis nos sites da Agência Nacional da Saúde (ANS) e da Organização Mundial da Saúde (OMS).

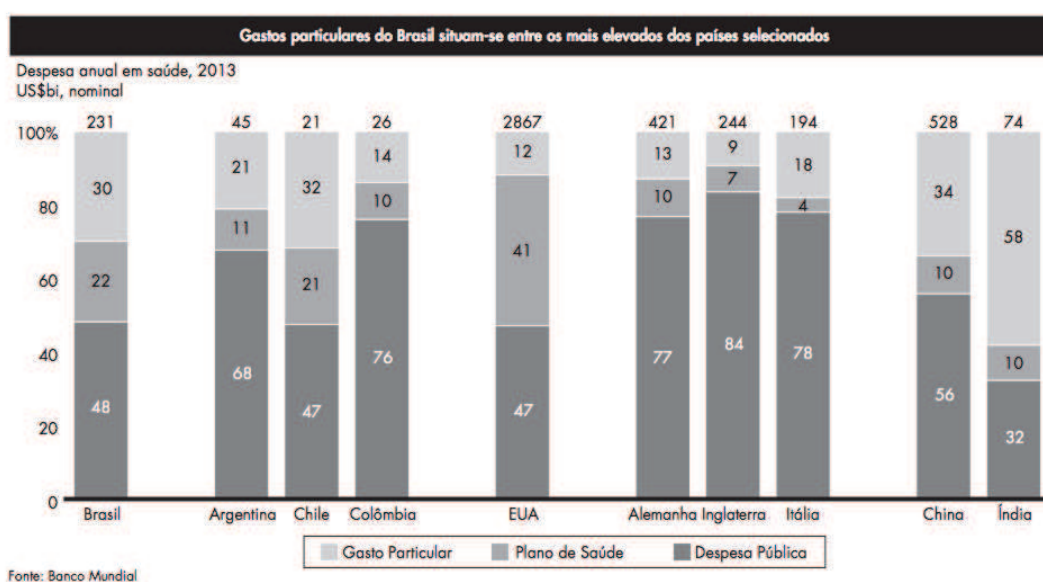
**Figura 4:** Indicadores de recursos em saúde



De acordo com dados do Ministério da Saúde, o Brasil possuía em 2015, 2,2 leitos hospitalares por 1.000 habitantes e 20 médicos por 10 mil habitantes, números abaixo da recomendação da Organização Mundial da Saúde (OMS), que considera um número de 30 médicos por 10 mil habitantes como infraestrutura mínima, sendo que essa realidade ainda varia bastante por estado (FIORENTINO, 2016).

A figura 5, apresenta a composição de gastos públicos, planos de saúde e gastos particulares de alguns países comparando com o Brasil.

**Figura 5:** Composição de gastos públicos, planos de saúde e gastos particulares



Podemos visualizar que os gastos particulares com saúde no Brasil, está entre os mais elevados em comparação com os países selecionados. Também é possível identificar que a despesa pública ainda prevalece no Brasil, devido à origem do sistema de saúde brasileiro, que se desenvolveu a partir da previdência social, entretanto, a saúde suplementar (setor privado) cresceu fortemente nos últimos anos e, hoje, o Brasil apresenta um dos maiores sistemas de saúde suplementar do mundo, de acordo com a ANS.

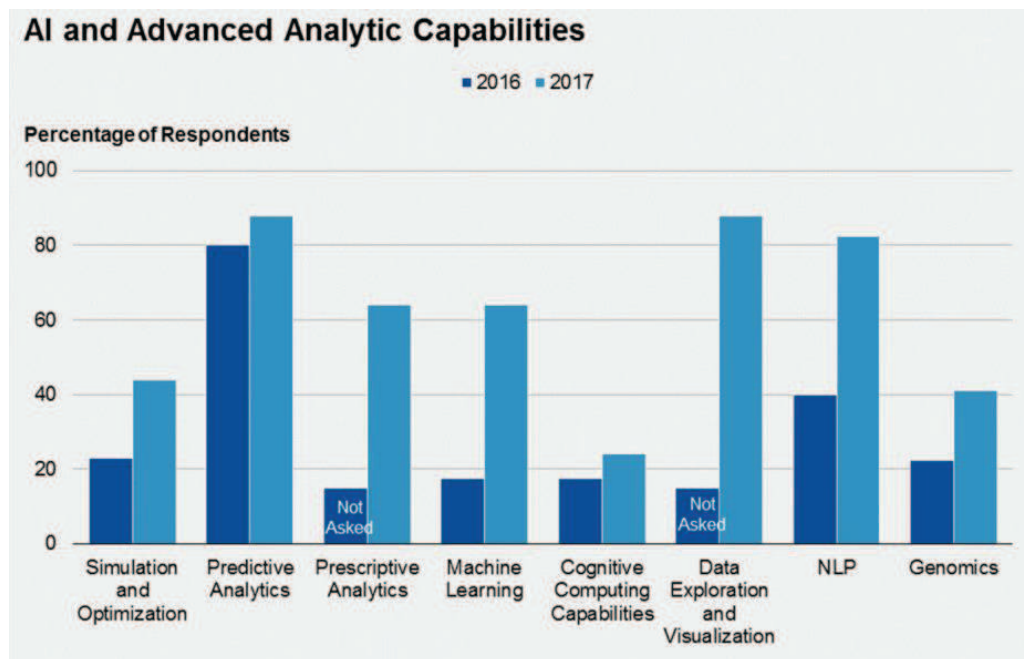
Lobo (2017) afirma que grandes gastos justificam-se pelos altos custos hospitalares em todo o mundo, sendo um dos motivos para esse aumento, a constante adoção de novas tecnologias na área da saúde, houve um grande avanço nas tecnologias médicas utilizadas no diagnóstico pela imagem: o ecocardiograma, o ultrassom, a ressonância, e o PET, que permite

verificar as áreas envolvidas em várias atividades cerebrais, dentre outros exames que possuem um custo altíssimo e fazem uso de equipamentos de ponta (LOBO, 2017).

De acordo com Craft (2018), analista da Gartner, a adoção de novas tecnologias, por parte de empresas de saúde privadas, também se dá pelo fato de que por meio de ferramentas que processam os dados dos clientes, as operadoras de saúde podem fazer previsões de gravidade em determinado grupo de risco, identificar pacientes aparentemente saudáveis, mas que estão prestes a se tornar onerosos ou pacientes de altos riscos. Identificar também oportunidades para redução de custos, melhorando os padrões que elevam o sucesso do tratamento, estabelecendo assim melhores práticas.

A Gartner(2018), realizou uma pesquisa em fevereiro de 2018, onde é possível verificar em quais tecnologias e técnicas para verificação e extração de informação, as empresas de saúde estão investindo. A figura 6, destaca as técnicas utilizadas por empresas de saúde.

**Figura 6:** Tecnologias utilizadas por empresas de saúde



**Fonte:** Gartner (2018)

Como podemos ver na figura 6, as empresas de saúde suplementar estão investindo em tecnologias para análise e extração de informação, é possível identificar um crescimento significativo, considerando os números no ano de 2016 para 2017.

Dentre as técnicas utilizadas, podemos verificar que a exploração de dados e visualização, linguagem de processamento natural e a análise de previsão, destacam-se dentre as mais utilizadas.

Além do processo de incorporação de novas tecnologias, também têm sido foco de discussão no Brasil os altos valores no mercado de materiais, como é o caso das próteses e órteses. Segundo a FenaSaúde(2017), o valor de uma prótese de joelho no Brasil pode oscilar entre R\$2 mil e R\$18 mil, e esta variação pode ser um desafio para os pagadores.

Em conjunto com a adoção de novas tecnologias e pesquisas realizadas para o tratamento das doenças, o setor de saúde tem adotado tecnologias para o processamento, armazenamento e controle de todos os dados que são gerados nos procedimentos em clínicas, hospitais, nos exames laboratoriais, etc.

Grandes quantidades de dados são geradas diariamente, devido a necessidade da manutenção de registros dos pacientes chamado de *EHR – Eletronic Health Records* ou *EMR – Eletronic Medical Records*, PEPs – Prontuários Eletrônicos dos Pacientes, narrativas de pacientes e médicos, que se encontram nos mais diversos formatos, sendo os dados apresentados em formatos desestruturados, o que torna o trabalho de extração e síntese dessas informações um desafio técnico (JIANG, 2015).

Além dos dados gerados pelos registros dos pacientes, tem sido introduzida na prática médica a utilização de *wearable devices*, obtendo informações contínuas sobre glicemia, ECG e movimento, por exemplo, que podem gerar ações automatizadas, como injetar insulina, dar uma descarga elétrica de um desfibrilador subcutâneo ou variar a dose de um medicamento em pacientes com doença de Parkinson. Informações desses *gadgets* são capturadas pelo celular do paciente e podem ser transmitidas ao seu médico, o que gera mais dados para serem processados. (LOBO, 2017).

O processamento dessas informações contidas nesses registros, em notas clínicas e em narrativas de médicos ou pacientes e os que são gerados por dispositivos, como o *wearable devices*, se torna crucial, sendo um aspecto importante a ordem temporal dos eventos clínicos.

O tempo é essencial na pesquisa clínica, pois nesses registros e anotações pode-se encontrar informações importantes relacionados à padrões de doenças, sua progressão, por meio da análise de eventos passados, pode-se explicar possíveis causas de uma situação ou estágio clínico, bem como a previsão de eventos futuros com base no status atual do paciente,

contribuindo para que essas informações ajudem os profissionais da área médica no auxílio ao tratamento de doenças, bem como de pesquisadores da área (JIANG, 2015).

Dembosky (2012), afirma que um objetivo importante para permitir o uso significativo do EHR é o desenvolvimento de aplicações de software que consigam perceber o seu verdadeiro potencial, para melhorar a segurança, a qualidade e a eficiência dos tratamentos médicos, já Lee (2017), afirma que os EHRs devem possibilitar e habilitar a extração e a aplicação do conhecimento.

Para auxiliar e permitir aos pesquisadores clínicos o trabalho com a dimensão temporal na análise de dados médicos, várias pesquisas foram realizadas nesse âmbito. Pesquisadores desenvolveram softwares e aplicações que se utilizam de vários métodos que realizam a síntese e a análise desses textos e narrativas (LI, 2012).

Dentre os métodos e aplicações desenvolvidas podemos citar as técnicas de processamento de linguagem natural (PLN), algoritmos para aprendizagem de máquina, técnicas de mineração de dados clínicos, softwares que realizam a leitura e extração de padrões e sequência de ocorrência dos eventos, dentre outros tipos de tratamentos para os dados temporais (ZHANG, 2016).

No entanto, a natureza dos dados clínicos nos EMR's, EHR's e nas narrativas clínicas tendem a ser heterogêneos, ruidosos e com uma linguagem médica complexa. Assim, o processamento automático desses dados para extrair informações de interesse ainda é um desafio para os métodos e softwares já desenvolvidos.

Wang et al. (2008) observaram que a sublinguagem expressa em narrativas de pacientes tende a ser de natureza descritiva, composta por substantivos e adjetivos pois o assunto principal é constituído por achados clínicos, tratamentos e procedimentos, que são expressos principalmente como substantivos. A sublinguagem de relatórios clínicos é única e, muitas vezes, caracteriza-se por frases incompletas e com abreviaturas.

Sujan et. al (2013) afirma que os dados existentes nos EMR's, EHR's e nos *wearable devices*, por sua vez também tendem a se tornar cada vez mais densos, devido ao envelhecimento da população e ao aumento da ocorrência de doenças, com múltiplos eventos sobrepostos, em diversos registros ou diversas narrativas, sendo necessária à sua combinação para extração de informações relevantes que auxiliem no diagnóstico ou tratamento dos pacientes. Para Sun et al. (2015), o raciocínio temporal é fundamental para determinar eventos e ações específicas ao longo da história do paciente.

O problema atualmente é o processamento de grandes volumes de informações, seja por meio de prontuários eletrônicos com dados dos pacientes, resultados de exames, diagnóstico proposto, prescrição de medicamentos e resultados dessas medicações, seja por digitação, considerando que dados podem não estar disponíveis ou podem estar incompletos. Algumas informações ainda podem ter que ser digitadas em decorrência da incompatibilidade de sistemas e a digitação poderá eventualmente introduzir um componente de erro humano. (SUN, 2015).

Diante da importância dos dados temporais contidos nos registros eletrônicos dos pacientes, bem como nas narrativas clínicas, e observando as dificuldades encontradas para se trabalhar com esses tipos de dados, a seguir detalhamos a revisão sistemática realizada nessa dissertação, para o conhecimento do estado da arte no que se trata de tratamentos e métodos existentes na literatura que tratem da questão temporal dos dados clínicos e também nas narrativas de pacientes ou médicos, visto a sua importância para a área.

