

MINERAÇÃO DE TEXTO APLICADO A LAUDOS MÉDICOS: UMA ANÁLISE BIBLIOMÉTRICA

Fábio Barbosa BATISTA^{1*}, Geórgia Regina Rodrigues GOMES², Helder Gomes COSTA³

¹ Universidade Candido Mendes – UCAM – Campos dos Goytacazes - Rio de Janeiro, Brasil. Mestrando em Pesquisa Operacional e Inteligência Computacional

² Universidade Candido Mendes – UCAM – Campos dos Goytacazes - Rio de Janeiro, Brasil. Doutora em Informática

³ Universidade Candido Mendes – UCAM – Campos dos Goytacazes - Rio de Janeiro, Brasil. Doutor em Engenharia Mecânica

*Autor para correspondência. fbio_barbosa@hotmail.com

RESUMO

Este trabalho refere-se em uma análise bibliométrica abordando a aplicação do estudo de mineração de textos em laudos médicos. O objetivo do artigo é a realização de um levantamento bibliométrico que foi realizado através do Portal Periódico Capes. A pesquisa foi realizada tendo como palavra chave de pesquisa “Mining Text Information Extraction Medical Reports”. A pesquisa inicial somente utilizando a palavra chave, obteve 1.190 registros, sendo 1.636 resultados refinado por nível superior de periódicos revisados por pares. Foram aplicados refinamentos diante dos resultados encontrados, assim resultando em 47 artigos, conforme é descrito no decorrer do trabalho. Ao final da análise completa dos artigos selecionados, foram considerados 10 trabalhos para compor o núcleo de partida para a pesquisa, análise esta que, demonstra e comprovam as aplicações efetivas e satisfatórias da utilização da mineração de texto aplicadas a laudos médicos, oportunizando uma discussão positiva da utilização da técnica de mineração de texto na área da saúde.

Palavras chave: Mineração de textos, bibliometria e laudos médicos.

ABSTRACT

This work refers to a bibliometric analysis of the study addressing the application of text mining in medical reports. The aim of the paper is to conduct a bibliometric survey that was conducted by Journal Portal Capes. The survey was conducted having as keyword research "Mining Text Information Extraction Medical Reports". Initial research using only the keyword got 1,190 records, 1,636 being refined by top-level peer-reviewed journals results. Refinements were applied before the results found, resulting in 47 items , as described in this work . At the end of thorough analysis of the selected articles, 10 papers were considered to form the core for your research, this analysis that demonstrates and proves the effective and satisfactory application of the use of text mining applied to medical reports, providing opportunities for a positive discussion using the technique of text mining in healthcare.

Keywords: Text Mining , bibliometrics and medical reports.

1 – Introdução

Diante do avanço tecnológico acelerado e a necessidade do ser humano ao consumo de informações, surge à criação, manipulação e utilização de mega textos, principalmente os armazenamentos disponíveis na Web, formando então um grande repositório de informações digitais. Com isso existe uma grande necessidade de técnicas e ferramentas que auxiliem na facilitação de obtenção destas informações e também como ferramentas em buscas de novos conhecimentos, automaticamente, através destes grandes bancos de textos.

A Mineração de Texto (MT) é a aplicação da tecnologia onde objetiva a descoberta do conhecimento em texto. A tecnologia permite recuperar informações, extrair dados, resumir documentos, descobrir padrões, associar regras e realizar análises qualitativas ou quantitativas em documentos de texto. A tecnologia de mineração de texto nasceu a partir da necessidade de descobrir, de forma automática, informações, padrões e anomalias em textos. (Aranha and Passos 2006).

Segue abaixo algumas definições sobre mineração de textos:

- Text Mining realiza várias funções de busca, análise linguística e categorização. Mecanismos de busca se restringem à Internet. (Chen, 2001).
- Text Mining é o estudo e a prática de extrair informações de textos usando os princípios da linguística computacional. (SULLIVAN, D., 2000)
- Text Mining é ideal para inspecionar mudanças no mercado, ou para identificar ideias. (BIGGS, M., 2005)
- Text Mining é uma forma de examinar uma coleção de documentos e descobrir informação não contida em nenhum dos documentos. (LUCAS, M., 2000).
- Text Mining como sendo Data Mining em dados textuais. Text Mining tem como objetivo extrair padrões e associações desconhecidas de um grande banco de dados textual. (THURASINGHAM, B., 1999).
- Text Mining, como análise de dados exploratória, é um método para apoiar pesquisadores a derivar novas e relevantes informações de uma grande coleção de textos. É um processo parcialmente automatizado onde o pesquisador ainda está envolvido, interagindo com o sistema. (HEARST, M. A., 1999).
- Pode-se então definir Descoberta de Conhecimento em Textos (KDT) ou Text Mining como sendo o processo de extrair padrões ou conhecimento, interessantes e não triviais, a partir de documentos textuais. (TAN, A.H., 1999).