## 4 REVISÃO SISTEMÁTICA

Esta seção apresenta as etapas da condução dessa revisão sistemática da literatura, que permitiram responder às questões de pesquisa deste trabalho.

### 4.1 Etapas da revisão sistemática

A fim de orientar o mapeamento da revisão sistemática da literatura desse trabalho, levantou-se as questões de pesquisa que atuam como fator de inclusão e exclusão dos estudos encontrados.

Posteriormente, definiu-se as bases de busca utilizados na pesquisa. Buscou-se incluir as principais fontes de trabalhos nas especialidades de ciência da computação e de saúde. As bases selecionadas são listadas no Quadro 2.

**Quadro 2:** Bases de periódicos utilizados nas revisões sistemáticas em Computação e Medicina

Base	Endereço
ACM Digital Library	<a href="http://dl.acm.org">http://dl.acm.org</a>
IEEE Xplore	<a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a>
PubMed	<a href="http://www.ncbi.nlm.nih.gov/pubmed">www.ncbi.nlm.nih.gov/pubmed</a>
Periódicos da Capes (Google Scholar)	<a href="https://scholar-google-com-br.ez1.periodicos.capes.gov.br/">https://scholar-google-com-br.ez1.periodicos.capes.gov.br/</a>

Fonte: Autora

Após a definição das bases onde são efetuadas as buscas, definiu-se os termos sinônimos para cada item das questões de pesquisa, além de agrupar os termos conforme o Quadro 3. Posteriormente, foram definidos os termos de busca (termos concatenados com operadores booleanos).

É preciso representar os conjuntos de termos encontrados nas questões de pesquisa a fim de recuperar os trabalhos relevantes nas bases pesquisadas. O Grupo 1 descreve sinônimos relacionados à problemas relacionados a dados temporais. O Grupo 2 trata de dados não estruturados em saúde, por fim o Grupo 3 ilustra sinônimos que envolvam textos da área da saúde.



**Quadro 3:** Termos sinônimos e grupos de pesquisa

	<b>Grupo1 - Tempo</b>	<b>Grupo2 – Dados não estruturados em saúde</b>	<b>Grupo3 – textos em saúde</b>
<b>Termo 1</b>	<i>Temporal Indeterminacy</i>	<i>Personal Health Record – PHR unstructured text</i>	Clinical Text
<b>Termo 2</b>	<i>Temporal Data</i>	Electronic Medical Record – EMR unstructured text	Patient Authored Text – PAT
<b>Termo 3</b>	<i>Valid-time</i>	Electronic Healthcare Record – EHR unstructured text	Clinical Narrative
<b>Termo 4</b>	<i>Clinical Temporal Information</i>		Clinical Natural Language Processing
<b>Termo 5</b>			Patient's stories

Fonte: Autora

A fim de responder as questões de pesquisa levantadas, combinou-se os grupos de sinônimos e definiu-se as *Strings* de Busca conforme o Quadro 4. Os Grupos 1 e 2 buscam identificar trabalhos que tratam de problemas relacionados a dados temporais e indeterminação temporal em dados não estruturados em saúde tais como, PHR – *Personal Health Record*, EMR – *Electronic Medical Record* e EHR – *Electronic Healthcare Record*. Por outro lado, os Grupos 2 e 3 buscam identificar trabalhos sobre dados não estruturados, textos clínicos, textos de pacientes, narrativas clínicas e processamento de linguagem natural.

**Quadro 4:** Strings de busca

<b>Identificação</b>	<b>Grupos Combinados</b>	<b>Strings de Busca</b>
1	Grupo1 e Grupo2	<i>(“Temporal Indeterminacy” AND (“Personal Health Record” OR “PHR”) AND “unstructured text”)</i>
2	Grupo1 e Grupo2	<i>(“Temporal Indeterminacy” AND (“Electronic Medical Record” OR “EMR”) AND “unstructured text”)</i>
3	Grupo1 e Grupo2	<i>(“Temporal Indeterminacy” AND (“Electronic Healthcare Record” OR “EHR”) AND “unstructured text”)</i>
4	Grupo1 e Grupo2	<i>(“Temporal Data” AND (“Personal Health Record” OR “PHR”) AND “unstructured text”)</i>

5	Grupo1 e Grupo2	<i>("Temporal Data" AND ("Electronic Medical Record" OR "EMR") AND "unstructured text")</i>
6	Grupo1 e Grupo2	<i>("Temporal Data" AND "Electronic Healthcare Record" OR "EHR") AND "unstructured text")</i>
7	Grupo1 e Grupo2	<i>("Valid-time" AND ("Personal Health Record OR "PHR") AND "unstructured text")</i>
8	Grupo1 e Grupo2	<i>("Valid-time" AND ("Electronic Medical Record OR "EMR") AND "unstructured text")</i>
9	Grupo1 e Grupo2	<i>("Valid-time" AND ("Electronic Healthcare Record" OR "EHR") AND "unstructured text")</i>
10	Grupo1 e Grupo2	<i>("Clinical Temporal Information" AND ("Personal Health Record" OR "PHR") AND "unstructured text")</i>
11	Grupo1 e Grupo2	<i>("Clinical Temporal Information" AND "Electronic Medical Record " OR "EMR") AND "unstructured text")</i>
12	Grupo1 e Grupo2	<i>("Clinical Temporal Information" AND ("Electronic Healthcare Record" OR "EHR") AND "unstructured text")</i>
13	Grupo1 e Grupo2	<i>(("Temporal Indeterminacy" OR "Clinical Temporal Information" OR "Temporal Data" OR "Valid-Time") AND ("Personal Health Record" OR "PHR") AND ("unstructured text") OR "Electronic Medical Record" OR "EMR") AND ("unstructured text") OR ("Electronic Healthcare Record" OR "EHR") AND ("unstructured text"))</i>
1	Grupo1 e Grupo3	<i>(("Temporal Indeterminacy" OR "Clinical Temporal Information" OR "Temporal Data" OR "Valid-Time") AND "Clinical Text" OR "Patient Authored Text" OR "PAT" OR "Clinical Narrative" OR "Clinical Natural Language Processing" OR "Patient's stories"))</i>
2	Grupo1 e Grupo3	<i>"Temporal Indeterminacy" AND "Clinical Text"</i>
3	Grupo1 e Grupo3	<i>"Temporal Indeterminacy" AND "Patient Authored Text" OR "PAT"</i>
4	Grupo1 e Grupo3	<i>"Temporal Indeterminacy" AND "Clinical Narrative "</i>
5	Grupo1 e Grupo3	<i>"Temporal Indeterminacy" AND "Clinical Natural Language Processing"</i>
6	Grupo1 e Grupo3	<i>"Temporal Indeterminacy" AND "Patient's stories"</i>
7	Grupo1 e Grupo3	<i>"Temporal Data" AND "Clinical Text"</i>
8	Grupo1 e Grupo3	<i>"Temporal Data" AND "Patient Authored Text" OR "PAT"</i>
9	Grupo1 e Grupo3	<i>"Temporal Data " AND "Clinical Narrative "</i>

10	Grupo1 e Grupo3	<i>“Temporal Data “ AND “Clinical Natural Language Processing“</i>
11	Grupo1 e Grupo3	<i>“Temporal Data “ AND “Patient’s stories“</i>
12	Grupo1 e Grupo3	<i>“Valid-time” AND “Clinical Text”</i>
13	Grupo1 e Grupo3	<i>“Valid-time” AND “Patient Authored Text” OR “PAT”</i>
14	Grupo1 e Grupo3	<i>“Valid-time“ AND “Clinical Narrative “</i>
15	Grupo1 e Grupo3	<i>“Valid-time“ AND “Clinical Natural Language Processing“</i>
16	Grupo1 e Grupo3	<i>“Valid-time“ AND “Patient’s stories“</i>
17	Grupo1 e Grupo3	<i>“Clinical Temporal Information” AND “Clinical Text”</i>
18	Grupo1 e Grupo3	<i>“Clinical Temporal Information” AND “Patient Authored Text” OR “PAT”</i>
19	Grupo1 e Grupo3	<i>“Clinical Temporal Information “ AND “Clinical Narrative “</i>
20	Grupo1 e Grupo3	<i>“Clinical Temporal Information “ AND “Clinical Natural Language Processing“</i>
21	Grupo1 e Grupo3	<i>“Clinical Temporal Information “ AND “Patient’s stories“</i>

**Fonte:** Autora

Após a utilização das *strings* de busca nas bases de publicações descritas, foram utilizados os seguintes critérios para selecionar os artigos para a revisão sistemática da literatura:

- ✓ Artigos publicados em revistas ou conferências revisadas que respondam uma ou mais questões de pesquisa;
- ✓ Artigos publicados nos últimos 10 anos, de 2007 até 2017;
- ✓ Estudos disponíveis para download nas bases de buscas definidas.
- ✓ Estudos duplicados foram excluídos da amostra
- ✓ Artigos que não atendiam o escopo desse trabalho, ou seja, trabalhos que não respondiam a uma ou todas as perguntas que orientam essa pesquisa.
- ✓ Para a verificar se um artigo atendia ou não ao escopo do trabalho, em um primeiro momento foram lidos os títulos e seus resumos e em um segundo momento, os artigos foram lidos na sua íntegra, a fim de identificar a qualidade dos estudos, para posteriori realizar a extração e síntese dos estudos.

A Revisão Sistemática da literatura foi realizada com o propósito de: reunir conhecimento sobre os problemas relacionados a temporalidade dos dados, bem como a indeterminação temporal em dados, registros clínicos e narrativas de pacientes e médicos por meio de estudos primários a respeito do tema e também métodos e tipos de tratamentos que estão sendo pesquisados e desenvolvidos na academia, quando trata-se de problemas relacionados à temporalidade dos dados clínicos.

Para cumprir os objetivos propostos neste trabalho, pesquisas foram realizadas tendo como principal fonte de referência os sites específicos da comunidade científica na Internet (ieee.org, acm.org, portal de periódicos da Capes, Pubmed, ACM e Google Acadêmico), para a seleção dos estudos, para que fosse possível responder as questões de pesquisa que orientam essa dissertação.

#### 4.2 Visão geral dos estudos

A partir das buscas realizadas nas bases de dados, com os grupos de *Strings* descritos, foram retornados 926 estudos, provenientes da busca automática nos engenhos eletrônicos, a esses estudos foram aplicados critérios de inclusão e exclusão (artigos duplicados, ano de publicação, adequação aos objetivos dessa dissertação, se respondiam às questões de pesquisa desse trabalho).

Do total de 926 artigos, 283 duplicados foram excluídos, restando 643 artigos, sendo que após um filtro para considerar apenas os artigos publicados nos últimos 10 anos, período esse entre os anos de 2007 e 2017, 200 artigos foram excluídos, totalizando 443 artigos para a fase da leitura dos títulos e resumos.

Após a leitura dos títulos e resumos restaram então 33 estudos potencialmente relevantes, sendo descartados nessa fase 413 artigos, que não respondiam as questões de pesquisa desse trabalho.

O quadro 5, mostra os 33 artigos selecionados nessa fase:

#### Quadro 5: Relação dos artigos selecionados

<b>Relação dos 33 artigos selecionados</b>
--

ID	Título	Autores	Fonte
1	<b>Towards Event Sequence Representation, Reasoning and Visualization for EHR Data</b>	Cui Tao and Krist Wongsuphasawat and Kimberly Clark and Catherine Plaisant and Ben Shneiderman and Christopher G. Chute	ACM
2	<b>Challenges in Understanding Clinical Notes: Why NLP Engines Fall Short and Where Background Knowledge Can Help</b>	Sujan Perera and Amit Sheth and Krishnaprasad Thirunarayan and Suhas Nair and Neil Shah	ACM
3	<b>Parsing Clinical Text: How Good Are the State-of-the-Art Parsers?</b>	Min Jiang and Yang Huang and Jun-wei Fan and Buzhou Tang and Josh C. Denny and Hua Xu	ACM
4	<b>Leveraging Natural Language Processing of Clinical Narratives for Phenotype Modeling</b>	Preethi Raghavan and Albert M. Lai	ACM
5	<b>Experiences with Mining Temporal Event Sequences from Electronic Medical Records: Initial Successes and Some Challenges</b>	Debpakash Patnaik and Patrick Butler and Naren Ramakrishnan and Laxmi Parida and Benjamin J. Keller and David A. Hanauer	ACM
6	<b>Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records</b>	Taowei David Wang and Catherine Plaisant and Alexander J. Quinn and Roman Stanchak and Shawn Murphy and Ben Shneiderman	ACM
7	<b>Multidimensional Temporal Mining in Clinical Data</b>	Shusaku Tsumoto and Shoji Hirano	ACM
8	<b>Temporal Phenotyping from Longitudinal Electronic Health Records: A Graph Based Framework</b>	Chuanren Liu and Fei Wang and Jianying Hu and Hui Xiong	ACM
9	<b>V-model: A New Innovative Model to Chronologically Visualize Narrative Clinical Texts</b>	Heekyong Park and Jinwook Choi	ACM
10	<b>A Corpus of Clinical Narratives Annotated with Temporal Information</b>	Lucian Galescu and Nate Blaylock	ACM
11	<b>A semi-supervised approach for temporal information extraction from clinical text</b>	G. Moharasar; T. B. Ho	IEEE
12	<b>Mining the Clinical Narrative: All Text are Not Equal</b>	K. Feldman; N. Hazekamp; N. V. Chawla	IEEE

13	<b>Towards comprehensive longitudinal healthcare data capture</b>	D. Cameron; V. Bhagwan; A. P. Sheth	IEEE
14	<b>Integrated Decision Support by Combining Textual Information Extraction, Facetted Search and Information Visualisation</b>	D. Sonntag; H. J. Profitlich	IEEE
15	<b>Patient information extraction in noisy tele-health texts</b>	M. Y. Kim; Y. Xu; O. Zaiane; R. Goebel	IEEE
16	<b>Extracting Modifiable Risk Factors from Narrative Preventive Healthcare Guidelines for EHR Integration</b>	S. Shah; X. Luo	IEEE
17	<b>MTPGraph: A Data-Driven Approach to Predict Medical Risk Based on Temporal Profile Graph</b>	S. Zhang; L. Liu; H. Li; Z. Xiao; L. Cui	IEEE
18	<b>Towards practical temporal relation extraction from clinical notes: An analysis of direct temporal relations</b>	H. J. Lee; Y. Zhang; J. Xu; C. Tao; H. Xu; M. Jiang	IEEE
19	<b>Task-Oriented Extraction of Temporal Information: The Case of Clinical Narratives</b>	R. Gaizauskas; H. Harkema; M. Hepple; A. Setzer	IEEE
20	<b>Design of an extensive information representation scheme for clinical narratives.</b>	Louise Deléger, Leonardo Campillos, Anne-Laure Ligozat, Aurélie Névél	Pubmed
21	<b>Normalization of relative and incomplete temporal expressions in clinical narratives.</b>	Weiyi Sun , Anna Rumshisky , Ozlem Uzuner	Pubmed
22	<b>CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives.</b>	Cui Tao, Harold R. Solbrig, Christopher G. Chute	Pubmed
23	<b>GUIR at SemEval-2017 Task 12: A Framework for Cross-Domain Clinical Temporal Information Extraction</b>	MacAvaney, Sean; Cohan, Arman; Goharian, Nazli;	Google
24	<b>Extracting temporal information from electronic patient records</b>	Li, Min; Patrick, Jon;	Google

25	<b>Valid-Time Indeterminacy in Temporal Relational Databases: A Family of Data Models</b>	L. Anselma; P. Terenziani; R. T. Snodgrass	IEEE
26	<b>Mining of indeterminacy temporal data based on Fuzziness</b>	R. Shuxia; Z. Zheng	IEEE
27	<b>A representational analysis of a temporal indeterminacy display in clinical events</b>	M. Madkour; Hsing-yi Song; Jingcheng Du; C. Tao	IEEE
28	<b>Query of Indeterminacy Temporal Data Based on CPR</b>	R. Shuxia; Z. Zheng	IEEE
29	<b>Coping with Events in Temporal Relational Databases</b>	P. Terenziani	IEEE
30	<b>A Comprehensive Approach to ‘Now’ in Temporal Relational Databases: Semantics and Representation</b>	L. Anselma; L. Piovesan; A. Sattar; B. Stantic; P. Terenziani	IEEE
31	<b>Nearly Periodic Facts in Temporal Relational Databases</b>	P. Terenziani	IEEE
32	<b>Valid-Time Indeterminacy in Temporal Relational Databases: Semantics and Representations</b>	L. Anselma; P. Terenziani; R. T. Snodgrass	IEEE
33	<b>Irregular Indeterminate Repeated Facts in Temporal Relational Databases</b>	P. Terenziani	IEEE

Fonte: Autora

Nessa fase os 33 estudos descritos foram submetidos a avaliação da qualidade, fase essa descrita a seguir.

### 4.3 Avaliação da qualidade dos artigos

De acordo com Cooper (2010), a avaliação da qualidade dos estudos consiste em examinar o grau de correspondência entre os métodos e técnicas utilizados e as conclusões observadas.



Após aplicar os critérios de inclusão e exclusão, foi realizada a avaliação da qualidade dos estudos resultantes da fase anterior para isso, foi utilizado um questionário adaptado de Dyba e Dingsoyr (2008).

Durante esta fase, com a leitura completa dos artigos, identificamos que alguns deles não se enquadravam nos critérios (inclusão e exclusão) de seleção e não respondiam às perguntas de pesquisa e, portanto, deveriam ter sido excluídos anteriormente.

Esses artigos foram então excluídos, não tendo sido feita sua avaliação de qualidade.

O quadro 6, apresenta as questões aplicadas. Para avaliar os artigos, foi utilizada a escala de três pontos de Likert (1932):

0: Não existe nada no artigo que atenda ao critério avaliado;

0.5: O artigo não deixa claro se atende ou não ao critério;

1: O artigo atende ao critério avaliado.

**Quadro 6:** Questões para avaliação dos artigos

Questões
1. É um artigo de pesquisa?
2. Existe uma descrição clara dos objetivos da pesquisa?
3. Existe uma descrição adequada do contexto em que o estudo foi realizado?
4. O desenho de pesquisa foi adequado para atender os objetivos da pesquisa?
5. A estratégia de seleção da amostragem foi adequada aos objetivos da pesquisa?
6. Os dados foram coletados de maneira adequada a responder as questões?
7. A análise dos dados foi suficientemente rigorosa?
8. A relação entre os pesquisadores e demais foi adequadamente considerada?
9. Há uma descrição clara dos resultados?
10. O estudo possui valor para a academia ou para a indústria?

**Fonte:** Adaptado de DYBÅ e DINGSØYR (2008)



A partir do somatório das notas de todos os critérios, os artigos foram classificados em quatro faixas de qualidade de acordo com a pontuação obtida, conforme apresentado na Quadro 7. Os artigos com somatório classificado nas faixas Média, Alta e Muito Alta foram encaminhados para extração, os demais foram descartados nesta etapa.

**Quadro 7:** Pontuação por faixa de qualidade

Baixa	Média	Alta	Muito Alta
$0 \leq N \leq 3$	$4 \leq N \leq 5,5$	$6 \leq N \leq 8,5$	$9 \leq N \leq 10$

**Fonte:** autora

A seguir temos o quadro 8, com os pontos obtidos, bem como os artigos que não foram avaliados, pois deveriam ter sido excluídos em fases anteriores.

**Quadro 8:** Pontuação dos Artigos

Avaliação dos Artigos					
ID	Título	Autores	Fonte	Pontuação	Qualidade
1	<b>Towards Event Sequence Representation, Reasoning and Visualization for EHR Data</b>	Cui Tao and Krist Wongsuphasawat and Kimberly Clark and Catherine Plaisant and Ben Shneiderman and Christopher G. Chute	ACM	7,5	Alta
2	<b>Challenges in Understanding Clinical Notes: Why NLP Engines Fall Short and Where Background Knowledge Can Help</b>	Sujan Perera and Amit Sheth and Krishnaprasad Thirunarayan and Suhas Nair and Neil Shah	ACM	8,5	Alta
3	<b>Parsing Clinical Text: How Good Are the State-of-the-Art Parsers?</b>	Min Jiang and Yang Huang and Jun-wei Fan and Buzhou Tang and Josh C. Denny and Hua Xu	ACM	Excluído	Não avaliado
4	<b>Leveraging Natural Language Processing of Clinical Narratives for Phenotype Modeling</b>	Preethi Raghavan and Albert M. Lai	ACM	9	Muito Alta
5	<b>Experiences with Mining Temporal Event Sequences from Electronic Medical Records:</b>	Debprakash Patnaik and Patrick Butler and Naren Ramakrishnan and Laxmi Parida and Benjamin J.	ACM	Excluído	Não avaliado

	<b>Initial Successes and Some Challenges</b>	Keller and David A. Hanauer			
6	<b>Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records</b>	Taowei David Wang and Catherine Plaisant and Alexander J. Quinn and Roman Stanchak and Shawn Murphy and Ben Shneiderman	ACM	Excluído	Não avaliado
7	<b>Multidimensional Temporal Mining in Clinical Data</b>	Shusaku Tsumoto and Shoji Hirano	ACM	Excluído	Não avaliado
8	<b>Temporal Phenotyping from Longitudinal Electronic Health Records: A Graph Based Framework</b>	Chuanren Liu and Fei Wang and Jianying Hu and Hui Xiong	ACM	Excluído	Não avaliado
9	<b>V-model: A New Innovative Model to Chronologically Visualize Narrative Clinical Texts</b>	Heekyong Park and Jinwook Choi	ACM	8,5	Alta
10	<b>A Corpus of Clinical Narratives Annotated with Temporal Information</b>	Lucian Galescu and Nate Blaylock	ACM	Excluído	Não avaliado
11	<b>A semi-supervised approach for temporal information extraction from clinical text</b>	G. Moharasar; T. B. Ho	IEEE	8	Alta
12	<b>Mining the Clinical Narrative: All Text are Not Equal</b>	K. Feldman; N. Hazekamp; N. V. Chawla	IEEE	Excluído	Não avaliado
13	<b>Towards comprehensive longitudinal healthcare data capture</b>	D. Cameron; V. Bhagwan; A. P. Sheth	IEEE	Excluído	Não avaliado
14	<b>Integrated Decision Support by Combining Textual Information Extraction, Facetted Search and</b>	D. Sonntag; H. J. Profitlich	IEEE	Excluído	Não avaliado

	<b>Information Visualisation</b>				
15	<b>Patient information extraction in noisy tele-health texts</b>	M. Y. Kim; Y. Xu; O. Zaiane; R. Goebel	IEEE	9	Muito Alta
16	<b>Extracting Modifiable Risk Factors from Narrative Preventive Healthcare Guidelines for EHR Integration</b>	S. Shah; X. Luo	IEEE	Excluído	Não avaliado
17	<b>MTPGraph: A Data-Driven Approach to Predict Medical Risk Based on Temporal Profile Graph</b>	S. Zhang; L. Liu; H. Li; Z. Xiao; L. Cui	IEEE	Excluído	Não avaliado
18	<b>Towards practical temporal relation extraction from clinical notes: An analysis of direct temporal relations</b>	H. J. Lee; Y. Zhang; J. Xu; C. Tao; H. Xu; M. Jiang	IEEE	7,5	Alta
19	<b>Task-Oriented Extraction of Temporal Information: The Case of Clinical Narratives</b>	R. Gaizauskas; H. Harkema; M. Hepple; A. Setzer	IEEE	Excluído	Não avaliado
20	<b>Design of an extensive information representation scheme for clinical narratives.</b>	Louise Deléger, Leonardo Campillos, Anne-Laure Ligozat, Aurélie Névoul	Pubmed	5,5	Média
21	<b>Normalization of relative and incomplete temporal expressions in clinical narratives.</b>	Weiyi Sun, Anna Rumshisky, Ozlem Uzuner	Pubmed	9,5	Muito Alta
22	<b>CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives.</b>	Cui Tao, Harold R. Solbrig, Christopher G. Chute	Pubmed	9,5	Muito Alta

23	<b>GUIR at SemEval-2017 Task 12: A Framework for Cross-Domain Clinical Temporal Information Extraction</b>	MacAvaney, Sean; Cohan, Arman; Goharian, Nazli;	Google	Excluído	Não avaliado
24	<b>Extracting temporal information from electronic patient records</b>	Li, Min; Patrick, Jon;	Google	9,5	Muito Alta
25	<b>Valid-Time Indeterminacy in Temporal Relational Databases: A Family of Data Models</b>	L. Anselma; P. Terenziani; R. T. Snodgrass	IEEE	8,5	Alta
26	<b>Mining of indeterminacy temporal data based on Fuzziness</b>	R. Shuxia; Z. Zheng	IEEE	7,5	Alta
27	<b>A representational analysis of a temporal indeterminacy display in clinical events</b>	M. Madkour; Hsing-yi Song; Jingcheng Du; C. Tao	IEEE	9	Muito Alta
28	<b>Query of Indeterminacy Temporal Data Based on CPR</b>	R. Shuxia; Z. Zheng	IEEE	9	Muito Alta
29	<b>Coping with Events in Temporal Relational Databases</b>	P. Terenziani	IEEE	Excluído	Não avaliado
30	<b>A Comprehensive Approach to ‘Now’ in Temporal Relational Databases: Semantics and Representation</b>	L. Anselma; L. Piovesan; A. Sattar; B. Stantic; P. Terenziani	IEEE	Excluído	Não avaliado
31	<b>Nearly Periodic Facts in Temporal Relational Databases</b>	P. Terenziani	IEEE	9,5	Muito Alta
32	<b>Valid-Time Indeterminacy in Temporal Relational Databases: Semantics and Representations</b>	L. Anselma; P. Terenziani; R. T. Snodgrass	IEEE	9,5	Muito Alta