Conforme COSTA *et al* (2012) o processo de descoberta de conhecimento textual é composto das etapas de pré-processamento, com a finalidade de preparar, transformar, organizar e melhorar a qualidade do texto para a etapa seguinte, o processamento (mineração), que é o objetivo, onde as técnicas aplicadas variam de acordo com a finalidade, recuperação da informação, indexação, extração da informação, associação de documentos, sumarização, clusterização e classificação/categorização. Por fim a etapa de pós-processamento avalia os resultados da mineração. Todo esse processo foi totalmente originado do modelo Knowledge Discovery in Databases (KDD – Mineração de dados), onde o processo é baseado em dados estruturados, diferente da mineração de texto, que é uma subárea da inteligência computacional que tem a finalidade de extrair padrões relevantes de dados não estruturados.

A proposta da análise bibliometria proposta por COSTA (2010), tem por finalidade aplicar um modelo de mineração de fontes bibliográficas baseados em ferramentas de acesso e busca a dados bibliográficos, baseados na Internet, objetivando a seleção de um conjunto de referências bibliográfica para servir como apoio a um determinado estudo.

O modelo proposto proporciona uma investigação entre as referências assim resultando em um estudo preliminar contendo trabalhos que vão compor um “núcleo de partida”, onde o pesquisador terá embasamento suficiente para o próprio estudo bibliográfico.

COSTA (2010) destaca o modelo em 6 partes:

- Definição da amostra;
- Pesquisa da Amostra;
- Identificação dos periódicos com maior número de artigos publicados;
- Identificação dos autores com maior número de publicações;
- Levantamento da cronologia da produção;
- Identificação do “núcleo de partida”.

Neste trabalho foi utilizada a pesquisa no sistema de busca do de artigos indexados no Portal de Periódicos da Capes em dezembro de 2013. A mineração de texto aplicado a laudos médicos foi o tema pesquisado na amostra sendo utilizadas as palavras chaves “*Mining Text Information Extraction Medical Reports*”, retornando 1.190 registros, onde a busca passou por refinamentos e obtendo 10 registros para análise final, conforme é detalhado no próximo item, como também as identificações dos periódicos, autores, cronologia e os registros que serviram com base para o núcleo de partida da análise bibliométrica.

Tem-se como objetivo final a identificação, após análises, artigos que possam identificar a aplicação da mineração de texto em laudos médicos. Destacando o grande interesse da área médica nesta tecnologia.

2 – Materiais e Métodos

A busca foi realizada através da palavra chave “*mining text information extraction medical reports*” obtendo 1.190 registros, sendo 1.636 resultados refinado por nível superior de periódicos revisados por pares, os quais são apresentados no quadro 01 quanto a distribuição dos tipos de documentos que foram retornados na pesquisa.

Quadro 1: Distribuição por tipos de publicação de registros encontrados.

Tipos de Publicação	Quantidade de registros
Artigos	1498
Resenhas	477
Recursos textuais	172
Atas de congressos	55

Na sequência são apresentados nas figuras (Figura 1), (Figura 2), (Figura 3), os tópicos relacionados à pesquisa, autores destacados e títulos dos períodos dos registros encontrados, respectivamente.