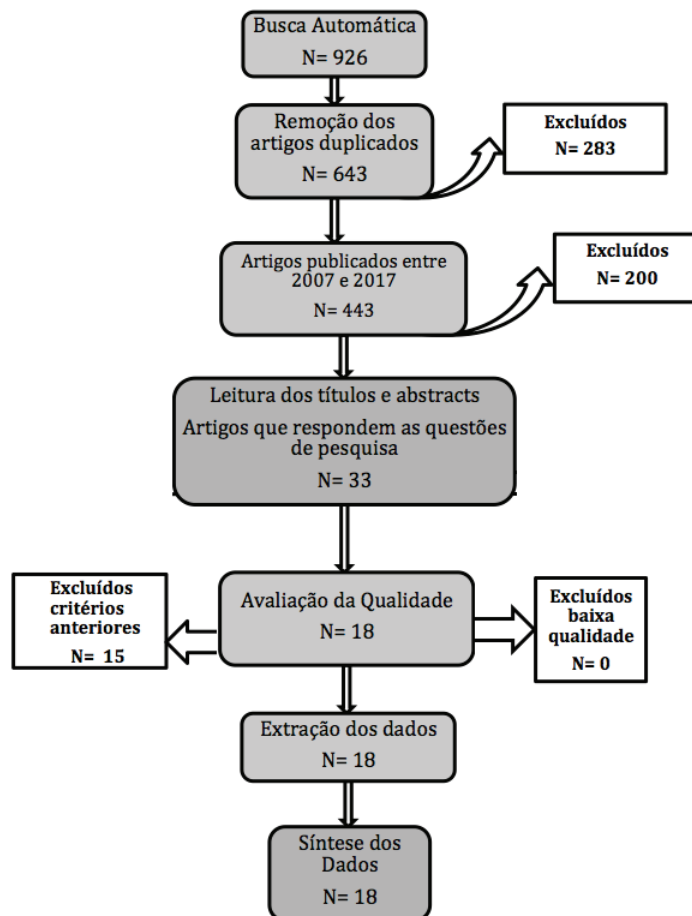
33	<b>Irregular Indeterminate Repeated Facts in Temporal Relational Databases</b>	P. Terenziani	IEEE	9,5	Muito Alta
----	--	---------------	------	-----	------------

Fonte: Autora

Como podemos visualizar o quadro acima, na fase de avaliação dos artigos, foram excluídos 15 estudos não por baixa qualidade, mas por ter sido identificado que não respondiam as questões de pesquisa e que deveriam ter sido excluídos em fases anteriores, restando 18 artigos que foram utilizados na extração e síntese dos dados dessa revisão sistemática.

A figura 7, apresenta as etapas do trabalho e os números para uma melhor compreensão.

**Figura 7:** Artigos por fase



Fonte: autora

Conforme podemos observar na Tabela 1, dos 18 artigos selecionados, 4 artigos (22%) foram provenientes da Base da ACM Digital Library, 10 artigos (55%) foram provenientes do IEEE Xplore, 3 artigos (17%) da Pubmed e 1 artigo (6%) proveniente da base do Google Acadêmico.

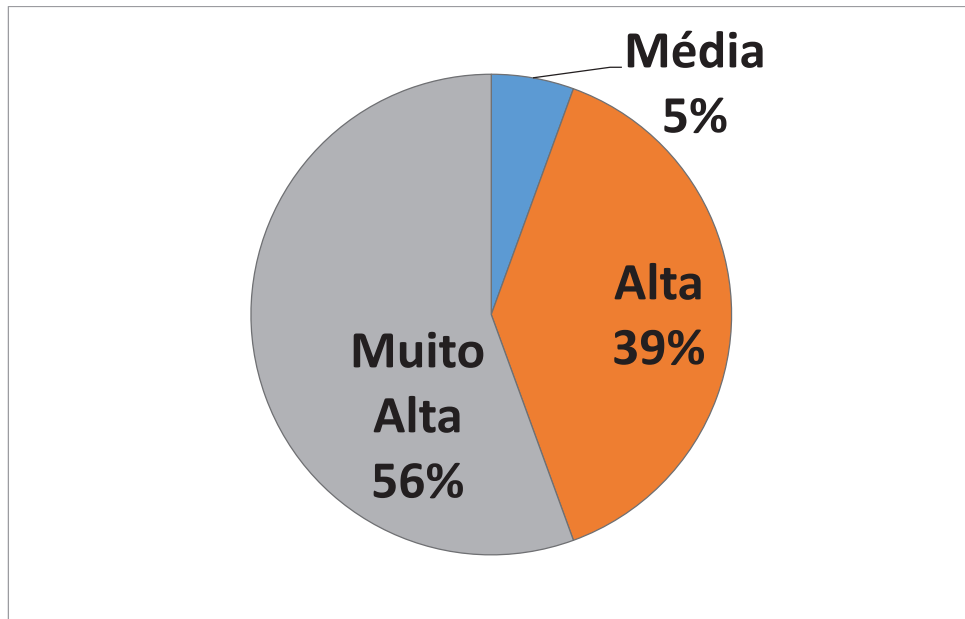
**Tabela 1:** Artigos por fonte de busca

Base	Seleção Inicial	Excluídos			Artigos Selecionados	
		Critérios de IN e EX	Baixa Qualidade	Total Excluídos	Qtd	% Seleção
ACM	592	588	0	588	4	22%
IEEE	134	124	0	124	10	55%
Pubmed	28	25	0	25	3	17%
Google Acadêmico	172	171	0	171	1	6%
<b>Totais</b>	<b>926</b>	<b>908</b>	<b>0</b>	<b>908</b>	<b>18</b>	<b>100%</b>

**Fonte:** autora

A Figura 8 mostra a qualidade dos 18 artigos selecionados, por faixa. Um estudo (1%) foi classificado com qualidade Média, sete (39%) com qualidade Alta e dez (56%) com qualidade Muito Alta.

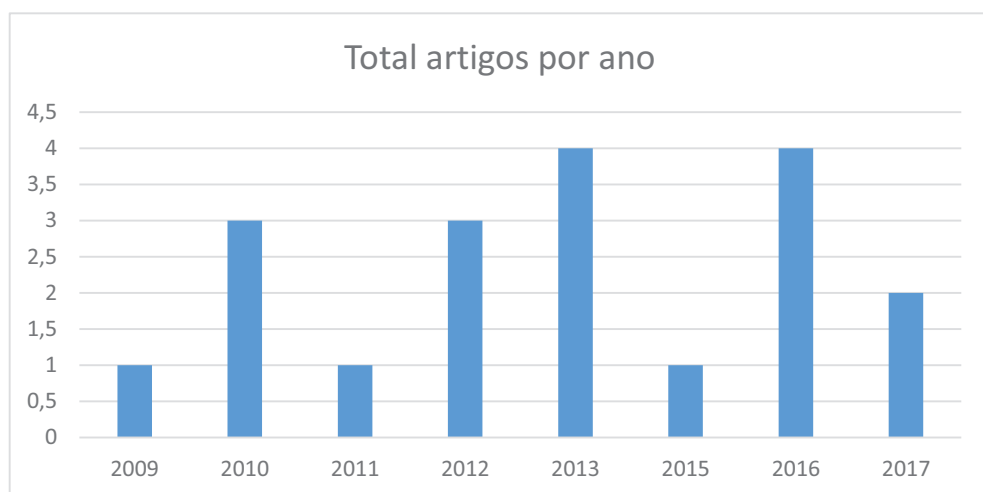
**Figura 8:** Qualidade dos artigos



Fonte: Autora

Além da boa qualidade, a grande parte dos artigos selecionados foram publicados nos últimos cinco anos, reforçando assim a relevância do assunto objeto deste estudo, conforme podemos visualizar na figura abaixo:

**Figura 9:** Total de artigos por ano



Fonte: Autora

Dos 18 artigos selecionados, 11 apresentam tratamentos e métodos realizados em dados temporais e 7 em indeterminação temporal, abaixo o quadro 9, apresenta os 11 artigos que trazem tratamentos e métodos para dados temporais em textos e narrativas clínicas.

**Quadro 9:** Tratamentos para dados Temporais

<b>Artigos com propostas para dados temporais</b>				
<b>ID</b>	<b>Título</b>	<b>Autores</b>	<b>Fonte</b>	<b>Tratament o</b>
<b>1</b>	<b>Towards Event Sequence Representation, Reasoning and Visualization for EHR Data</b>	Cui Tao and Krist Wongsuphasawat and Kimberly Clark and Catherine Plaisant and Ben Shneiderman and Christopher G. Chute	ACM	Dado Temporal
<b>2</b>	<b>Challenges in Understanding Clinical Notes: Why NLP Engines Fall Short and Where Background Knowledge Can Help</b>	Sujan Perera and Amit Sheth and Krishnaprasad Thirunarayan and Suhas Nair and Neil Shah	ACM	Dado Temporal
<b>3</b>	<b>Leveraging Natural Language Processing of Clinical Narratives for Phenotype Modeling</b>	Preethi Raghavan and Albert M. Lai	ACM	Dado Temporal
<b>4</b>	<b>V-model: A New Innovative Model to Chronologically Visualize Narrative Clinical Texts</b>	Heekyong Park and Jinwook Choi	ACM	Dado Temporal
<b>5</b>	<b>A semi-supervised approach for temporal information extraction from clinical text</b>	G. Moharasar; T. B. Ho	IEEE	Dado Temporal
<b>6</b>	<b>Patient information extraction in noisy tele-health texts</b>	M. Y. Kim; Y. Xu; O. Zaiane; R. Goebel	IEEE	Dado Temporal
<b>7</b>	<b>Towards practical temporal relation extraction from clinical notes: An analysis of direct temporal relations</b>	H. J. Lee; Y. Zhang; J. Xu; C. Tao; H. Xu; M. Jiang	IEEE	Dado Temporal



8	<b>Design of an extensive information representation scheme for clinical narratives.</b>	Louise Deléger, Leonardo Campillos, Anne-Laure Ligozat, Aurélie Névéol	Pubmed	Dado Temporal
9	<b>Normalization of relative and incomplete temporal expressions in clinical narratives.</b>	Weiyi Sun , Anna Rumshisky , Ozlem Uzuner	Pubmed	Dado Temporal
10	<b>CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives.</b>	Cui Tao, Harold R. Solbrig, Christopher G. Chute	Pubmed	Dado Temporal
11	<b>Extracting temporal information from electronic patient records</b>	Li, Min; Patrick, Jon;	Google	Dado Temporal

Fonte: Autora

Abaixo podemos visualizar os 7 artigos selecionados que trazem estudos para tratamento da questão da indeterminação temporal.

#### Quadro 10: Tratamentos para Indeterminação Temporal

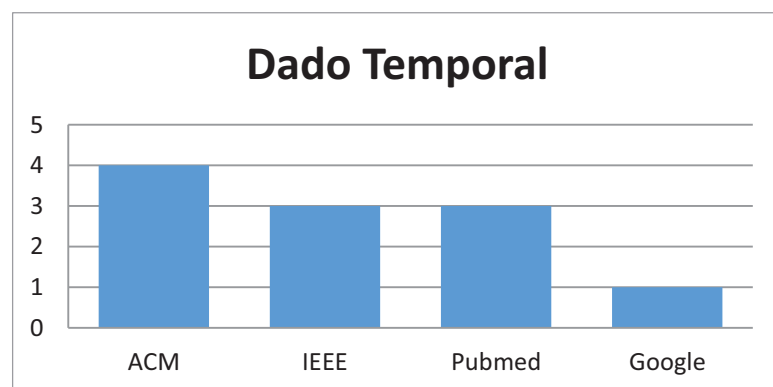
<b>Artigos com tratamentos da Indeterminação Temporal</b>				
<b>ID</b>	<b>Título</b>	<b>Autores</b>	<b>Fonte</b>	<b>Tratamento</b>
1	<b>Valid-Time Indeterminacy in Temporal Relational Databases: A Family of Data Models</b>	L. Anselma; P. Terenziani; R. T. Snodgrass	IEEE	Indeterminação Temporal
2	<b>Mining of indeterminacy temporal data based on Fuzziness</b>	R. Shuxia; Z. Zheng	IEEE	Indeterminação Temporal

3	<b>A representational analysis of a temporal indeterminacy display in clinical events</b>	M. Madkour; Hsing-yi Song; Jingcheng Du; C. Tao	IEEE	Indeterminação Temporal
4	<b>Query of Indeterminacy Temporal Data Based on CPR</b>	R. Shuxia; Z. Zheng	IEEE	Indeterminação Temporal
5	<b>Nearly Periodic Facts in Temporal Relational Databases</b>	P. Terenziani	IEEE	Indeterminação Temporal
6	<b>Valid-Time Indeterminacy in Temporal Relational Databases: Semantics and Representations</b>	L. Anselma; P. Terenziani; R. T. Snodgrass	IEEE	Indeterminação Temporal
7	<b>Irregular Indeterminate Repeated Facts in Temporal Relational Databases</b>	P. Terenziani	IEEE	Indeterminação Temporal

Fonte: Autora

Dos 11 artigos que apresentam tratamentos realizados em dados temporais, 4 artigos foram selecionados da base da ACM, 3 artigos da base da IEEE, 3 artigos da base da Pubmed e apenas 1 artigo da base do Google Acadêmico, conforme apresentado na figura 10.

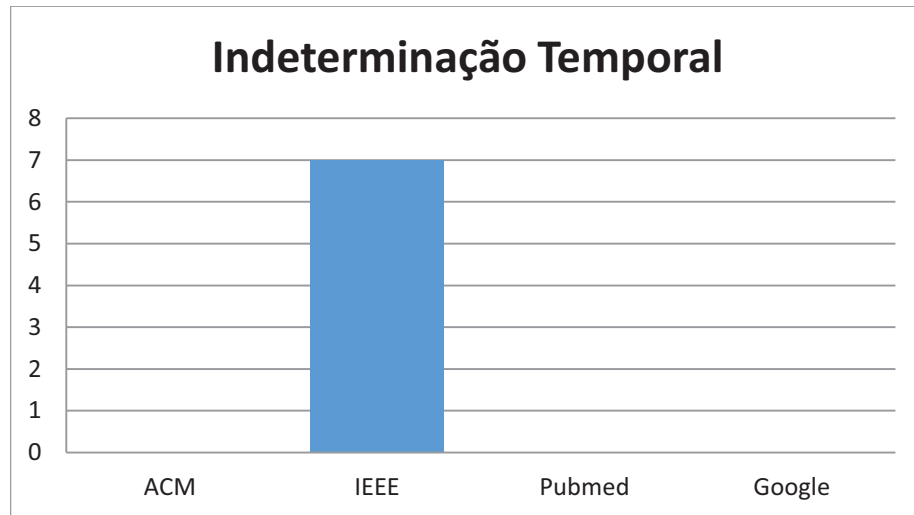
**Figura 10:** Artigos Dado Temporal



Fonte: autora

Já os 7 artigos que traziam tratamentos e métodos para o tratamento da indeterminação temporal, foram selecionados da base da IEEE, conforme apresentado na figura 11.

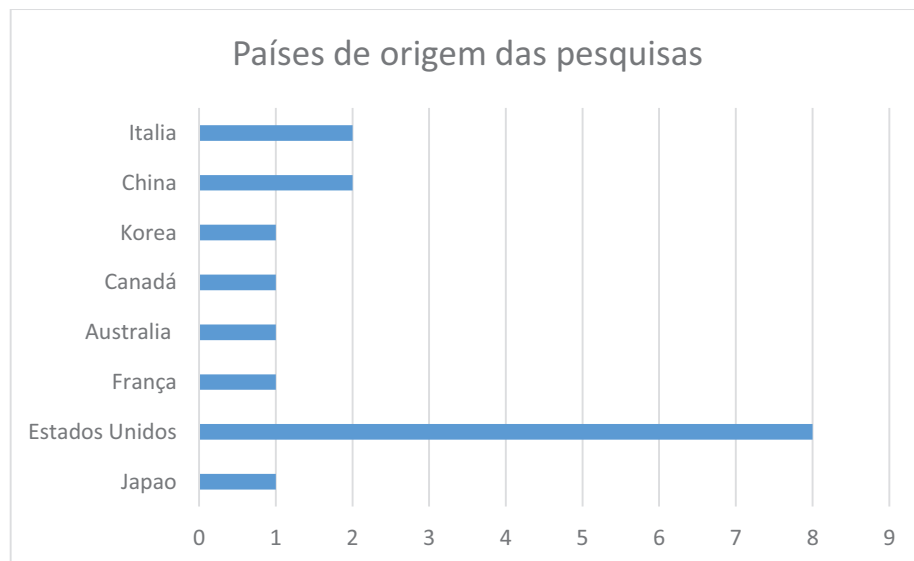
**Figura 11:** Artigos Indeterminação Temporal



**Fonte:** autora

Em relação aos países envolvidos, podemos perceber como mostra a figura 12, que os Estados Unidos é o país que mais se publica artigos abordando essa temática, seguido por China e Itália.

**Figura 12:** Países de Origem das Pesquisas



**Fonte:** autora

Por meio das figuras e tabelas que foram apresentados, compreende-se as características dos 18 artigos selecionados. A seguir, é descrito, como os dados dos artigos selecionados foram extraídos, para a análise dos resultados.

#### 4.4 Extração dos dados

Nesta fase foi realizada a extração dos dados dos artigos resultantes da fase anterior. O processo consistiu em extrair os dados de forma estruturada, utilizando um formulário padrão, conforme a figura 13.

Foram extraídos dados dos trabalhos selecionados, tais como tipo de estudo, métodos de pesquisa, análise dos dados, tamanho da amostra, requisitos de software, trabalhos futuros e limitações, objetivando responder as questões de pesquisa, conforme sugerido por Cruzes e Dybå (2011).

**Figura 13:** Formulário Padrão

Id:		Título:	
Descrição Geral:			
Método de Pesquisa:			
Tratamento:			
Trabalhos Futuros:			
QP1: Quais são os tipos de tratamento realizados em dados temporais e na indeterminação temporal em textos da área da saúde?			
QP2: Quais são os métodos computacionais existentes que tratem da questão tempo e a falta dele em narrativas clínicas e textos de autorias de pacientes?			
QP3: Quais são os métodos computacionais existentes para o tratamento da questão temporal?			

QP4: Quais são os métodos computacionais existentes para encontrar e tratar a indeterminação temporal.

**Fonte:** Autora (2018)

A síntese e a análise dos dados foram construídas quase que paralelamente, baseadas em uma abordagem qualitativa, para resumir as evidências extraídas dos estudos primários incluídos nesta pesquisa (KITCHENHAM; CHARTERS, 2007).

A seguir temos os resultados obtidos com a revisão sistemática da literatura, bem como a análise dos dados extraídos dos artigos selecionados.

## 5 RESULTADOS E DISCUSSÃO

Nesta seção é apresentada a síntese dos resultados da revisão sistemática considerando as respostas às questões de pesquisa abaixo:

**QP1: Quais são os tipos de tratamento realizados em dados temporais e na indeterminação temporal em textos da área da saúde?**

**QP2: Quais são os métodos computacionais existentes que tratem da questão tempo e a falta dele em narrativas clínicas e textos de autorias de pacientes?**

**QP3: Quais são os métodos computacionais existentes para o tratamento da questão temporal?**

**QP4: Quais são os métodos computacionais existentes para encontrar e tratar a indeterminação temporal.**

O processamento de Registros Médicos Eletrônicos (EMR), exige de acordo com os estudos selecionados, o tratamento preciso e a superação de vários desafios técnicos e não técnicos, especificamente quando em 80% dos dados hospitalares e da saúde não são estruturados.

Analisar esses registros para fins de pesquisa ou na tomada de decisão diante de um evento ou paciente, faz com que o estudo e a compreensão dos mesmos sejam de extrema importância. Um aspecto importante dos registros clínicos é justamente a ordem temporal em que os eventos ocorrem, sendo o tempo um fator essencial na saúde, permitindo a visualização e a compreensão da progressão de uma doença, as possíveis causas, o quanto os eventos passados podem explicar as situações atuais e a previsão de eventos futuros.

Nos textos selecionados, podemos observar os diversos tipos de tratamentos que são utilizados nessa questão temporal (dados temporais) e até mesmo na indeterminação temporal, que é quando em narrativas ou em registros eletrônicos, a questão tempo não fica explícita nos textos ou nas falas dos pacientes, sendo necessário assim alguns métodos para extração e ligação dos fatos temporais.

Dentre os tratamentos selecionados, em um deles os autores desenvolveram um sistema denominado LifeFlow, que trata a saída de dados de uma API já existente a CNTRO, permitindo assim uma interface mais intuitiva e agradável ao usuário final e que a partir de

dados temporais em textos e narrativas clínicas existentes em EHR, encontram padrões em ocorrências de doenças, medicação, tendências e sequenciamento de eventos (TAO et al., 2012).

A LifeFlow agrupa os registros relacionados aos pacientes, identificando o fluxo de eventos e as sequencias das ocorrências ao longo do tempo, facilitando assim a visualização em um único diagrama, além de possuir ferramentas para interação do usuário final com a ferramenta, permitindo zoom para alterar a granularidade do tempo, seleção de registros e dados específicos, dentre outros recursos. De acordo com os autores, a ferramenta demonstrou-se viável para visualização dos dados inferidos pelo sistema CNTRO, mas a ferramenta ainda conta com algumas possibilidades de melhorias tais como, lidar com incertezas.

Em outro estudo os autores deixam claro a complexidade dos problemas relacionados às questões temporais em textos e narrativas clínicas, devido às deficiências das técnicas de Processamento de Linguagem Natural (SUJAN, et. al., 2013).

Citam que existem algumas ferramentas tais como CTAKES, MedLEE e MetaMap, que trabalham com processamento de linguagem natural, mas são deficientes em alguns aspectos, tais como quando se deparam com declarações na negativa, condicionais, incertas, históricas e declarações sobre histórico familiar do paciente, quando trabalham com dados não-estruturados.

Sendo assim, os autores construíram uma base de dados extraídos de EMR, criaram versões estruturadas, utilizando documentos XML e realizaram os inputs nas ferramentas citadas, apresentando assim melhoras nas respostas retornadas após o processamento dos dados por parte dos algoritmos de Processamento de Linguagem Natural.

Os autores em outro estudo selecionado, afirmam que analisar as diversas fontes de dados clínicos de um paciente, ou seja, sua história médica não é nada trivial, pois grande parcela dos textos e relatórios clínicos tendem a ser desestruturados ou não estruturados, ruidosos e em uma linguagem médica nada usual. Por esse motivo, eles propuseram primeiramente o estudo às estruturas dos diferentes tipos de narrativas de pacientes e utilizaram posteriormente algoritmos de aprendizado de máquina para auxiliar na extração e classificação de dados e características clínicas dos pacientes (RAGHAVAN, et. al., 2010).

Em um outro estudo, os autores desenvolveram uma modelagem temporal denominada V-Model, apropriada para visualizar cronologicamente o histórico de um paciente, a partir dos dados que estão descritos em textos ou narrativas clínicas (PARK, etl. al., 2012).

Uma abordagem semi-supervisionada para extração de eventos temporais em narrativas e textos clínicos, também é apresentada em um outro estudo. Para o reconhecimento de expressões temporais contaram com o auxílio da ferramenta HeidelTime, que extrai expressões temporais de documentos e os normaliza de acordo com o padrão de notação denominado TIMEX3, onde a partir desses dados, desenvolveram um novo framework para extrair eventos temporais, estendendo assim os modelos já existentes, mas que trabalhavam somente com a abordagem supervisionada. Como testes de sua aplicação utilizaram os dados contidos em I2B2<sup>1</sup>. A ferramenta apresentou bom desempenho na detecção de eventos, mas não conseguiu detectar abreviaturas, sendo assim algo a se solucionar em trabalhos futuros (MOHARASAR, et. al., 2016).

Neste artigo os autores exploram métodos eficientes e eficazes para a extração de informação em narrativas clínicas, realizadas pelo sistema público de consultoria em saúde chamado HealthLink. Os dados disponíveis consistem em diálogos realizados e transcritos por enfermeiras durante atendimento em conversas telefônicas. De acordo com os autores, nesses tipos de troca de mensagens, há dois tipos de ruídos, o explícito e o implícito, o qual propõem soluções para cada um dos ruídos identificados. Descrevem no artigo dois métodos, um que trata o ruído explícito, o denominado Identificação de entidade biomédica baseada em modelo de reconhecimento de linguagem e o método que trata os ruídos implícitos, onde extraem somente informações dos pacientes e removem as informações ou frases na negativa. Também apresentam um sistema de detecção de temporalidade, que fornece um suporte a dados e expressões temporais (KIM, et. al., 2013).

Em um outro estudo selecionado os autores se concentram na identificação de relações temporais em textos e narrativas clínicas, por ser essencial na compreensão da informação temporal e na construção de uma linha de tempo dos eventos.