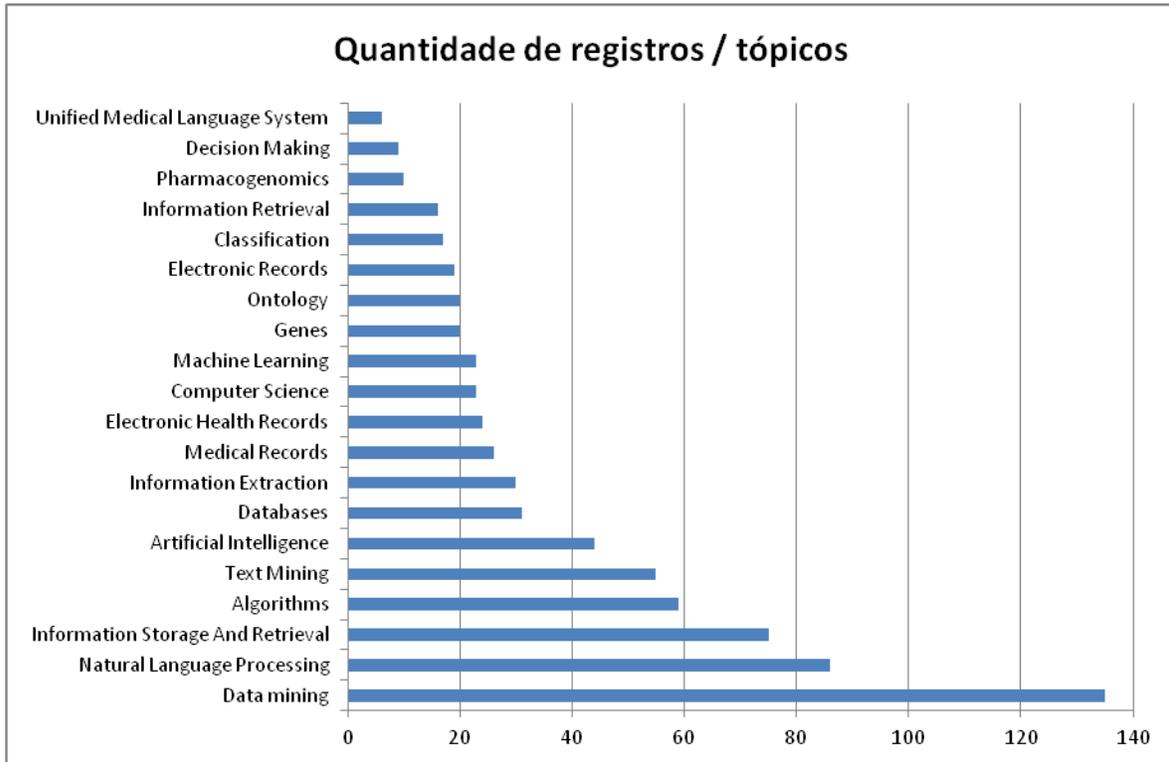


Figura 1: Tópicos relacionados às palavras chaves pesquisadas.

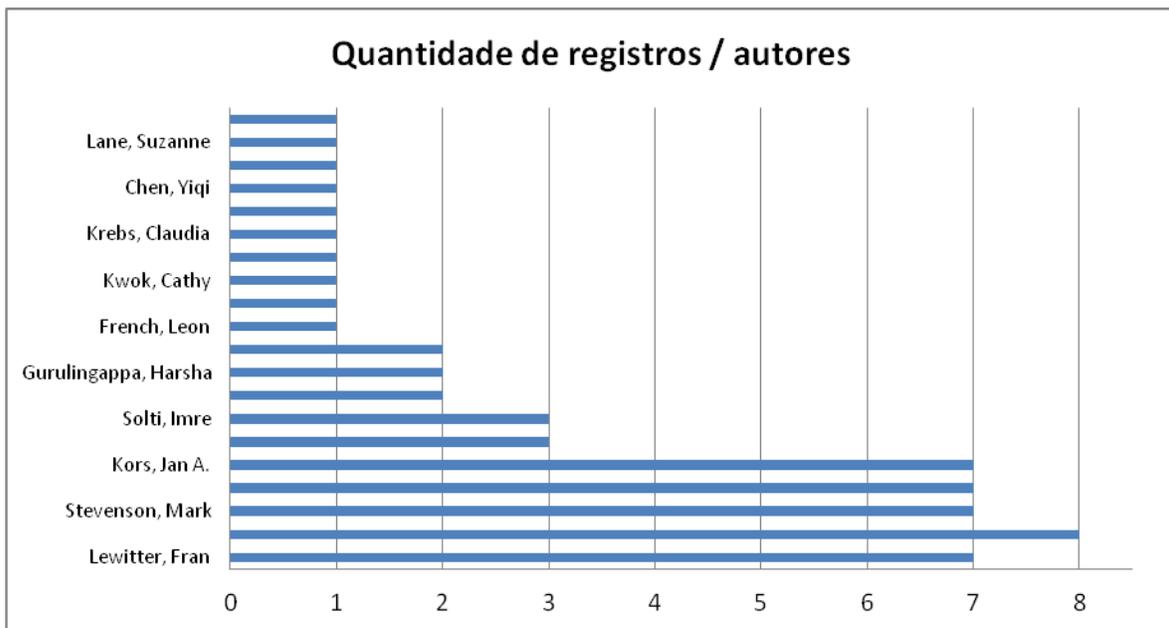


Figura 2: Distribuição de Autores dos registros encontrados

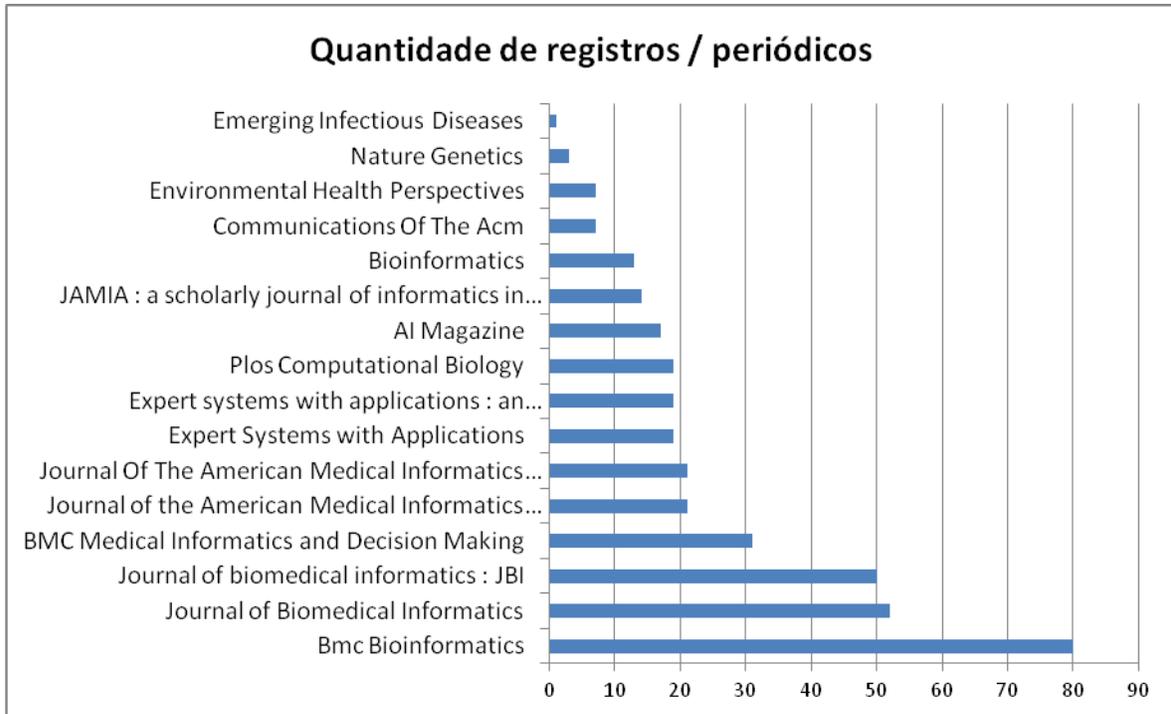


Figura 3: Títulos dos períodos dos registros encontrados

O idioma predominante entre as publicações encontradas na pesquisa foi de origem da língua inglesa, isto é 94,5%. O que é destacado na figura 05 é a ordem cronológica das referentes publicações encontradas, onde demonstra uma grade disseminação do tema após a década de 2000.



Figura 4 – Análise por anos de publicação

3 - Resultados

Para a realização da análise dos artigos foram realizados alguns outros refinamentos entre os registros encontrados: Foram consideradas somente as publicações em artigos, após o ano de 2009 e restringindo aos tópicos relacionados à Information Extraction (Extração da Informação), Text Mining (Mineração de Texto) e Natural Language Processing (Processamento de Linguagem Natural) relacionados a laudos médicos, assim obtendo um conjunto de 47 artigos selecionados, entretanto, após a análise individual dos temas específicos dos artigos, a relação de selecionados foi finalizada em 10 artigos, conforme quadro 2.