A maior dificuldade apontada pelos autores, são os dados temporais implícitos, que estão de certa forma mascarados ou embutidos em meio à narrativas, prescrições e descrições de pacientes, sendo que vários desafios já foram criados para tentar solucionar tais problemas.

---

<sup>1</sup> Uma das referências ao desenvolvimento de novas abordagens na área de mineração de textos em saúde, ou mais amplamente na área de Processamento de Linguagem Natural, são os desafios da i2B2, os i2B2 Challenges que ocorrem desde 2006. O Informatics for Integrating Biology and the Bedside (i2b2) é um centro de pesquisa financiado pelo NIH – National Institutes of Health, para fomentar o desenvolvimento dos EUA na infra-estrutura computacional para informática biomédica.



Algumas relações temporais são identificáveis apenas por inferência, pois as informações por muitas vezes se encontram espalhadas por um documento (LEE, et. al., 2017).

Para realização dessa avaliação, os autores utilizaram o sistema Vanderbilt, para identificar padrões e relações temporais implícitos e explícitos, para tal o sistema utiliza técnicas de *machine learning* para realizar essa classificação.

Em um outro estudo os autores descrevem o desenvolvimento de um esquema de representação de informações clínicas, extraíndo dados de narrativas, a fim de responder questões relacionadas à dados temporais. Eles propõem um esquema unificador que suporte uma grande variedade de questões clínicas, e realizam uma comparação de seu esquema, com outros esquemas já desenvolvidos (DELÉGER, et. al., 2017).

Neste artigo os autores descrevem o RI-TIMEXes (Expressões Temporais Relativas e Incompletas). Para os autores a normalização de dados RI-TIMEX é uma tarefa desafiadora e eles propõem nesse trabalho a normalização desse tipo de dados e apresentam um sistema que extraem das narrativas clínicas com dados RI-TIMEX, informações que se relacionam com o contexto onde esses dados estão inseridos.

O software CNTRO 2.0 (*Clinical Narrative Temporal Relation Ontology*) é apresentado em um outro estudo e foi desenvolvido para extrair de textos e narrativas clínicas, dados e aspectos temporais. O CNTRO 2.0 é uma evolução do CNTRO 1.0, o qual na nova versão foi incorporada uma nova lista de ontologias de tempo existentes, ou denominadas ontologias de alto nível, baseada em OWL (*Ontology Web Language*) de relações temporais para pesquisa clínica (TAO, et. al., 2011).

Em um outro trabalho os autores descrevem um novo método de extração de informações temporais em narrativas clínicas. Este método criado, foi integrado a um componente de processamento de documentos de um sistema de atendimento de perguntas clínicas existente, o qual se concentra em responder perguntas específicas de pacientes. Os autores utilizaram uma abordagem supervisionada de aprendizagem de máquina, que obteve resultados significativos em relação à outras abordagens existentes (LI, et. al., 2012).

Em dois trabalhos o autor propõe uma abordagem de referência (modelo de dados e álgebra) em que todos os possíveis cenários temporais induzidos por indeterminação de tempo válido podem ser modelados de forma extensiva e especifica uma família de dezesseis modelos de dados representacionais mais compactos.

Neste estudo o autor amplia o método proposto anteriormente por Snodgrass (2008), adicionando duas novas definições temporais de indeterminação instantânea, bem como intervalos de indeterminação: "*BeforeI*" e "simultaneidade". Ao apresentar um argumento chamado "*Fuzziness*" para descrever o grau de indeterminação, um DBMS (Gerenciador de Banco de Dados) temporal baseado em registros de pacientes com base em computador (CPR) é construído para implementar a mineração de dados temporais médicos de indeterminação (ANSELMA, et. al., 2017).

Em um outro estudo os autores apresentam uma proposição para representação da indeterminação temporal em narrativas e eventos clínicos, utilizando-se de funções de associação de conjuntos difusos, sendo esta abordagem responsável pela representação gráfica do tempo e sua granularidade usando lógica *fuzzy* (SHUXIA, 2009).

Para o processamento e consulta de dados temporais indeterminados, em um dos estudos foi construído um sistema de gerenciamento de banco de dados temporal para alguns hospitais. Para o processamento desses dados foram utilizados conceitos de ambiguidade e uso de método probabilístico (MADKOUR, et. al., 2016).

Em outro trabalho, os autores abordam a questão de eventos periódicos anuais, ou seja, fatos ou eventos que ocorrem em intervalos de tempo que se repetem periodicamente, afirmam que o tratamento nesses tipos de eventos é algo desafiador, pois envolve o tratamento integrado de três aspectos: o número de repetições, a periodicidade e a indeterminação temporal dos fatos. Neste trabalho, eles apresentam um novo modelo de dados e novas definições para operadores algébricos, utilizando-se de álgebra relacional para tratamento desses conjuntos de dados, onde suas repetições e periodicidade por muitas é vezes desconhecida (TERENZIANI, 2017).

E por fim um novo modelo de dados e novas definições de operadores algébricos relacionais que lidam com os problemas relacionados a repetição de dados e indeterminação temporal é testada a propriedade de redutibilidade e como ela pode ser integrada com outros modelos na literatura (TERENZIANI, 2017).

Como podemos observar, nos estudos apresentados, em cada um deles encontramos respostas para as questões de pesquisa que orientam esse trabalho, a seguir temos as conclusões desse trabalho.

## 6 CONCLUSÕES

Este trabalho dedicou-se a pesquisar, coletar, selecionar e avaliar por meio da revisão sistemática os estudos que traziam os tratamentos para dados temporais e de indeterminação temporal em narrativas clínicas e de autorias de pacientes, a fim de responder as questões de pesquisa que orientaram esse trabalho, possibilitando que os tratamentos que se encontravam dispersos, constassem em um único estudo e documento, resultando em maior visibilidade e também contribuindo para a área médica.

O avanço tecnológico, possibilitou que os dispositivos de armazenamento fossem capazes de guardar grandes quantidades de dados, também com a evolução dos sistemas de bancos de dados e a criação de gerenciadores mais robustos, permitiram que a quantidade de dados, variedade e velocidade aumentassem substancialmente, gerando o que ficou conhecido pelo nome *Big Data*.

Muitas áreas foram responsáveis por esse aumento e geração de dados e a área da saúde é uma delas, pois devido a sua natureza, os dados gerados, necessitam ser guardados como fonte de consulta para estudos, bem como para fins de legislação onde a sua manutenção é obrigatória. Também por meio desses registros é possível a extração de informações importantes do estado clínico, seu avanço, da evolução do quadro clínico, resultados de exames dos pacientes, evolução de doenças, dentre outras informações que são importantes para tomada de decisão e auxílio no tratamento.

E para que o tratamento seja efetivo, se faz necessário conhecer a evolução dos sintomas, quadro clínico, resultados dos exames, etc., é necessário que a sequências dos eventos sejam conhecidas, sendo o fator temporal e dados temporais imprescindíveis para organização e conhecimento das informações que estão armazenadas nos registros clínicos denominados *EHR (Eletronical Health Record)*, *EMR (Eletronic Medical Records)*, PEPs – Prontuários Eletrônicos dos Pacientes, narrativas de pacientes e médicos, que se encontram nos mais diversos formatos e são apresentados em formatos desestruturados, o que torna o trabalho de extração e síntese dessas informações um desafio técnico, o que foi explorado como objeto de estudo desse trabalho.

Pela análise e discussão dos resultados alcançados, foi possível responder tanto ao objetivo principal quanto aos específicos e à questão fundamental desse trabalho que era

analisar e avaliar os métodos computacionais para o tratamento de dados temporais e da indeterminação temporal em narrativas clínicas ou textos de autoria de pacientes.

Foi possível identificar as características e os tipos de tratamento que são feitos em dados temporais e na indeterminação temporal em textos da área da saúde onde por meio da revisão sistemática podemos perceber que existem alguns trabalhos abordando tratamentos em textos e narrativas clínicas.

Alguns dos trabalhos apontaram ferramentas de análise temporal e extração de textos que utilizam técnicas de Processamento de Linguagem Natural, como a CTAKES, MedLEE, MetaMap e HeidelTime, mas relataram que essas ferramentas são deficientes quando se trata de declarações na negativa, condicionais, incertas, históricas e declarações sobre o histórico familiar do paciente, quando se deparam com dados não estruturados.

Por isso, muitas pesquisas envolvem o aprimoramento desses programas, propondo o tratamento dos dados antes de submetê-los a análise das ferramentas. Também há trabalhos que facilitam a comunicação entre os usuários e os softwares de processamento de linguagem natural, em um dos trabalhos, os autores desenvolveram um sistema chamado LifeFlow, para o tratamento de dados de saída de uma API já existente a CNTRO, facilitando a interpretação dos dados por parte dos usuários.

Em outros tratamentos encontramos a utilização de algoritmos de aprendizado de máquina, para auxiliar na extração e classificação dos dados e características clínicas dos pacientes, bem como propostas de modelagem temporal, a fim de visualizar cronologicamente o histórico de um paciente que está descrito em textos ou narrativas clínicas.

Foi possível a identificação de tratamentos para o problema específico da indeterminação temporal, porém o que se observou é que quando se trata de dados temporais, a indeterminação está implícita, pois em grande parte dos trabalhos selecionados, os autores consideraram no seu tratamento os dados indeterminados, além dos dados temporais.

No tratamento da indeterminação temporal, foi observado que em muitos dos trabalhos, há o desenvolvimento e o aprimoramento da lógica temporal com a utilização de lógica Fuzzy, novos modelos de dados, álgebra relacional e método probabilístico.

Em vários trabalhos, foi possível a identificação de desafios e obstáculos para que os atributos temporais sejam identificados de forma correta e que os eventos sejam classificados e organizados de forma precisa. Um dos desafios é o fato de que os dados são desestruturados,

além dos ruídos presentes nas narrativas e também na dificuldade de se trabalhar com dados implícitos.

Além das dificuldades apontadas, o estudo também mostrou que, levando-se em consideração a importância do tema para a área médica, ainda assim são poucos trabalhos que abordam essas questões, visto a quantidade de trabalhos que foram selecionadas previamente.

Por meio dos trabalhos selecionados, podemos perceber uma tendência dos autores em tentarem apresentar um complemento ou uma ampliação aos tratamentos já existentes, visando o aprimoramento das técnicas. Visto que os tratamentos propostos não são 100% eficientes, e sempre apresentam alguma falha dependendo dos tipos de dados o qual trabalham.

Dentro disso, é adequado afirmar que existem alguns tratamentos quando se fala de dados temporais e da indeterminação temporal em dados clínicos e nas narrativas de pacientes e de médicos, mas por conta da natureza dos dados, sua variedade e quantidade, a pesquisa no tratamento desses dados ainda enfrenta desafios, pois em nenhum dos tratamentos o resultado foi satisfatório, permitindo a extração e síntese na totalidade dos dados temporais, sendo uma área de estudo ainda a ser explorada.

Contudo, apesar da pesquisa ter conseguido responder os seus objetivos, ela apresenta alguns pontos fracos, pois poderia ter sido realizada utilizando-se de outras fontes e bases, para complementar a busca por mais trabalhos e também se utilizar de outros termos de busca, deixando a pesquisa mais ampla.

Como sugestão para trabalhos futuros podemos destacar a ampliação do estudo, considerando tratamentos em dados temporais e de indeterminação temporal em outros tipos de dados clínicos ou até mesmo os coletados por meio de *wearable devices* e utilizar os tratamentos selecionados para expandir e detalhar na prática, sua aplicação e desempenho na análise de dados temporais.

## REFERÊNCIAS

ALVES, A. H. Avaliando as características de ferramentas para a gestão do conhecimento. Dissertação (Bacharelado em Sistemas de Informação) — Instituto Superior Tupy, Joinville, 2005.

ANS – Agência Nacional de Saúde Suplementar. Disponível em: < <http://www.ans.gov.br/> > Acesso em: 15 de Julho de 2017.

ANSELMA, L; PIOVESAN, L; SATTAR, A; STANTIC, B; TERENCEZIANI, P. **A comprehensive approach to 'now' in temporal relational databases:** Semantics and representation. Disponível em: < <http://ieeexplore.ieee.org/document/7506328/>>. Acesso em: 10 de Novembro de 2017.

ANSELMA, L., TERENCEZIANI, P., SNODGRASS, R. **Valid-Time Indeterminacy in Temporal Relational Databases:** A Family of Data Models. Disponível em: < <http://ieeexplore.ieee.org/document/5601875/>>. Acesso em: 10 de Novembro de 2017.

ANSELMA, L., TERENCEZIANI, P., SNODGRASS, R. **Valid-Time Indeterminacy in Temporal Relational Databases:** Semantics and Representation. Disponível em: <http://ieeexplore.ieee.org/document/6329892/>. Acesso em: 10 de Novembro de 2017.