Quadro 2: Artigos encontrados após refino

Extraction of Radiology Reports using Text mining A.v.krishna Prasad ; Dr.s.ramakrishna; Dr. d. sravan Kumar ; Dr.b.padmaja Rani International Journal on Computer Science and Engineering, 2010, Vol.2(5), p.1558
Text mining for traditional Chinese medical knowledge discovery: A survey Zhou, Xuezhong ; Peng, Yonghong ; Liu, Baoyan Journal of Biomedical Informatics, 2010, Vol.43(4), pp.650-660
Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure Garvin, Jennifer H ; Duvall, Scott L ; South, Brett R ; Bray, Bruce E ; Bolton, Daniel ; Heavirland, Julia ; Pickard, Steve ; Heidenreich, Paul ; Shen, Shuying ; Weir, Charlene ; Samore, Matthew ; Goldstein, Mary K Journal of the American Medical Informatics Association : JAMIA, 2012, Vol.19(5), pp.859-66
Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text George Steven Bova Journal of the American Medical Informatics Association, 2013, Vol.20(5), p.898
Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010 de Bruijn, Berry; Cherry, Colin; Kiritchenko, Svetlana; Martin, Joel; Zhu, Xiaodan Journal of the American Medical Informatics Association, 2011, Vol.18(5), p.557
Detecting temporal expressions in medical narratives Reeves, Ruth M. ; Ong, Ferdo R. ; Matheny, Michael E. ; Denny, Joshua C. ; Aronsky, Dominik ; Gobbel, Glenn T. ; Montella, Diane ; Speroff, Theodore ; Brown, Steven H. International Journal of Medical Informatics, 2013, Vol.82(2), pp.118-127
A review on Relation Extraction with an eye on Portuguese Abreu, Sandra ; Bonamigo, Tiago ; Vieira, Renata Journal of the Brazilian Computer Society, 2013, Vol.19(4), pp.553-571
Automating classification of free-text electronic health records for epidemiological studies Schuemie, Martijn J ; Sen, Emine ; 't Jong, Geert W ; Van Soest, Eva M ; Sturkenboom, Miriam C ; Kors, Jan APharmacoepidemiology and drug safety, 2012, Vol.21(6), pp.651-8
Using rule-based natural language processing to improve disease normalization in biomedical text Ning Kang, Bharat Singh, Zubair Afzal, Erik M van Mulligen, Jan A Kors J Am Med Inform Assoc. 2013 September; 20(5): 876–881.

Importante destacar que apenas um artigo foi selecionado na língua portuguesa, utilizando as palavras chaves “Mineração de texto laudos”, apresentado no quadro 3.

Quadro 3: Artigo encontrados na língua portuguesa

Uso de mineração de texto como ferramenta de avaliação da qualidade informacional em laudos eletrônicos de mamografia. Paulo Roberto Barbosa Serapião ; Kátia Mitiko Firmino Suzuki ; Paulo Mazzoncini De Azevedo Marques. Radiol Bras vol.43 no.2 São Paulo Mar./Apr. 2010, pg:103- 107

4 – Discussão

Após a análise dos registros, foram selecionados 10 (dez) artigos com maior relevância para uma análise bibliométrica detalhada. Ficam destacados os artigos que realmente são relacionados às palavras chaves de interesse. Abaixo está relacionado um resumo de cada artigo selecionado, como também o artigo encontrado publicado em língua portuguesa através de uma busca utilizando a palavra chave em português.

Prasad *et al* (2010) propuseram um sistema de mineração de texto para extrair e utilizar a informação em relatórios de radiologia. O sistema é composto por três módulos principais: extrator constatação médica, relatórios e imagem retrieve. O módulo de extração de constatação médica extrai automaticamente descobertas médicas e modificadores associados para estruturar relatórios de radiologia. A estruturação dos relatórios de texto livre preenche a lacuna entre os usuários e banco de dados relatório, faz com que a informação contida nos relatórios de fácil acesso. Também serve como resultado intermediário de outros componentes do sistema. O módulo de recuperação analisa consulta do usuário e retorna os relatórios e as imagens que correspondem a consulta. Os resultados são satisfatórios overall evaluation, porém são necessários mais testes e uma avaliação. Nosso trabalho futuro inclui a melhoria do desempenho atual do sistema e implementação do sistema de geração de relatórios de radiologia utilizando abordagem de tradução automática estatística, para o qual nós projetamos a arquitetura geral.

Text mining for traditional Chinese medical knowledge Discovery: Este trabalho aborda a extração de informações e conhecimentos de texto livre, onde destaca que o tema de interesse considerável nos campos de aprendizado de máquina e mineração. Destaca também que os avanços mais significativos na área de mineração de texto na biomédica durante os últimos anos demonstraram a sua grande promessa para apoiar os cientistas no desenvolvimento de novas hipóteses e novos conhecimentos a partir da literatura biomédica. O artigo apresenta a Medicina tradicional chinesa (TCM) como fonte significativa de informação complementar para ciências biomédicas modernas e que a Literatura TCM obtido a partir do período histórico e de estudos clínicos modernos foi recentemente transformada em dados digitais, na forma de bancos de dados relacionais ou documentos de texto, que fornecem uma plataforma eficaz para a partilha de informação e recuperação. Isso motiva e facilita a pesquisa e desenvolvimento de abordagens de descoberta de conhecimento e para a modernização do TCM. A fim de contribuir para esta área continua a crescer , este artigo apresenta uma introdução comparativa ao TCM e biomedicina moderna , um levantamento das fontes de informação do TCM , uma revisão e discussão do estado da arte e o desenvolvimento de técnicas de mineração de texto com aplicações à TCM, uma discussão sobre as questões de pesquisa ao redor TCM mineração de texto e suas direções futuras. Xuezhong; Yonghong; Baoyan (2010).

Garvin *et al* (2012) apresenta um trabalho que objetivou em construir um sistema de processamento de linguagem natural para extrair o EF (Fração de ejeção ventricular esquerda) a partir de relatórios do eco cardiograma de texto livre para automatizar relatórios de medição e para validar a precisão do sistema usando um padrão de referência de comparação desenvolvido através de revisão humana. Foi criado um conjunto de expressões regulares e regras para capturar o EF com uma amostra aleatória de 765 eco cardiogramas de sete centros médicos VA (Departamento de Assuntos de Veteranos). Os documentos foram distribuídos aleatoriamente em dois conjuntos: um conjunto de 275 usado para treinamento e um segundo conjunto de 490 usado para testes e validação. Para estabelecer o padrão de referência, dois revisores independentes anotaram todos os documentos em ambos os conjuntos, um terceiro revisor adjudicada desentendimentos.

Bova, G. S. (2013) testou a viabilidade da utilização de mineração de texto para descrever de forma significativa a experiência da dor em pacientes com câncer de próstata metastático, para identificar novos fenótipos dor e propor métodos para visualização longitudinal do status dor. Painéis gráficos revelaram os dores ‘paisagem’ descritos nos registros textuais e confirmou dramaticamente o aumento dos níveis de dor nos últimos anos de vida em tudo, mas dois pacientes, todos eles morreram de câncer metastático. A dor severa foi associada com a recepção de opioides e radioterapia paliativa.

Bruijn *et al* (2011) descrevem o design e desempenho de três aplicações de mineração de texto state-of-the-art do Conselho Nacional de Pesquisa do Canadá em avaliações dentro do desafio i2b2 2010. Os três sistemas executaram três passos fundamentais na extração de informações clínicas: a extração de problemas médicos, exames e tratamentos, a partir de resumos de alta e notas de progresso, classificação das afirmações feitas sobre os problemas de saúde, classificação das relações entre conceitos médicos. Sistemas de aprendizado de máquina executadas essas tarefas usando sacos de grandes dimensões de recursos, como derivado de tanto o próprio texto e de fontes externas: UMLS, cTAKES e Medline. Os sistemas de classificação alta entre todos os sistemas apresentados resultaram em respostas bastante satisfatórias. Verificou-se também que a introdução de uma ampla gama de recursos foi fundamental para o sucesso.

A tarefa de Extração de Relação é um dos principais desafios em Processamento de Língua Natural. Apresentaram uma revisão do estado-da-arte para Extração Relação de textos livres, abordando os avanços e dificuldades da área, e situando Português nesse quadro. Discutimos os diferentes aspectos relacionados com esta tarefa, considerando as principais estratégias computacionais, os recursos utilizados, bem como os métodos de avaliação aplicados. Também damos atenção especial à literatura para obter ferramentas portuguesas, que precisam de mais progressos. No melhor de nosso conhecimento, este é o primeiro estudo abrangente sobre Extração Relação de incluir o estado do trabalho realizado para o Português. Abreu; Bonamigo; Vieira (2013).