ASTRAHAN, M., BLASGEN, W., CHAMBERLIN, D., ESWARAN, P., GRAY, J., GRIFFITHS, P., KING, F., LORIE, A., MCJONES, R., MEHL, W., PUTZOLU, R., TRAIGER, L., WADE, W., WATSON. V. 1976. **System R:** relational approach to database management. ACM Trans. Database. Disponível em: < <https://dl.acm.org/citation.cfm?id=320457> >. Acesso em: 22 de Junho de 2017.

BAIN & COMPANY. Big data Organizational Challenge. Disponível em: [http://www.bain.com/publications/articles/big\\_data\\_the\\_organizational\\_challenge.aspx](http://www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx). Acesso em: 15 de Julho de 2017.

BANCO MUNDIAL. Disponível em: < <https://data.worldbank.org/country/BR?locale=pt>>. Acesso em: 20 de Janeiro de 2018.

BAPTISTA, M.; OLIVEIRA, J.; SEZOES, C. Business Intelligence. 2008. Disponível em: <<http://www.slideshare.net/nesi/business-intelligence-20391>>. Acesso em: 26 Outubro. 2017.

BATAL, I., VALIZADEGAN, H., COOPER, G. F., HAUSKRECHT, M. A Pattern Mining Approach for Classifying Multivariate Temporal Data. Disponível em: < <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261774/> >. Acesso em: 26 Outubro. 2017.

BIAN, J., TOPALOGLU, U; YU, F. Towards large-scale twitter mining for drug-related adverse events. Disponível em: < <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5619871/> >. Acesso em: 26 Outubro. 2017.

BARBIERI, C. BI - Business Intelligence Modelagem e Tecnologia. Rio de Janeiro: Axcel Books, 2001.

BELLAZZI, R, Sacchi, L.; S. Concaro, "Methods and tools for mining multivariate temporal data in clinical and biomedical applications," 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, 2009, pp. 5629-5632. doi: 10.1109/IEMBS.2009.5333788

BURGHARD, C. - Big Data and Analytics Key to Accountable Care Success. 2012, IDC Health Insights

BORKAR, V., CAREY, M., LI, C. Big data platforms: what's next? **XRDS: Crossroads, The ACM Magazine for Students**. Disponível em: < <https://dl.acm.org/citation.cfm?id=2331042&picked=prox> >. Acesso em: 30 de Outubro de 2016.

CHAPPLE, M. - Data Mining: An Introduction. Disponível em: <<http://databases.about.com/od/datamining/a/datamining.htm>> 2008. Acesso em: 23 Nov. 2015

COMBI, C. - "Discovering Quantitative Temporal Functional Dependencies on Clinical Data," 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, 2017, pp. 248-257. doi: 10.1109/ICHI.2017.80

CAMERON, Delroy; BHAGWAN, Varun; SHETH, Amit P.. Towards comprehensive longitudinal healthcare data capture. **2012 Ieee International Conference On Bioinformatics And Biomedicine Workshops**, [s.l.], p.240-247, out. 2012. IEEE. <http://dx.doi.org/10.1109/bibmw.2012.6470310>.

COOPER, H. (2010). **Research synthesis and meta-analysis: A step-by-step approach** (3. ed.). Thousand Oaks, CA: Sage.

DATE, C. J. Introdução aos Sistemas de Bancos de Dados. Campus, 2000.

DELÉGER, L. Design of an extensive information representation scheme for clinical narratives. **Journal of Biomedical Semantics**, [s.l.], v. 8, n. 1, p.1-18, 11 set. 2017. Springer Nature. <http://dx.doi.org/10.1186/s13326-017-0135-z>.

DAVENPORT, T.; PRUSAK, L. **Conhecimento empresarial: como as organizações gerenciam seu capital intelectual**. Rio de Janeiro: Campus, 1998.

DAVENPORT, T. **A report by Harvard Business Review Analytic Services: Big Data: The Opportunity and the Challenge for Competitive Advantage**. Harvard Bussiness School Publishing- 2013.

DAVENPORT, T.; PRUSAK, L. **Conhecimento empresarial: como as organizações gerenciam seu capital intelectual**. Rio de Janeiro: Campus, 1998.

DEMBOSKY A: **Data Prescription for Better Healthcare**. Financial Times, December 12, 2012, p. 19. 2012. Disponível em: <<http://www.ft.com/intl/cms/s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz2W9cuwajK>>. Acesso em: 30 de Outubro de 2016.

DYBA, T., DINGSOYR, T. (2008) Empirical Studies of Agile Software Development: A Systematic Review. *Information and Software Technology*, 50, 833-859. <http://dx.doi.org/10.1016/j.infsof.2008.01.006>.

EDELWEISS, N. **Banco de Dados Temporais: Teoria e Prática**, 1998. Belo Horizonte.

EDELWEISS, N. Banco de Dados Temporais: Teoria e Prática. In: JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA, 17., 1998, Belo Horizonte. Anais do XVII Congresso Nacional da Sociedade Brasileira de Computação. Belo Horizonte: SBC, 1998. p. 225-282.

EDELWEISS, N.; OLIVEIRA, J. P. M. Modelagem de Aspectos Temporais de Sistemas de Informação. Recife: UFPE-DI, 1994. 163 p.

ELMASRI, R.; NAVATHE, S. B. *Fundamentals of Databases Systems*. Benjamin – Cummins, 1989.

FENASAÚDE – Federação Nacional de Saúde Suplementar. Disponível em: <<http://cnseg.org.br/fenasaude/publicacoes/boletim-da-saude-suplementar/>>. Acesso em: 15 de Julho de 2017.



FELDMAN, B, MARTIN, E., SKOTNES, T: Big Data in Healthcare Hype and Hope. Disponível em: <<http://www.west-info.eu/files/big-data-in-healthcare.pdf>> Acesso em: 30 de Outubro de 2016.

FELDMAN, K; HAZEKAMP, N; CHAWLA, N. Mining the Clinical Narrative: All Text are Not Equal. **2016 Ieee International Conference on Healthcare Informatics (ichi)**, [s.l.], p.271-280, out. 2016. IEEE. <http://dx.doi.org/10.1109/ichi.2016.37>.

FIORENTINO, G., SEBASTIÃO, B., MATTOS, L., GRASS, K. Tendências do Setor de Saúde no Brasil. Disponível em: < [http://www.bain.com/offices/saopaulo/en\\_us/Images/healthcare-trends-in-brazil\\_por.PDF](http://www.bain.com/offices/saopaulo/en_us/Images/healthcare-trends-in-brazil_por.PDF)> Acesso em: 30 de novembro de 2017.

FROST E SULLIVAN: Drowning in Big Data? Reducing Information Technology Complexities and Costs for Healthcare Organizations- 2010. Disponível em: <<http://www.emc.com/collateral/analyst-reports/-sullivan-reducing-information-technology-complexities-ar.pdf>>. Acesso em: 30 de Outubro de 2016.

GAIZAUSKAS, R. et al. Task-Oriented Extraction of Temporal Information: The Case of Clinical Narratives. **Thirteenth International Symposium On Temporal Representation And Reasoning (time'06)**, [s.l.], p.1-8, 2006. IEEE. <http://dx.doi.org/10.1109/time.2006.27>.

GALESCU, L; BLAYLOCK, N. A corpus of clinical narratives annotated with temporal information. **Proceedings of The 2nd Acm Sighit Symposium On International Health Informatics - Ihi '12**, [s.l.], p.715-719, 2012. ACM Press. <http://dx.doi.org/10.1145/2110363.2110446>.

GARTNER. **Answering Big Data's 10 Biggest Planning and Implementation Questions**. Disponível em: <<https://www.gartner.com/doc/2957917/answering-big-datas--biggest>>. Acesso em: 30 de Outubro de 2016.

GARTNER. **Major Myths About Big Data's Impact on Analytics**. Disponível em: <<https://www.gartner.com/doc/2846318/major-myths-big-datas-impact>>. Acesso em: 30 de Outubro de 2016.

CRAFT, L. Market Guide for Healthcare Provider Population Health Management Solutions: 2018 Update.

GINNI, R.- CEO IBM – **IBM: O mundo na era cognitiva – 2015**. Disponível em: <<https://www.ibm.com/blogs/robertoa/2015/10/ibm-o-mundo-na-era-cognitiva>>. Acesso em: 30 de Outubro de 2016.

GOLDSCHMIDT, R.; PASSOS, E. **Data Mining – Um Guia Prático**. Campus 2005.

GORALWALLA, A., ABUDULLAH, U., TANSEL, M., ÖZSU, T. - 1995. Experimenting with temporal relational databases. In Proceedings of the fourth international conference on Information and knowledge management (CIKM '95), Niki Pissinou, Avi Silberschatz, E. K. Park, and Kia Makki (Eds.). ACM, New York, NY, USA, 296-303. DOI: <https://doi.org/10.1145/221270.221597>

**HARVARD BUSINESS REVIEW – A report by Harvard Business Review Analytic Services:** Big Data: The future of Information and Business. Harvard Business School Publishing- 2012. Disponível em: < <https://hbr.org/2012/10/big-data-the-management-revolution>>. Acesso em: 15 de Setembro de 2016.

IQBAL A. Goralwalla, Abudullah U. Tansel, ; M. Tamer Özsu. 1995. Experimenting with temporal relational databases. In Proceedings of the fourth international conference on Information and knowledge management (CIKM '95), Niki Pissinou, Avi Silberschatz, E. K. Park, and Kia Makki (Eds.). ACM, New York, NY, USA, 296-303. DOI: <https://doi.org/10.1145/221270.221597>

JENSEN, C.S., SNODGRASS, R.T. **Temporal Database Entries for the Springer Encyclopedia of Database Systems: A TIMECENTER Technical Report**, Aalborg University & University of Arizona (2008).

JIANG, M. - Parsing clinical text: how good are the state-of-the-art parsers? **Bmc Medical Informatics and Decision Making**, [s.l.], v. 15, n. 1, p.1-6, 20 Maio 2015. Springer Nature. <http://dx.doi.org/10.1186/1472-6947-15-s1-s2>.

KITCHENHAN, B. **Systematic literature reviews in software engineering:** A systematic literature review, Information and Software Technology. Technical report. Nov. 2009. Disponível em: < <https://www.sciencedirect.com/science/article/abs/pii/S0950584908001390>>. Acesso em: 18 de Agosto de 2017,

KIM, M. et al. Patient information extraction in noisy tele-health texts. **2013 Ieee International Conference on Bioinformatics And Biomedicine**, [s.l.], p.326-329, dez. 2013. IEEE. <http://dx.doi.org/10.1109/bibm.2013.6732511>.

LEE, H. et al. Towards practical temporal relation extraction from clinical notes: An analysis of direct temporal relations. **2017 Ieee International Conference on Bioinformatics And**

**Biomedicine (bibm)**, [s.l.], p.1272-1275, nov. 2017. IEEE. <http://dx.doi.org/10.1109/bibm.2017.8217842>.

LEAVITT, N. Will NoSQL Databases Live Up to Their Promise? Disponível em: <<http://ieeexplore.ieee.org/document/5410700/>>

LING Li., ZSU. T. Encyclopedia of Database Systems (1st ed.). Springer Publishing Company, Incorporated.

LI, M; PATRICK, J. Extracting temporal information from electronic patient records. AMIA Annual Symposium Proceedings. vol. 2012. American Medical Informatics Association, 2012: 542

LIKERT, R. (1932). A technique for measurement of attitudes. Archives of Psychology, 140, 5-55.

LIU, C. et al. Temporal Phenotyping from Longitudinal Electronic Health Records. **Proceedings Of The 21th Acm Sigkdd International Conference On Knowledge Discovery And Data Mining - Kdd '15**, [s.l.], p.705-714, 2015. ACM Press. <http://dx.doi.org/10.1145/2783258.2783352>.

LOBO, L. Inteligência Artificial e Medicina. **Revista Brasileira de Educação Médica**, [s.l.], v. 41, n. 2, p.185-193, jun. 2017. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1981-52712015v41n2esp>.

LOHR, S. **The age of big data**, New York Times, vol. 11, 2012: Disponível em: <[http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?\\_r=0](http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=0)>. Acesso em: 18 de Agosto de 2017.

LÓSCIO, B. F., OLIVEIRA, H. R., PONTES, J. C. S. NoSQL no desenvolvimento de aplicações Web colaborativas., VIII Simpósio Brasileiro de Sistemas Colaborativos. Disponível em: [http://www.addlabs.uff.br/sbse\\_site/SBSC2011\\_NoSQL.pdf](http://www.addlabs.uff.br/sbse_site/SBSC2011_NoSQL.pdf). Acesso em: 18 de Agosto de 2017.

MACAVANEY, S; COHAN, Arman; GOHARIAN, Nazli. GUIR at SemEval-2017 Task 12: A Framework for Cross-Domain Clinical Temporal Information Extraction. **Proceedings Of The 11th International Workshop On Semantic Evaluation (semeval-2017)**, [s.l.], p.1024-1029, 2017. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/s17-2180>.