Cada vez mais, as informações do paciente são armazenadas em registros médicos eletrônicos, que podem ser reutilizados para a pesquisa, afirmam Schuemie *et al* (2012). Muitas vezes, estes registros incluem dados narrativos não estruturados, que são complicados de analisar. Os autores investigaram se a mineração de texto pode tornar esses dados adequados para estudos epidemiológicos e comparou uma abordagem de reconhecimento de conceito e uma variedade de técnicas de aprendizado de máquina que requerem um conjunto de treinamento anotado

manualmente. Os autores mostram como este conjunto de treinamento pode ser criado com o mínimo esforço usando uma consulta de banco de dados amplo.

Ning *et al* (2013) compararam o desempenho de dois sistemas biomédicos conceito de normalização, MetroMeta e Peregrine, no Arizona Disease Corpus, com e sem o uso de um módulo de PNL baseada em regras. O desempenho foi avaliado para exata e inexata correspondente limite das anotações do sistema com as do padrão-ouro e de correspondência de identificador conceito. Sem o módulo de PNL, MetroMeta e Peregrine atingiu F -score de 61,0% e 63,9%, respectivamente, para correspondência exata fronteira, e 55,1% e 56,9% para correspondência identificador conceito. Com a ajuda do módulo de PNL, os F- dezenas de MetroMeta e Peregrine melhorou para 73,3% e 78,0% para a correspondência de fronteira, e para 66,2% e 69,8% para a correspondência de identificador conceito. Para inexata correspondente limite, performances aumentada para 85,5% e 85,4%, e 73,6% e 73,3% para a correspondência de identificador conceito. O valor acrescentado da PNL para o reconhecimento e normalização de doenças com MetroMeta e Peregrine. O módulo PNL é geral e pode ser aplicado em combinação com qualquer sistema conceito normalização. Se o seu uso para o conceito do que outros tipos de doença são restos igualmente vantajosos para serem investigados.

Serapião *et al* (2010) investigaram o uso da técnica de mineração de texto como forma de avaliar a qualidade Informacional de laudos eletrônicos de mamografia, tendo como parâmetro de qualidade a adesão ao léxico BI-RADS®. **MATERIAIS E MÉTODOS:** Foram extraídos 22.247 laudos de mamografia do banco de dados do sistema de informação em radiologia do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, no período de janeiro de 2000 até junho de 2006. Foram realizados dois experimentos, um buscando-se verificar a utilização mais correta dos termos do léxico - experimento 1 (especificidade do método de mineração), e outro buscando-se verificar toda e qualquer tentativa de uso ou alusão ao léxico - experimento 2 (sensibilidade do método de mineração). **RESULTADOS:** Experimento 1: variação entre 11% e 61% de laudos contendo termos do léxico em sua conclusão, distribuída de forma aleatória ao longo do tempo, a partir do ano de 2001. Experimento 2: variação entre 44% e 100% de laudos que se referem de alguma forma ao léxico em sua conclusão. Os resultados indicam um bom potencial da aplicação da ferramenta de mineração de texto para a avaliação da qualidade das informações contidas em laudos eletrônicos de mamografia.

5 – Conclusão

Este trabalho apresentou uma pesquisa sobre a aplicação da tecnologia de mineração de texto a laudos médicos, com o objetivo de apresentar uma seleção de artigos como ponto de partida para uma análise de trabalhos mais relevantes referentes ao tema proposto.

Os resultados dos artigos selecionados foram obtidos através do portal periódicos Capes, apesar de ter sido satisfatória, sendo encontrados trabalhos muito relevantes na área, entretanto, vale resaltar a existência de dificuldades no sistema de busca, que podem ser melhores em outros portais.

O modelo proposto por Costa (2010), demonstrou ser muito eficiente, o qual proporciona ao pesquisador uma base fundamental para iniciar uma pesquisa, onde sua pesquisa é fundamentada em trabalhos com maiores credibilidades, com resultados obtidos a partir de refinamentos identificando trabalhos com maior números de artigos já publicados, maior número

de autores com publicações relacionadas e ainda podendo identificar e analisar a cronologia das produções.

Em relação à aplicação da tecnologia de mineração de texto, a área de saúde demonstra um avançado interesse na tecnologia. Foi possível identificar centenas de trabalhos da área de saúde voltada à utilização da tecnologia, inclusive como nossa pesquisa, o caso da aplicabilidade da tecnologia de mineração de texto em laudos médicos. Como tratado no início deste trabalho, com o avanço exorbitante da tecnologia causando mega textos, principalmente na web, todos os setores/áreas estão se beneficiando da tecnologia de mineração de texto, onde demonstra, não só a facilidade de manipulação, como também a necessidade de obter novos conhecimentos a partir destes materiais que cada vez cresce mais.

6 - Referências

ABREU, Sandra; BONAMIGO, Tiago; VIEIRA, Renata. A review on Relation Extraction with an eye on Portuguese. *Journal of the Brazilian Computer Society*, Vol.19(4), pp.553-571, 2013.

ARANHA, Christian; PASSOS, Emmanuel. A Tecnologia de Mineração de Textos. *RESI-Revista Eletrônica de Sistemas de Informação*, N°2-2006. Lab.ICA Elétrica PUC-Rio, 2006.

BIGGS, M. (2005). Resurgent text-mining technology can greatly increase your firm's 'intelligence' factor. *InfoWorld* 11(2), 52

BOVA, G.S. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. *Journal of the American Medical Informatics Association*, Vol.20(5), p.898, 2013.

BRUIJN, Berry; CHERRY, Colin; KIRITCHENKO, Svetlana; MARTIN, Joel; ZHU, Xiaodan. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2. *Journal of the American Medical Informatics Association*, 2011, Vol.18(5), p.557, 2010.

CHEN, H. (1994). *A Textual Database/Knowledge-Base Coupling Approach to Creating Computer-Supported Organizational Memory*. MIS Department. University of Arizona.

COSTA, Helder. Modelo para webibliomining: proposta e caso de aplicação. *Rev. FAE*, Curitiba, v13, n.1, p. 115-126, jan./jun.2010.

GARVIN, J. H. *et al.* Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure *Journal of the American Medical Informatics Association: JAMIA*, Vol.19(5), pp.859-66, 2012.

GOMES, Georgia Regina Rodrigues; XAVIER, Bruno Missi; SILVA, Alcione Dias da; COSTA, Helder. Mineração de Texto e suas aplicações na literatura científica – Estudo Bibliométrico, *Perspectivas Online: Exatas e Engenharia*, vol. 2, No 4, 2012.

HEARST, M. A. (1999). *Untangling Text Data Mining*. Proceedings of the ACL'99: the 37th Annual Meeting of the Association

KANG, N. *et al.*, Using rule-based natural language processing to improve disease normalization in biomedical text *Am Med Inform Assoc*. 2013 September; 20(5): 876–881.

LUCAS, M. (2000). Mining in textual mountains, an interview with Marti Hearst. *Mappa Mundi Magazine*, Trip-M, 005, 1-3.

PRASAD, A. V. K. *et al.* Extraction of Radiology Reports using Text mining International. *Journal on Computer Science and Engineering*, Vol.2(5), p.1558, 2010.

REEVES, Ruth M. ; ONG, Ferdo R. ; MATHENY, Michael E. ; DENNY, Joshua C. ; ARONSKY, Dominik ; GOBBEL, Glenn T. ; MONTELLA, Diane ; SPEROFF, Theodore ; BROWN, Steven H. Detecting temporal expressions in medical narratives. *International Journal of Medical Informatics*, Vol.82(2), pp.118-127, 2013.

SCHUEMIE, Martijn J; SEN, Emine; 'T JONG, Geert W; VAN SOEST, Eva M; STURKENBOOM, Miriam C; KORS, Jan A. Automating classification of free-text electronic health records for epidemiological studies. *Pharmacoepidemiology and drug safety*, Vol.21(6), pp.651-8, 2012,

SERAPIÃO, Paulo Roberto Barbosa; SUZUKI, Kátia Mitiko Firmino; MARQUES, Paulo Mazzoncini de Azevedo. Uso de mineração de texto como ferramenta de avaliação da qualidade informacional em laudos eletrônicos de mamografia. *Radiol Bras* vol.43 no.2, pg:103- 107, São Paulo Mar./Apr. 2010.

SULLIVAN, D. (2000). *The need for text mining in business intelligence*. DM Review, Dec. 2000.

TAN, A.-H. (1999). Text mining: The state of the art and the challenges. In Proceedings, PAKDD'99 workshop on Knowledge Discovery from Advanced Databases, *Beijing*, pp.65-70

THURASINGHAM, B. (1999). *Data mining: technologies, techniques, tools, and trends*. CRC Press, Boca Raton, Florida.

ZHOU, Xuezhong; PENG, Yonghong; LIU, Baoyan. Text mining for traditional Chinese medical knowledge discovery: *Journal of Biomedical Informatics*, Vol.43(4), pp.650-660, 2010.