MADKOUR, M. et al. A representational analysis of a temporal indeterminacy display in clinical events. **2016 Ieee International Conference on Bioinformatics And Biomedicine (bibm)**, [s.l.], p.1089-1095, dez. 2016. IEEE. <http://dx.doi.org/10.1109/bibm.2016.7822673>.

MANIYKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C., BYERS A. H. **Big data: The next frontier for innovation, competition, and productivity**. Disponível em: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>>. Acesso em: 18 de Agosto de 2017.

MEEKER, M. **THE DATA: Next Big Wave** - Leveraging this unlimited connectivity & storage to collect / aggregate / correlate / interpret all of this data to improve people's lives & enable enterprises to operate more efficiently. Trends Report 2016. Disponível em: [http://dq756f9pzlyr3.cloudfront.net/file/2016\\_internet\\_trends\\_report\\_final.pdf](http://dq756f9pzlyr3.cloudfront.net/file/2016_internet_trends_report_final.pdf)>. Acesso em: 18 de Agosto de 2017.

McAFEE, A.; BRYNJOLFSSON, E. **Big Data: The Management Revolution**. Harvard Business Review. Disponível em: <https://hbr.org/2012/10/big-data-the-management-revolution/ar>>. Acesso em: 18 de Agosto de 2017.

MCKINSEY E COMPANY. Global Media Report. Disponível em: <http://www.mckinsey.com/industries/media-and-entertainment/our-insights/global-media-report-2014> >. Acesso em: 18 de Agosto de 2017.

MOHARASAR, G., HO, T. A semi-supervised approach for temporal information extraction from clinical text. **2016 Ieee Rivf International Conference on Computing & Communication Technologies, Research, Innovation, And Vision For The Future (rivf)**, [s.l.], p.7-12, nov. 2016. IEEE. <http://dx.doi.org/10.1109/rivf.2016.7800261>.

MORGENTHAL. J. - **Evolution in Data Integration From EII to Big Data**. Disponível em: <http://www.infoq.com/articles/DataIntegrationFromEIItoBigData>. Acesso em: 22 de Junho de 2017.

OMS – Organização Mundial da Saúde. Disponível em: <http://www.who.int/countries/bra/en/>>. Acesso em: 15 de Julho de 2017

PATNAIK, Debprakash et al. Experiences with mining temporal event sequences from electronic medical records. **Proceedings of The 17th Acm Sigkdd International Conference On Knowledge Discovery And Data Mining - Kdd '11**, [s.l.], p.360-368, 2011. ACM Press. <http://dx.doi.org/10.1145/2020408.2020468>.

PARK, Heekyong; CHOI, Jinwook. V-model. **Proceedings Of The 2012 Acm Annual Conference On Human Factors In Computing Systems - Chi '12**, [s.l.], p.453-462, 2012. ACM Press. <http://dx.doi.org/10.1145/2207676.2207739>.

PEARSON, T., WEGENER, R. **Bain Insights**. Bain & Company, 2013. Disponível em: <[http://www.bain.com/publications/articles/big\\_data\\_the\\_organizational\\_challenge.aspx](http://www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx)>. Acesso em: 22 de Junho de 2017.

POKORNY, J. - **NoSQL databases**. A step to database scalability in web environment. Disponível em: <<https://www.semanticscholar.org/paper/NoSQL-databases%3A-a-step-to-database-scalability-in-Pokorn%C3%BD/c7ef1e1c9ec1e8ed17b6b634b6963b37838b9614>>. Acesso em: 22 de Junho de 2017

RAGHAVAN, P.; LAI, A. - Leveraging natural language processing of clinical narratives for phenotype modeling. **Proceedings of The 3rd Workshop On Ph.d. Students In Information And Knowledge Management - Pilm '10**, [s.l.], p.57-65, 2010. ACM Press. <http://dx.doi.org/10.1145/1871902.1871913>.

RAGHUPATHI, W., RAGHUPATHI, V. **Big data analytics in healthcare: promise and potential** (2014). Disponível em: <<http://hissjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3>>. Acesso em: 22 de Junho de 2017.

RAGHUPATHI W: **Data Mining in Health Care**. Healthcare Informatics: Improving Efficiency and Productivity. Edited by: Kudyba S. 2010, Taylor & Francis, 211-223. Disponível em: <<http://www.crcnetbase.com/doi/10.1201/9781439809792-c11>>. Acesso em: 15 de Julho de 2017.

RAGHUPATHI, W., RAGHUPATHI, V. An overview of health analytics. **Journal of Health & Medical Informatics**, v. 2013, 2013. Disponível em: < <https://www.omicsonline.org/an-overview-of-health-analytics-2157-7420.1000132.pdf>>. Acesso em: 15 de Julho de 2017.

RAMAKRISHNAN, R.; GEHRKE, J. Database Management Systems- 3<sup>rd</sup> ed. McGraw- Hill, 2003.

RANSBOTHAM, S., KIRON, D., PRENTICE, P. K. **Beyond The Hype: The Hard Work Behind Analytics Success**. MIT Sloan Management Review. Disponível em: <https://sloanreview.mit.edu/projects/the-hard-work-behind-data-analytics-strategy/> . Acesso em: 15 de Julho de 2017.

SAKAI, R. - Extensão de um SGBD para Incluir o Gerenciamento da informação Temporal – São Paulo, 2007.

SAKR, S., LIU, A., BATISTA, D. M., ALOMARI, M. (2011) - A Survey of Large Scale Data Management Approaches in Cloud Environments. Disponível em: <<http://ieeexplore.ieee.org/document/5742778/>>. Acesso em: 15 de Julho de 2017.

SAVAGE, N. - Digging for drug facts. **Communications of the ACM**, v. 55, n. 10, p. 11-13, 2012.

SAMPAIO, R., MANCINI, M. **Estudos de Revisão Sistemática: Um guia para síntese criteriosa da evidência científica**. Rev. Bras. Fisioter. São Carlos, v. 11, n. 1, p. 83-89, jan. 2007.

SHAH, S; LUO, X. Extracting Modifiable Risk Factors from Narrative Preventive Healthcare Guidelines for EHR Integration. **2017 Ieee 17th International Conference on Bioinformatics And Bioengineering (bibe)**, [s.l.], p.514-519, out. 2017. IEEE. <http://dx.doi.org/10.1109/bibe.2017.000-2>.

SENGE, P. A dança das mudanças. Rio de Janeiro: Campus, 1999.

SERRA, L. A essência do Business Intelligence. São Paulo: Berkeley, 2002

SETZER, W. **Dado, informação, conhecimento e Competência**. Disponível em: <<http://www.ime.usp.br/~vwsetzer/dado-info-Folha.html>> Publicado no jornal do Grupo Folha, Folha Educação No. 27, out./nov. 2004, pp. 6 e 7). Acesso em: 15 de Julho de 2017.

SOUBHIK Paul, Sandip Bhaumik: A context and knowledge based health event storage for efficient clinical information retrieval. BIBM Workshops 2012: 822-823

SMITH, J. **jStart: How Big Data Analytics Reduced Medicaid Re-admissions**. A jStart Case Study. 2012. Disponível em: <<http://www-01.ibm.com/software/ebusiness/jstart/portfolio/uncMedicaidCaseStudy.pdf>>. Acesso em: 15 de Julho de 2017.

SOUZA, S. Especificação de um Modelo de Dados Bitemporal Orientado a Objetos. Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo (EPUSP), São Paulo, 1998.

STEWART, T. A. **Riqueza do conhecimento. O capital Intelectual e a organização do século XXI**. 1. ed. Rio de Janeiro: Campus, 2002.

SUJAN, P., AMIT, S., KRISHNAPRASAD, T. **Challenges in understanding clinical notes: why nlp engines fall short and where background knowledge can help.** Proceedings of ACM Conference of Information and Management, Burlingame, November 1, 2013.

SHUXIA R., ZHENG, Z. - Query of Indeterminacy Temporal Data Based on CPR. (2009) International Forum on Computer Science-Technology and Applications, Chongqing, 2009, pp. 163-167. doi: 10.1109/IFCSTA.2009.279

SHUXIA R., ZHENG, Z. -Mining of indeterminacy temporal data based on Fuzziness," 2010 2nd IEEE International Conference on Information Management and Engineering, Chengdu, 2010, pp. 391-393. doi: 10.1109/ICIME.2010.5478314

SNODGRASS, R. T. The TSQL2 Temporal Query Language. Kluwer Academic Publishers, 674+xxiv pages, 1995.

SNODGRASS, R.T. (1992) Temporal databases. In: Frank A.U., Campari I., Formentini U. (eds) Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science, vol 639. Springer, Berlin, Heidelberg

SONNTAG, D; PROFITLICH, Hans-jurgen. Integrated Decision Support by Combining Textual Information Extraction, Facetted Search and Information Visualisation. **2017 Ieee 30th International Symposium on Computer-based Medical Systems (cbms)**, [s.l.], p.95-100, jun. 2017. IEEE. <http://dx.doi.org/10.1109/cbms.2017.119>.

SUN, W; RUMSHISKY, A; UZUNER, Ozlem. Normalization of relative and incomplete temporal expressions in clinical narratives. **Journal of The American Medical Informatics Association**, [s.l.], v. 22, n. 5, p.1001-1008, 12 abr. 2015. Oxford University Press (OUP). <http://dx.doi.org/10.1093/jamia/ocu004>.

TAO. C., SOLBRIG. HR, CHUTE CG. **CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives.** AMIA Summits Transl Sci Proc. 2011; 2011:64-8. Epub 2011 Mar 07. PMID:22211182. PMCID:3248753.

TAO, C. et al. Towards event sequence representation, reasoning and visualization for EHR data. **Proceedings of The 2nd Acm Sighit Symposium On International Health Informatics - Ihi '12**, [s.l.], p.801-805, 2012. ACM Press. <http://dx.doi.org/10.1145/2110363.2110461>.



TERENZIANI, P. - Coping with Events in Temporal Relational Databases. Disponível em: <Coping with Events in Temporal Relational Databases>. Acesso em: 15 de Julho de 2017.

TERENZIANI, P. - Irregular Indeterminate Repeated Facts in Temporal Relational Databases," Disponível em: < <http://ieeexplore.ieee.org/document/7360163/>>. Acesso em: 15 de Julho de 2017

TERENZIANI, P. Nearly Periodic Facts in Temporal Relational Databases. **Ieee Transactions on Knowledge And Data Engineering**, [s.l.], v. 28, n. 10, p.2822-2826, 1 out. 2016. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/tkde.2016.2585483>.

TABLEAU. **Principais Tendencias para o Big Data**. Disponível em: <http://get.tableau.com/pt-br/asset/top-8-trends-big-data-2016.html> . Acesso em: 15 de Julho de 2017

TSUMOTO, S; HIRANO, Shoji. Multidimensional temporal mining in clinical data. **Proceedings of The 2nd Acm Sighit Symposium On International Health Informatics - Ihi '12**, [s.l.], p.563-572, 2012. ACM Press. <http://dx.doi.org/10.1145/2110363.2110426>.

ZENGER, B. Can Big Data Solve Healthcare's Big Problems? Disponível em: <https://www.equityhealthcare.com/docs/librariesprovider2/news-item-documents/eh-blog-on-analytics.pdf>>. Acesso em: 15 de Julho de 2017

ZHANG, S. et al. MTPGraph: A Data-Driven Approach to Predict Medical Risk Based on Temporal Profile Graph. **2016 Ieee Trustcom/bigdatase/ispa**, [s.l.], p.1174-1181, ago. 2016. IEEE. <http://dx.doi.org/10.1109/trustcom.2016.0191>.

ZANIOLO, C., CERI, S., FALOUTSOS, C., SNODGRASS, R.T., SUBRAHMANIAN, V.S., ZICARI, R. *Advanced Database Systems*. San Francisco, CA: Morgan Kaufmann Publishers, 1997. 574p.

WU, U.; JOJODIA, S.; WANG, X.S. Temporal Database Bibliography Update – 1997. Disponível em: < <http://isse.gmu.edu/~csis/tdb/bib97/bib97.html>>. Acesso em: 15 de Julho de 2017.

WANG, X., CHUSED, A., ELHADAD, N., FRIEDMAN, C., MARKATOU, M. Automated Knowledge Acquisition from Clinical Narrative Reports. *AMIA Annual Symposium Proceedings*. 2008; 2008:783-787.



WANG, T. et al. Aligning temporal data by sentinel events. **Proceeding of The Twenty-sixth Annual Chi Conference On Human Factors In Computing Systems - Chi '08**, [s.l.], p.457-466, 2008. ACM Press. <http://dx.doi.org/10.1145/1357054.1357129>.